IBM Spectrum Scale
Version 5.0.1

Problem Determination Guide

SC27-9264-03

IBM Spectrum Scale
Version 5.0.1

Problem Determination Guide

SC27-9264-03

Note
Before using this information and the product it supports, read the information in "Notices" on page 717.

This edition applies to version 5 release 0 modification 1 of the following products, and to all subsequent releases and modifications until otherwise indicated in new editions:

- IBM Spectrum Scale ordered through Passport Advantage® (product number 5725-Q01)
- IBM Spectrum Scale ordered through AAS/eConfig (product number 5641-GPF)
- IBM Spectrum Scale for Linux on Z (product number 5725-S28)
- IBM Spectrum Scale for IBM ESS (product number 5765-ESS)

Significant changes or additions to the text and illustrations are indicated by a vertical line (|) to the left of the change.

IBM welcomes your comments; see the topic "How to send your comments" on page xxv. When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

Chapter 16. Upgrade issues 313

Chapter 17. Network issues 315

Chapter 18. File system issues 317

Tables

About this information

This edition applies to IBM Spectrum Scale™ version 5.0.1 for AIX®, Linux, and Windows.

IBM Spectrum Scale is a file management infrastructure, based on IBM® General Parallel File System (GPFS™) technology, which provides unmatched performance and reliability with scalable access to critical file data.

To find out which version of IBM Spectrum Scale is running on a particular AIX node, enter:

```
lslpp -l gpfs\*
```

To find out which version of IBM Spectrum Scale is running on a particular Linux node, enter:

```
rpm -qa | grep gpfs     (for SLES and Red Hat Enterprise Linux)
dpkg -l | grep gpfs     (for Ubuntu Linux)
```

To find out which version of IBM Spectrum Scale is running on a particular Windows node, open **Programs and Features** in the control panel. The IBM Spectrum Scale installed program name includes the version number.

Which IBM Spectrum Scale information unit provides the information you need?

The IBM Spectrum Scale library consists of the information units listed in Table 1 on page xii.

To use these information units effectively, you must be familiar with IBM Spectrum Scale and the AIX, Linux, or Windows operating system, or all of them, depending on which operating systems are in use at your installation. Where necessary, these information units provide some background information relating to AIX, Linux, or Windows. However, more commonly they refer to the appropriate operating system documentation.

Note: Throughout this documentation, the term "Linux" refers to all supported distributions of Linux, unless otherwise specified.

Table 1. IBM Spectrum Scale library information units

Information unit	Type of information	Intended users
IBM Spectrum Scale: Concepts, Planning, and Installation Guide	This guide provides the following information: **Product overview** • Overview of IBM Spectrum Scale • GPFS architecture • Protocols support overview: Integration of protocol access methods with GPFS • Active File Management • AFM-based Asynchronous Disaster Recovery (AFM DR) • Data protection and disaster recovery in IBM Spectrum Scale • Introduction to IBM Spectrum Scale GUI • IBM Spectrum Scale management API • Introduction to Cloud services • Introduction to file audit logging • IBM Spectrum Scale in an OpenStack cloud deployment • IBM Spectrum Scale product editions • IBM Spectrum Scale license designation • Capacity based licensing • IBM Spectrum Storage™ Suite **Planning** • Planning for GPFS • Planning for protocols • Planning for Cloud services • Firewall recommendations • Considerations for GPFS applications	System administrators, analysts, installers, planners, and programmers of IBM Spectrum Scale clusters who are very experienced with the operating systems on which each IBM Spectrum Scale cluster is based

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Concepts, Planning, and Installation Guide	**Installing** • Steps for establishing and starting your IBM Spectrum Scale cluster • Installing IBM Spectrum Scale on Linux nodes and deploying protocols • Installing IBM Spectrum Scale on AIX nodes • Installing IBM Spectrum Scale on Windows nodes • Installing Cloud services on IBM Spectrum Scale nodes • Installing and configuring IBM Spectrum Scale management API • Installing Active File Management • Installing and upgrading AFM-based Disaster Recovery • Installing call home • Installing file audit logging • Steps to permanently uninstall GPFS and/or Protocols	System administrators, analysts, installers, planners, and programmers of IBM Spectrum Scale clusters who are very experienced with the operating systems on which each IBM Spectrum Scale cluster is based

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Concepts, Planning, and Installation Guide	**Upgrading** • IBM Spectrum Scale supported upgrade paths • Upgrading to IBM Spectrum Scale 5.0.x from IBM Spectrum Scale 4.2.y • Upgrading to IBM Spectrum Scale 4.2.y from IBM Spectrum Scale 4.1.x • Upgrading to IBM Spectrum Scale 4.1.1.x from GPFS V4.1.0.x • Upgrading from GPFS 3.5 • Online upgrade support for protocols and performance monitoring • Upgrading AFM and AFM DR • Upgrading object packages • Upgrading NFS packages • Upgrading SMB packages • Upgrading call home configuration • Manually upgrading pmswift • Manually upgrading the performance monitoring tool • Manually upgrading the IBM Spectrum Scale management GUI • Upgrading Cloud services • Upgrading to IBM Cloud Object Storage software level 3.7.2 and above • Manually upgrading file audit logging • Upgrading IBM Spectrum Scale components with the installation toolkit • Migrating from Express Edition to Standard Edition • Completing the upgrade to a new level of IBM Spectrum Scale • Reverting to the previous level of IBM Spectrum Scale • Coexistence considerations • Compatibility considerations • Considerations for IBM Spectrum Protect™ for Space Management • GUI user role considerations • Applying maintenance to your GPFS system	System administrators, analysts, installers, planners, and programmers of IBM Spectrum Scale clusters who are very experienced with the operating systems on which each IBM Spectrum Scale cluster is based

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Administration Guide	This guide provides the following information: **Configuring** • Configuring the GPFS cluster • Configuring the CES and protocol configuration • Configuring and tuning your system for GPFS • Parameters for performance tuning and optimization • Ensuring high availability of the GUI service • Configuring and tuning your system for Cloud services • Configuring file audit logging • Configuring Active File Management • Configuring AFM-based DR • Tuning for Kernel NFS backend on AFM and AFM DR **Administering** • Performing GPFS administration tasks • Verifying network operation with the mmnetverify command • Managing file systems • File system format changes between versions of IBM Spectrum Scale • Managing disks • Managing protocol services • Managing protocol user authentication • Managing protocol data exports • Managing object storage • Managing GPFS quotas • Managing GUI users • Managing GPFS access control lists • Considerations for GPFS applications • Accessing a remote GPFS file system	System administrators or programmers of IBM Spectrum Scale systems

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Administration Guide	• Information lifecycle management for IBM Spectrum Scale • Creating and maintaining snapshots of file systems • Creating and managing file clones • Scale Out Backup and Restore (SOBAR) • Data Mirroring and Replication • Implementing a clustered NFS environment on Linux • Implementing Cluster Export Services • Identity management on Windows • Protocols cluster disaster recovery • File Placement Optimizer • Encryption • Managing certificates to secure communications between GUI web server and web browsers • Securing protocol data • Cloud services: Transparent cloud tiering and Cloud data sharing • Managing file audit logging • Highly-available write cache (HAWC) • Local read-only cache • Miscellaneous advanced administration • GUI limitations	System administrators or programmers of IBM Spectrum Scale systems

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Problem Determination Guide	This guide provides the following information: **Monitoring** • Performance monitoring • Monitoring system health through the IBM Spectrum Scale GUI • Monitoring system health by using the mmhealth command • Monitoring events through callbacks • Monitoring capacity through GUI • Monitoring AFM and AFM DR • GPFS SNMP support • Monitoring the IBM Spectrum Scale system by using call home • Monitoring remote cluster through GUI • Monitoring file audit logging **Troubleshooting** • Best practices for troubleshooting • Understanding the system limitations • Collecting details of the issues • Managing deadlocks • Installation and configuration issues • Upgrade issues • Network issues • File system issues • Disk issues • Security issues • Protocol issues • Disaster recovery issues • Performance issues • GUI issues • AFM issues • AFM DR issues • Transparent cloud tiering issues • File audit logging issues • Maintenance procedures • Recovery procedures • Support for troubleshooting • References	System administrators of GPFS systems who are experienced with the subsystems used to manage disks and who are familiar with the concepts presented in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Command and Programming Reference	This guide provides the following information: **Command reference** • gpfs.snap command • mmaddcallback command • mmadddisk command • mmaddnode command • mmadquery command • mmafmconfig command • mmafmctl command • mmafmlocal command • mmapplypolicy command • mmaudit command • mmauth command • mmbackup command • mmbackupconfig command • mmblock command • mmbuildgpl command • mmcachectl command • mmcallhome command • mmces command • mmcesdr command • mmchattr command • mmchcluster command • mmchconfig command • mmchdisk command • mmcheckquota command • mmchfileset command • mmchfs command • mmchlicense command • mmchmgr command • mmchnode command • mmchnodeclass command • mmchnsd command • mmchpolicy command • mmchpool command • mmchqos command • mmclidecode command • mmclone command • mmcloudgateway command • mmcrcluster command • mmcrfileset command • mmcrfs command • mmcrnodeclass command • mmcrnsd command • mmcrsnapshot command	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Command and Programming Reference	• mmdefedquota command • mmdefquotaoff command • mmdefquotaon command • mmdefragfs command • mmdelacl command • mmdelcallback command • mmdeldisk command • mmdelfileset command • mmdelfs command • mmdelnode command • mmdelnodeclass command • mmdelnsd command • mmdelsnapshot command • mmdf command • mmdiag command • mmdsh command • mmeditacl command • mmedquota command • mmexportfs command • mmfsck command • mmfsctl command • mmgetacl command • mmgetstate command • mmhadoopctl command • mmhealth command • mmimgbackup command • mmimgrestore command • mmimportfs command • mmkeyserv command • mmlinkfileset command • mmlsattr command • mmlscallback command • mmlscluster command • mmlsconfig command • mmlsdisk command	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Command and Programming Reference	• mmlsfileset command • mmlsfs command • mmlslicense command • mmlsmgr command • mmlsmount command • mmlsnodeclass command • mmlsnsd command • mmlspolicy command • mmlspool command • mmlsqos command • mmlsquota command • mmlssnapshot command • mmmigratefs command • mmmount command • mmmsgqueue command • mmnetverify command • mmnfs command • mmnsddiscover command • mmobj command • mmperfmon command • mmpmon command • mmprotocoltrace command • mmpsnap command • mmputacl command • mmquotaoff command • mmquotaon command • mmremotecluster command • mmremotefs command • mmrepquota command • mmrestoreconfig command • mmrestorefs command • mmrestripefile command • mmrestripefs command • mmrpldisk command • mmsdrrestore command • mmsetquota command • mmshutdown command • mmsmb command • mmsnapdir command • mmstartup command • mmtracectl command • mmumount command • mmunlinkfileset command • mmuserauth command • mmwinservctl command • spectrumscale command	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Command and Programming Reference	**Programming reference** • IBM Spectrum Scale Data Management API for GPFS information • GPFS programming interfaces • GPFS user exits • IBM Spectrum Scale management API commands	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Big Data and Analytics Guide	This guide provides the following information: **IBM Spectrum Scale support for Hadoop** • HDFS transparency • Supported IBM Spectrum Scale storage modes • Hadoop cluster planning • Installation and configuration of HDFS transparency • Application interaction with HDFS transparency • Upgrading the HDFS Transparency cluster • Rolling upgrade for HDFS Transparency • Security • Advanced features • Hadoop distribution support • Limitations and differences from native HDFS • Problem determination **BigInsights® 4.2.5 and Hortonworks Data Platform 2.6** • **Planning** – Hardware requirements – Preparing the environment – Preparing a stanza file • **Installation** – Set up – Installation of software stack – BigInsights value-add services on IBM Spectrum Scale • **Upgrading software stack** – Migrating from BI IOP to HDP – Upgrading IBM Spectrum Scale service MPack – Upgrading HDFS Transparency – Upgrading IBM Spectrum Scale file system – Upgrading to BI IOP 4.2.5	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Table 1. IBM Spectrum Scale library information units (continued)

Information unit	Type of information	Intended users
IBM Spectrum Scale: Big Data and Analytics Guide	• **Configuration** – Setting up High Availability [HA] – IBM Spectrum Scale configuration parameter checklist – Dual-network deployment – Manually starting services in Ambari – Setting up local repository – Configuring LogSearch – Setting IBM Spectrum Scale configuration for BigSQL – Hadoop Kafka/Zookeeper and IBM Spectrum Scale Kafka/Zookeeper • **Administration** – IBM Spectrum Scale-FPO deployment – Ranger – Kerberos – Short-circuit read (SSR) – Disabling short circuit write – IBM Spectrum Scale service management – Ambari node management – Restricting root access – IBM Spectrum Scale management GUI – IBM Spectrum Scale versus Native HDFS • **Troubleshooting** – Snap data collection • **Limitations** – Limitations and information • **FAQ** – General – Service fails to start – Service check failures	• System administrators of IBM Spectrum Scale systems • Application programmers who are experienced with IBM Spectrum Scale systems and familiar with the terminology and concepts in the XDSM standard

Prerequisite and related information

For updates to this information, see IBM Spectrum Scale in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html).

For the latest support information, see the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

Conventions used in this information

Table 2 describes the typographic conventions used in this information. UNIX file name conventions are used throughout this information.

Note: Users of IBM Spectrum Scale for Windows must be aware that on Windows, UNIX-style file names need to be converted appropriately. For example, the GPFS cluster configuration data is stored in the `/var/mmfs/gen/mmsdrfs` file. On Windows, the UNIX namespace starts under the `%SystemDrive%\cygwin64` directory, so the GPFS cluster configuration data is stored in the `C:\cygwin64\var\mmfs\gen\mmsdrfs` file.

Table 2. Conventions

Convention	Usage
bold	**Bold** words or characters represent system elements that you must use literally, such as commands, flags, values, and selected menu options. Depending on the context, **bold** typeface sometimes represents path names, directories, or file names.
bold underlined	**bold underlined** keywords are defaults. These take effect if you do not specify a different keyword.
`constant width`	Examples and information that the system displays appear in `constant-width` typeface. Depending on the context, `constant-width` typeface sometimes represents path names, directories, or file names.
italic	*Italic* words or characters represent variable values that you must supply. *Italics* are also used for information unit titles, for the first use of a glossary term, and for general emphasis in text.
<*key*>	Angle brackets (less-than and greater-than) enclose the name of a key on the keyboard. For example, **<Enter>** refers to the key on your terminal or workstation that is labeled with the word *Enter*.

\	In command examples, a backslash indicates that the command or coding example continues on the next line. For example:
	`mkcondition -r IBM.FileSystem -e "PercentTotUsed > 90" \`
	`-E "PercentTotUsed < 85" -m p "FileSystem space used"`

{*item*}	Braces enclose a list from which you must choose an item in format and syntax descriptions.
[*item*]	Brackets enclose optional items in format and syntax descriptions.
<**Ctrl-***x*>	The notation <**Ctrl-***x*> indicates a control character sequence. For example, <**Ctrl-c**> means that you hold down the control key while pressing <**c**>.
item...	Ellipses indicate that you can repeat the preceding item one or more times.
|	In *synopsis* statements, vertical lines separate a list of choices. In other words, a vertical line means *Or*. In the left margin of the document, vertical lines indicate technical changes to the information.

Note: CLI options that accept a list of option values delimit with a comma and no space between values. As an example, to display the state on three nodes use **mmgetstate -N** *NodeA*,*NodeB*,*NodeC*. Exceptions to this syntax are listed specifically within the command.

How to send your comments

Your feedback is important in helping us to produce accurate, high-quality information. If you have any comments about this information or any other IBM Spectrum Scale documentation, send your comments to the following e-mail address:

mhvrcfs@us.ibm.com

Include the publication title and order number, and, if applicable, the specific location of the information about which you have comments (for example, a page number or a table number).

To contact the IBM Spectrum Scale development organization, send your comments to the following e-mail address:

gpfs@us.ibm.com

Summary of changes

This topic summarizes changes to the IBM Spectrum Scale licensed program and the IBM Spectrum Scale library. Within each information unit in the library, a vertical line (|) to the left of text and illustrations indicates technical changes or additions that are made to the previous edition of the information.

Summary of changes
for IBM Spectrum Scale version 5.0.1
as updated, August 2018

This release of the IBM Spectrum Scale licensed program and the IBM Spectrum Scale library includes the following improvements. All improvements are available after an upgrade, unless otherwise specified.

AFM and AFM DR-related changes

- A secondary site can be converted into the primary site, if the primary goes down. This role can be reversed permanently. See *Role reversal* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.
- For enhanced data-in-flight security, a new configuration parameter added - **afmEnableNFSSec**. See *Configuration parameters for AFM* and *Configuration parameters for AFM-based DR* in *IBM Spectrum Scale: Administration Guide*.
- An administrator can stop replication activity of filesets from cache to home during planned downtime. See *Stop and start replication on a fileset* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Authentication-related changes

The *mmuserauth command* can use passwords saved in a Stanza file.

Big data and analytics changes

For information on changes in IBM Spectrum Scale Big Data and Analytics support, see Big data and analytics - summary of changes.

Cloud services changes

Cloud services has the following updates:

- Support for backup and restore using SOBAR. For more information, see the *Scale out backup and restore (SOBAR) for Cloud services* topic in the *IBM Spectrum Scale: Administration Guide*.
- Support for automated Cloud services maintenance service for the following operations:
  - Background removal of deleted files from the object storage
  - Backing up the Cloud services full database to the cloud
  - Reconciling the Cloud services database
- Support for setting up a customized maintenance window, overriding the default values
- Support for multi-threading within a file in a single node to improve large file recall latency.

File audit logging updates

File audit logging has the following updates:

- If a file system with file audit logging is unmounted and then remounted, the producer reacquires the password. In addition, if a node mounts a file system after file audit logging has already been enabled, the producer on that node acquires the password. Previously, in both cases, the GPFS daemon would have to be restarted for the producer to acquire the new password.
- File audit logging authentication is changed from SASL PLAINTEXT to SASL SCRAM for improved security.

- If the Object protocol is enabled, no file system activity occurring within the primary object fileset is audited.
- Added **Services** > **File Auditing** page in the management GUI to enable monitoring file auditing through GUI.

File system core improvements

File systems: Integration with systemd is broader
You can now use systemd to manage IBM Spectrum Scale systemd services on configured systems. IBM Spectrum Scale automatically installs and configures GPFS as a suite of systemd services on systems that have systemd version 219 or later installed. Support for the IBM Spectrum Scale Cluster Configuration Repository (CCR) is included. For more information, see the topic *Planning for systemd* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

File systems: Traditional NSD nodes and servers can use checksums
NSD clients and servers that are configured with IBM Spectrum Scale can use checksums to verify data integrity and detect network corruption of file data that the client reads from or writes to the NSD server. For more information, see the **nsdCksumTraditional** and **nsdDumpBuffersOnCksumError** attributes in the topic *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference*.

File systems: Concurrent updates to small shared directories are faster
Fine-grained directory locking significantly improves the performance of concurrent updates to small directories that are accessed by more than one node. "Concurrent" updates means updates by multiple nodes within a 10-second interval. A "small" directory is one with fewer than 8 KiB entries.

File systems: NSD disk discovery on Linux now detects NVMe devices
The default script for NSD disk discovery on Linux, `/usr/lpp/mmfs/bin/mmdevdiscover`, now automatically detects NVM Express (NVMe) devices. It is no longer necessary to create an `nsddevices` user exit script to detect NVMe devices on a node. For more information, see the topic *NSD disk discovery* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* and the topic *nsddevices user exit* in the *IBM Spectrum Scale: Command and Programming Reference*.

`mmapplypolicy` command: New default values are available for the parameters `-N` (helper nodes) and -g (global work directory)

- If the **-N** parameter is not specified and the **defaultHelperNodes** attribute is not set, then the list of helper nodes defaults to the **managerNodes** node class. The target file system must be at format version 5.0.1 (format number 19.01) or later.
- If the **-g** parameter is not specified, then the global work directory defaults to the path (absolute or relative) that is stored in the new **sharedTmpDir** attribute. The target file system can be at any supported format version.

For more information, see the topics *mmapplypolicy command* and *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference*.

`mmbackup` command: A new default value is available for the -g (global work directory) parameter
If the **-g** parameter is not specified, then the global work directory defaults to the path (absolute or relative) that is stored in the new **sharedTmpDir** attribute. The target file system can be at any supported format version. For more information, see the topics *mmbackup command* and *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference*.

`mmcachectl` command: You can list the file and directory entries in the local page pool
You can display the number of bytes of file data that are stored in the local page pool for each file in a set of files, along with related information. You can display information for a single file, for the files in a fileset, for all the files in a file system, or for all the file

systems that are mounted by the node. For more information, see *mmcachectl command* in the *IBM Spectrum Scale: Command and Programming Reference*.

IBM Spectrum Scale functionality to support GDPR requirements

To understand the requirements of EU General Data Protection Regulation (GDPR) compliance that are applicable to unstructured data storage and how IBM Spectrum Scale helps to address them, see the IBM Spectrum Scale functionality to support GDPR requirements technote.

IBM Spectrum Scale management API changes

Added the following API commands:

- GET /nodes/{name}/services
- GET /nodes/{name}/services/{serviceName}
- PUT /nodes/{name}/services/{serviceName}
- POST /filesystems/{filesystemName}/filesets/{filesetName}/afmctl
- GET /filesystems/{filesystemName}/policies
- PUT /filesystems/{filesystemName}/policies
- GET /perfmon/sensors
- GET /perfmon/sensors/{sensorName}
- PUT /perfmon/sensors/{sensorName}
- GET /cliauditlog

For more information on the API commands, see *IBM Spectrum Scale management API commands* in *IBM Spectrum Scale: Command and Programming Reference*. You can also access the documentation corresponding to each API command from the GUI itself. The API documentation is available in the GUI at: *https://<IP address or host name of API server>:<port>/ibm/api/explorer/*. For example: *https://scalegui.ibm.com:443/ibm/api/explorer*.

IBM Spectrum Scale GUI changes

The following changes are made to the GUI:

- Added a new Services page that provides options to monitor, configure, and manage various services that are available in the IBM Spectrum Scale system. You can monitor and manage the following services from the Services page:
  - GPFS daemon
  - GUI
  - CES
  - CES network
  - Hadoop connector
  - Performance monitoring
  - NFS
  - SMB
  - Object
  - File auditing
  - Message queue
  - File authentication
  - Object authentication
- Added a new **Access** > **Command Audit Log** page that lists the various actions that are performed through the CLI. This page helps the system administrator to audit the commands and tasks the users and administrators are performing. These logs can also be used to troubleshoot issues that are reported in the system.

- Moved the NFS Service, SMB Service, Object Service, and Object Administrator pages from the Settings menu to the newly created Services page.
- Removed GUI Preferences page and moved the options in that page to the GUI section of the Services page.
- New option is added in the GUI section of Services page to define session timeout for the GUI users.
- Support for creating and installing a self-signed or CA-certified SSL certificates is added in the GUI section of the Services page.
- Remote cluster monitoring capabilities are added. You can now create customized performance charts in the **Monitoring** > **Statistics** page and use them in the **Monitoring** > **Dashboard** page. If a file system is mounted on the remote cluster node, the performance of the remote node can be monitored through the detailed view of file systems in **Files** > **File Systems** page.
- Modified the **Files** > **Transparent Cloud Tiering** page to display details of the container pairs and cloud account.
- Added support for creating and modifying encryption rules in the **Files** > **Information Lifecycle** page. You can now create and manage the following types of encryption rules as well:
  - Encryption
  - Encryption specification
  - Encryption exclude
- Added ILM policy run settings in the **Files** > **Information Lifecycle** page.
- Added the **Provide Feedback** option in the user menu that is available at the upper right corner of the GUI.
- In the **Monitoring** > **Events** page, the events that are occurred multiple times are now grouped under the newly introduced **Event Groups** tab and the number of occurrences of the events are indicated in the **Occurrences** column. The **Individual Events** tab lists all the events irrespective of the multiple occurrences.
- Added a report for fileset growth and size distribution in the **Files** > **Filesets** page.
- Added the GPFSFilesystemAPI-sensor based performance monitoring in the File Systems and Nodes pages.
- NFS performance monitoring metrics are added in the detailed view of the **Files** > **Active File Management** page.

Installation toolkit changes

- The installation toolkit supports the installation and the deployment of IBM Spectrum Scale on Red Hat Enterprise Linux 7.5 on x86_64, PPC64, and PPC64LE.
- The installation toolkit supports the installation and the deployment of IBM Spectrum Scale on Ubuntu 16.04.4 and 18.04 (x86_64).
- The installation toolkit config populate option now supports call home and file audit logging.
- The installation toolkit performance monitoring configuration for protocols sensors has been improved.

mmhealth command: Enhancements

New options have been added to the **`mmhealth node show`** and **`mmhealth cluster show`** commands. For more information, see the topic *mmhealth command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Object changes

- In Cluster Export Services (CES), the Pike release of OpenStack is used for Swift, Keystone, and their dependent packages.

NFS changes

- NFS Daemon rename - improved compatibility with Red Hat Selinux environment.

- CES NFS logs its crash stack trace, in case of an abnormal termination.

SMB changes

The following enhancements are done:

- **winbind** queuing improvements under high load
- **winbind** enhancements to re-establish domain controller connection on reboot
- Support for hardware encryption on Power®
- Enhanced debug messages
- Enhanced share mode handling with Microsoft Excel temporary files
- IBM Spectrum Scale by default does no longer allow anonymous access via SMB. With this change, it is no longer possible to enumerate local users without credentials. If this change causes issues with customer application requiring anonymous access, the original behavior can be restored by issuing the `mmsmb config change --option "restrict anonymous = 0"` command.

System Health changes

- Improved configuration options for performance monitoring tools
- Enhanced checking of Infiniband port state: monitoring of port speed and width
- Upgrade: better shutdown/unmount orchestration to avoid upgrade problems
- Recovery: Automatic restart of nfs-ganesha
- Added performance data to call home package
- CES improvements
- Network monitoring for CES IPs when using node affinity
- Identification of unassigned CES IPs
- Improve error messages for easier problem determination
- Enhanced reporting for CES IP moves and rebalancing

Introduction of global group for call home

Call home now uses a global group as a default group that contains the global values that are applied to all groups. For more information on this group and its uses, see the *Understanding call home* topic in the *IBM Spectrum Scale: Problem Determination Guide*.

Automatic single node assignment for performance monitoring

A single node is automatically selected by the system to run certain sensors. If the selected node becomes unresponsive, the system reconfigures a healthy node to act as the singleton sensor node. For more information on assigning single node sensors, see the *Automatic assignment of single node sensors* topic in the *IBM Spectrum Scale: Problem Determination Guide*.

Upgrades to call home configuration

Upgrading call home to a higher version requires specific steps. For more information, see the *Upgrading call home configuration* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

New options to monitor node health

The CESIP option can be used to monitor a cluster manager node. Fore more information, see the *Monitoring the health of a node* topic in the *IBM Spectrum Scale: Problem Determination Guide*.

Documented commands, structures, and subroutines

The following section lists the modifications to the documented commands, structures, and subroutines:

New commands

The following command is new in this release:

- **mmcachectl**

New structures
There are no new structures.

New subroutines
There are no new subroutines.

Changed commands
The following commands were changed:

- **mmapplypolicy**
- **mmbackup**
- **mmces**
- **mmcallhome**
- **mmchconfig**
- **mmchfileset**
- **mmcloudgateway**
- **mmfsck**
- **mmhealth**
- **mmlsqos**
- **mmlssnapshot**
- **mmnfs**
- **mmsmb**
- **mmuserauth**
- **spectrumscale**

Changed structures
There are no changed structures.

Changed subroutines
There are no changed subroutines.

Deleted commands
There are no deleted commands.

Deleted structures
There are no deleted structures.

Deleted subroutines
There are no deleted subroutines.

Messages
The following are the new, changed, and deleted messages:

New messages
6027-307, 6027-2402, 6027-2403, 6027-2404, 6027-2405, 6027-2406, 6027-2407, and 6027-3932.

Changed messages
None.

Deleted messages
None.

Changes in documentation

List of documentation changes in product guides and respective Knowledge Center sections

The following is a list of documentation changes including changes in topic titles, changes in placement of topics, and deleted topics:

Table 3. List of changes in documentation

Guide	Knowledge center section	List of changes
Concepts, Planning, and Installation Guide	Installing	• Upgrading-related information removed from the *Installing AFM* topic in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*. • Use case for arping binary for Ubuntu removed from the *Software requirements* topic in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.
	Upgrading	• Upgrading-related information removed from the *Installing AFM* topic added to a new topic *Upgrading AFM and AFM DR* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*. • The following object upgrade subtopics are removed and the object upgrade information is consolidated in *Upgrading object packages* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* – *Upgrading object packages to 5.0.x from 4.2.x* – *Upgrading object packages to version 4.2.3.x from 4.2.2.x* – *Upgrading object packages to version 4.2.2.x from 4.2.1.x*
Administration Guide	Administering	• Removed the following topic from *Administering files for Transparent cloud tiering* section: – "Restoring Transparent cloud tiering service on a backup cluster" • Changed the titles of the following topics: – *Configuring encryption with SKLM v2.7 or later* in the *IBM Spectrum Scale: Administration Guide*. – *Installing Windows IDMU* in the *IBM Spectrum Scale: Administration Guide*. – *Configuring ID mappings in IDMU* in the *IBM Spectrum Scale: Administration Guide*.
Problem Determination Guide	Monitoring	The *Threshold monitoring use cases* is now a sub-topic of the *System health monitoring use cases* section in the *IBM Spectrum Scale: Problem Determination Guide*.

Chapter 1. Performance monitoring

With IBM Spectrum Scale, system administrators can monitor the performance of GPFS and the communications protocols that it uses.

Network performance monitoring

Network performance can be monitored with Remote Procedure Call (RPC) statistics.

The GPFS daemon caches statistics relating to RPCs. Most statistics are related to RPCs sent to other nodes. This includes a set of up to seven statistics cached per node and one statistic that is cached per size of the RPC message. For RPCs received from other nodes, one statistic is cached for each type of RPC message. The counters are measured in seconds and milliseconds

The statistics cached per node are the following:

Channel wait time
The amount of time the RPC must wait for access to a communication channel to the target node.

Send time TCP
The amount of time to transfer an RPC message to an Ethernet interface.

Send time verbs
The amount of time to transfer an RPC message to an InfiniBand interface.

Receive time TCP
The amount of time to transfer an RPC message from an Ethernet interface into the daemon.

Latency TCP
The latency of the RPC when sent and received over an Ethernet interface.

Latency verbs
The latency of the RPC when sent and received over an InfiniBand interface.

Latency mixed
The latency of the RPC when sent over one type of interface (Ethernet or InfiniBand) and received over the other (InfiniBand or Ethernet).

If an InfiniBand network is not configured, no statistics are cached for send time verbs, latency verbs, and latency mixed.

The latency of an RPC is defined as the round-trip time minus the execution time on the target node. The round-trip time is measured from the start of writing the RPC message to the interface until the RPC reply is completely received. The execution time is measured on the target node from the time the message is completely received until the time the reply is sent. The latency, therefore, is the amount of time the RPC is being transmitted and received over the network and is a relative measure of the network performance as seen by the GPFS daemon.

There is a statistic associated with each of a set of size ranges, each with an upper bound that is a power of 2. The first range is 0 through 64, then 65 through 128, then 129 through 256, and then continuing until the last range has an upper bound of twice the **maxBlockSize**. For example, if the **maxBlockSize** is 1 MB, the upper bound of the last range is 2,097,152 (2 MB). For each of these ranges, the associated statistic is the latency of the RPC whose size falls within that range. The size of an RPC is the amount of data sent plus the amount of data received. However, if one amount is more than 16 times greater than the other, only the larger amount is used as the size of the RPC.

The final statistic associated with each type of RPC message, on the node where the RPC is received, is the execution time of the RPC.

Each of the statistics described so far is actually an aggregation of values. By default, an aggregation consists of 60 one-second intervals, 60 one-minute intervals, 24 one-hour intervals, and 30 one-day intervals. Each interval consists of a sum of values accumulated during the interval, a count of values added into the sum, the minimum value added into the sum, and the maximum value added into the sum. Sixty seconds after the daemon starts, each of the one-second intervals contains data and every second thereafter the oldest interval is discarded and a new one entered. An analogous pattern holds for the minute, hour, and day periods.

As each RPC reply is received, the following information is saved in a *raw statistics buffer*:

- channel wait time
- send time
- receive time
- latency
- length of data sent
- length of data received
- flags indicating if the RPC was sent or received over InfiniBand
- target node identifier

As each RPC completes execution, the execution time for the RPC and the message type of the RPC is saved in a *raw execution buffer*. Once per second these raw buffers are processed and the values are added to the appropriate aggregated statistic. For each value, the value is added to the statistic's sum, the count is incremented, and the value is compared to the minimum and maximum, which are adjusted as appropriate. Upon completion of this processing, for each statistic the sum, count, minimum, and maximum values are entered into the next one-second interval.

Every 60 seconds, the sums and counts in the 60 one-second intervals are added into a one-minute sum and count. The smallest of the 60 minimum values is determined, and the largest of the 60 maximum values is determined. This one-minute sum, count, minimum, and maximum are then entered into the next one-minute interval.

An analogous pattern holds for the minute, hour, and day periods. For any one particular interval, the sum is the sum of all raw values processed during that interval, the count is the count of all values during that interval, the minimum is the minimum of all values during that interval, and the maximum is the maximum of all values during that interval.

When statistics are displayed for any particular interval, an average is calculated from the sum and count, then the average, minimum, maximum, and count are displayed. The average, minimum and maximum are displayed in units of milliseconds, to three decimal places (one microsecond granularity).

The following **mmchconfig** attributes are available to control the RPC buffers and intervals:

- **rpcPerfRawStatBufferSize**
- **rpcPerfRawExecBufferSize**
- **rpcPerfNumberSecondIntervals**
- **rpcPerfNumberMinuteIntervals**
- **rpcPerfNumberHourIntervals**
- **rpcPerfNumberDayIntervals**

The **mmdiag** command with the **--rpc** parameter can be used to query RPC statistics.

For more information, see the topics *mmchconfig command*, *mmnetverify command* and *mmdiag command* in the *IBM Spectrum Scale: Administration Guide*.

Monitoring networks using GUI

The Network page provides an easy way to monitor the performance, health status, and configuration aspects of all available networks and interfaces that are part of the networks.

A dedicated network is used within the cluster for certain operations. For example, the system uses the administration network when an administration command is issued. It is also used for sharing administration-related information. This network is used for node-to-node communication within the cluster. The daemon network is used for sharing file system or other resources data. Remote clusters also establish communication path through the daemon network. Similarly, the dedicated network types like CES network and external network can also be configured in the cluster.

The performance of network is monitored by monitoring the data transfer managed through the respective interfaces. The following types of network interfaces can be monitored through the GUI:

- IP interfaces on Ethernet and InfiniBand adapters.
- Remote Direct Memory Access (RDMA) interfaces on InfiniBand adapters with Open Fabrics Enterprise Distribution (OFED) drivers.

The Network page also exposes adapters and IPs that are not specifically bound to a service, to provide a full view of the network activity on a node.

The details of the networks and their components can be obtained both in graphical as well as tabular formats. The Network page provides the following options to analyze the performance and status of networks and adapters:

1. A quick view that gives graphical representation of overall IP throughput, overall RDMA throughput, IP interfaces by bytes sent and received, and RDMA interfaces by bytes sent and received. You can access this view by selecting the expand button that is placed next to the title of the page. You can close this view if not required.

   Graphs in the overview are refreshed regularly. The refresh intervals of the top three entities are depended on the displayed time frame as shown below:

   - Every minute for the 5-minutes time frame
   - Every 15 minutes for the 1-hour time frame
   - Every 6 hours for the 24 hours time frame
   - Every two days for the 7 days time frame
   - Every seven days for the 30 days time frame
   - Every four months for the 365 days time frame

   If you click a block in the IP interfaces charts, the corresponding details are displayed in the IP interfaces table. The table is filtered by the IP interfaces that are part of the selected block. You can remove the filter by clicking on the link that appears above the table header row.
2. A table that provides different performance metrics that are available under the following tabs of the table.

   IP Interfaces
   Shows all network interfaces that are part of the Ethernet and InfiniBand networks in the cluster. To view performance details in graphical format or to see that the events reported against the individual adapter, select the adapter in the table and then select **View Details** from the **Actions** menu.

   RDMA Interfaces
   Shows the details of the InfininBand RDMA networks that are configured in the cluster. To

view performance details in graphical format or to see that the events reported against the individual adapter, select the adapter in the table and then select **View Details** from the **Actions** menu.

The system displays the RDMA Interfaces tab only if there are RDMA interfaces available.

Networks
Shows all networks in the cluster and provides information on network types, health status, and number of nodes and adapters that are part of the network.

IP Addresses
Lists all IP addresses that are configured in the cluster.

To find networks or adapters with extreme values, you can sort the values that are displayed in the tables by different performance metrics. Click the performance metric in the table header to sort the data based on that metric. You can select the time range that determines the averaging of the values that are displayed in the table and the time range of the charts in the overview from the time range selector, which is placed in the upper right corner. The metrics in the table do not update automatically. The refresh button above the table allows to refresh the table content with more recent data.

3. A detailed view of performance aspects and events reported against each adapter. To access this view, select the adapter in the table and then select **View Details** from the **Actions** menu. The detailed view is available for both IP and RDMA interfaces.

Monitoring GPFS I/O performance with the mmpmon command

Use the **mmpmon** command to monitor GPFS performance on the node in which it is run, and other specified nodes.

Before attempting to use the **mmpmon** command, review the command documentation in the *IBM Spectrum Scale: Administration Guide*.

Next, read all of the following relevant **mmpmon** topics.

- "Overview of mmpmon"
- "Specifying input to the mmpmon command" on page 5
- "Example mmpmon scenarios and how to analyze and interpret their results" on page 34
- "Other information about mmpmon output" on page 43

Overview of mmpmon

The **mmpmon** facility allows the system administrator to collect I/O statistics from the point of view of GPFS servicing application I/O requests.

The collected data can be used for many purposes, including:

- Tracking I/O demand over longer periods of time - weeks or months.
- Recording I/O patterns over time (when peak usage occurs, and so forth).
- Determining if some nodes service more application demand than others.
- Monitoring the I/O patterns of a single application which is spread across multiple nodes.
- Recording application I/O request service times.

Figure 1 on page 5 shows the software layers in a typical system with GPFS. **mmpmon** is built into GPFS.

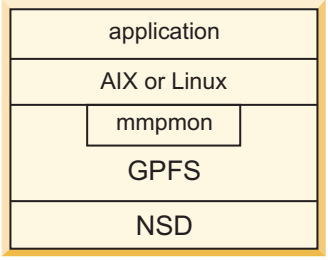

Figure 1. Node running mmpmon

Specifying input to the mmpmon command

The input requests to the **mmpmon** command allow the system administrator to collect I/O statistics per mounted file system (**fs_io_s**) or for the entire node (**io_s**).

The **mmpmon** command must be run using root authority. For command syntax, see **mmpmon** in the *IBM Spectrum Scale: Administration Guide*.

The **mmpmon** command is controlled by an input file that contains a series of requests, one per line. This input can be specified with the **-i** flag, or read from standard input (stdin). Providing input using stdin allows **mmpmon** to take keyboard input or output piped from a user script or application.

Leading blanks in the input file are ignored. A line beginning with a pound sign (#) is treated as a comment. Leading blanks in a line whose first non-blank character is a pound sign (#) are ignored.

Table 4 describes the **mmpmon** input requests.

Table 4. Input requests to the **mmpmon** *command*

Request	Description
fs_io_s	"Display I/O statistics per mounted file system" on page 6
io_s	"Display I/O statistics for the entire node" on page 8
nlist add *name*[*name*...]	"Add node names to a list of nodes for mmpmon processing" on page 10
nlist del	"Delete a node list" on page 11
nlist new *name*[*name*...]	"Create a new node list" on page 12
nlist s	"Show the contents of the current node list" on page 12
nlist sub *name*[*name*...]	"Delete node names from a list of nodes for mmpmon processing" on page 13
once *request*	Indicates that the request is to be performed only once.
reset	"Reset statistics to zero" on page 16
rhist nr	"Changing the request histogram facility request size and latency ranges" on page 19
rhist off	"Disabling the request histogram facility" on page 21. This is the default.
rhist on	"Enabling the request histogram facility" on page 22
rhist p	"Displaying the request histogram facility pattern" on page 23
rhist reset	"Resetting the request histogram facility data to zero" on page 26
rhist s	"Displaying the request histogram facility statistics values" on page 27
rpc_s	"Displaying the aggregation of execution time for Remote Procedure Calls (RPCs)" on page 30

Table 4. Input requests to the **mmpmon** *command (continued)*

Request	Description
rpc_s size	"Displaying the Remote Procedure Call (RPC) execution time according to the size of messages" on page 32
source *filename*	"Using request *source* and prefix directive *once*" on page 37
ver	"Displaying mmpmon version" on page 34
vio_s	"Displaying vdisk I/O statistics". See *IBM Spectrum Scale RAID: Administration* for more information.
vio_s_reset	"Resetting vdisk I/O statistics". See *IBM Spectrum Scale RAID: Administration* for more information.

Running mmpmon on multiple nodes

Invoke **mmpmon** list requests on a single node for mmpmon request processing on multiple nodes in a local cluster.

The **mmpmon** command may be invoked on one node to submit requests to multiple nodes in a local GPFS cluster by using the **nlist** requests. See "Understanding the node list facility" on page 9.

Running mmpmon concurrently from multiple users on the same node

Multiple instances of **mmpmon** can run on the same node so that different performance analysis applications and scripts can use the same performance data.

Five instances of **mmpmon** may be run on a given node concurrently. This is intended primarily to allow different user-written performance analysis applications or scripts to work with the performance data. For example, one analysis application might deal with **fs_io_s** and **io_s** data, while another one deals with **rhist** data, and another gathers data from other nodes in the cluster. The applications might be separately written or separately maintained, or have different sleep and wake-up schedules.

Be aware that there is only one set of counters for **fs_io_s** and **io_s** data, and another, separate set for **rhist** data. Multiple analysis applications dealing with the same set of data must coordinate any activities that could reset the counters, or in the case of **rhist** requests, disable the feature or modify the ranges.

Display I/O statistics per mounted file system

The **fs_io_s** input request to the **mmpmon** command allows the system administrator to collect I/O statistics per mounted file system.

The **fs_io_s** (file system I/O statistics) request returns strings containing I/O statistics taken over all mounted file systems as seen by that node, and are presented as total values for each file system. The values are cumulative since the file systems were mounted or since the last **reset** request, whichever is most recent. When a file system is unmounted, its statistics are lost.

Read and write statistics are recorded separately. The statistics for a given file system are for the file system activity on the node running **mmpmon**, not the file system in total (across the cluster).

Table 5 describes the keywords for the **fs_io_s** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 5. Keywords and values for the **mmpmon fs_io_s** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
_rc_	Indicates the status of the operation.

Table 5. Keywords and values for the **mmpmon fs_io_s** *response (continued)*

Keyword	Description
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.
_cl_	Name of the cluster that owns the file system.
_fs_	The name of the file system for which data are being presented.
_d_	The number of disks in the file system.
_br_	Total number of bytes read, from both disk and cache.
_bw_	Total number of bytes written, to both disk and cache.
_oc_	Count of **open()** call requests serviced by GPFS. This also includes **creat()** call counts.
_cc_	Number of **close()** call requests serviced by GPFS.
rdc	Number of application read requests serviced by GPFS.
_wc_	Number of application write requests serviced by GPFS.
dir	Number of **readdir()** call requests serviced by GPFS.
_iu_	Number of inode updates to disk.

Example of mmpmon fs_io_s request

This is an example of the **fs_io_s** input request to the **mmpmon** command and the resulting output that displays the I/O statistics per mounted file system.

Assume that **commandFile** contains this line:

```
fs_io_s
```

and this command is issued:

```
mmpmon -p  -i commandFile
```

The output is two lines in total, and similar to this:

```
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1066660148 _tu_ 407431 _cl_ myCluster.xxx.com
_fs_ gpfs2 _d_ 2 _br_ 6291456 _bw_ 314572800 _oc_ 10 _cc_ 16 _rdc_ 101 _wc_ 300 _dir_ 7 _iu_ 2
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1066660148 _tu_ 407455 _cl_ myCluster.xxx.com
_fs_ gpfs1 _d_ 3 _br_ 5431636 _bw_ 173342800 _oc_ 6 _cc_ 8 _rdc_ 54 _wc_ 156 _dir_ 3 _iu_ 6
```

The output consists of one string per mounted file system. In this example, there are two mounted file systems, **gpfs1** and **gpfs2**.

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:    myCluster.xxx.com
filesystem: gpfs2
disks: 2
timestamp: 1066660148/407431
bytes read: 6291456
bytes written: 314572800
opens: 10
closes: 16
reads: 101
writes: 300
readdir: 7
inode updates: 2

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:    myCluster.xxx.com
filesystem: gpfs1
```

```
disks: 3
timestamp: 1066660148/407455
bytes read: 5431636
bytes written: 173342800
opens: 6
closes: 8
reads: 54
writes: 156
readdir: 3
inode updates: 6
```

When no file systems are mounted, the responses are similar to:

```
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 1 _t_ 1066660148 _tu_ 407431 _cl_ - _fs_ -
```

The **_rc_** field is nonzero and the both the **_fs_** and **_cl_** fields contains a minus sign. If the **-p** flag is not specified, the results are similar to:

```
mmpmon node 199.18.1.8 name node1 fs_io_s status 1
no file systems mounted
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Display I/O statistics for the entire node

The **io_s** input request to the **mmpmon** command allows the system administrator to collect I/O statistics for the entire node.

The **io_s** (I/O statistics) request returns strings containing I/O statistics taken over all mounted file systems as seen by that node, and are presented as total values for the entire node. The values are cumulative since the file systems were mounted or since the last **reset**, whichever is most recent. When a file system is unmounted, its statistics are lost and its contribution to the total node statistics vanishes. Read and write statistics are recorded separately.

Table 6 describes the keywords for the **io_s** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 6. Keywords and values for the **mmpmon io_s** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.
_br_	Total number of bytes read, from both disk and cache.
_bw_	Total number of bytes written, to both disk and cache.
_oc_	Count of **open()** call requests serviced by GPFS. The open count also includes **creat()** call counts.
_cc_	Number of **close()** call requests serviced by GPFS.
rdc	Number of application read requests serviced by GPFS.
_wc_	Number of application write requests serviced by GPFS.
dir	Number of **readdir()** call requests serviced by GPFS.
_iu_	Number of inode updates to disk. This includes inodes flushed to disk because of access time updates.

Example of mmpmon io_s request

This is an example of the **io_s** input request to the **mmpmon** command and the resulting output that displays the I/O statistics for the entire node.

Assume that **commandFile** contains this line:

```
io_s
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is one line in total, and similar to this:

```
_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1066660148 _tu_ 407431 _br_ 6291456
_bw_ 314572800 _oc_ 10 _cc_ 16 _rdc_ 101 _wc_ 300 _dir_ 7 _iu_ 2
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 io_s OK
timestamp: 1066660148/407431
bytes read: 6291456
bytes written: 314572800
opens: 10
closes: 16
reads: 101
writes: 300
readdir: 7
inode updates: 2
```

Understanding the node list facility

The node list facility can be used to invoke **mmpmon** on multiple nodes and gather data from other nodes in the cluster. The following table describes the nlist requests for the **mmpmon** command.

Table 7. nlist requests for the **mmpmon** *command*

Request	Description
nlist add *name*[*name...*]	"Add node names to a list of nodes for mmpmon processing" on page 10
nlist del	"Delete a node list" on page 11
nlist new *name*[*name...*]	"Create a new node list" on page 12
nlist s	"Show the contents of the current node list" on page 12
nlist sub *name*[*name...*]	"Delete node names from a list of nodes for mmpmon processing" on page 13

When specifying node names, keep these points in mind:

1. A node name of '.' (dot) indicates the current node.
2. A node name of '*' (asterisk) indicates all currently connected local cluster nodes.
3. The nodes named in the node list must belong to the local cluster. Nodes in remote clusters are not supported.
4. A node list can contain nodes that are currently down. When an inactive node comes up, **mmpmon** will attempt to gather data from it.
5. If a node list contains an incorrect or unrecognized node name, all other entries in the list are processed. Suitable messages are issued for an incorrect node name.
6. When **mmpmon** gathers responses from the nodes in a node list, the full response from one node is presented before the next node. Data is not interleaved. There is no guarantee of the order of node responses.

7. The node that issues the **mmpmon** command need not appear in the node list. The case of this node serving only as a collection point for data from other nodes is a valid configuration.

Add node names to a list of nodes for mmpmon processing

The **nlist add** (node list add) request is used to add node names to a list of nodes for **mmpmon** to collect their data. The node names are separated by blanks.

Table 8 describes the keywords for the **nlist add** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 8. Keywords and values for the **mmpmon nlist add** *response*

Keyword	Description
_n_	IP address of the node processing the node list. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `add`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.
_c_	The number of nodes in the user-supplied list.
_ni_	Node name input. A user-supplied node name from the offered list of names.
_nx_	Node name translation. The preferred GPFS name for the node.
nxip	Node name translated IP address. The preferred GPFS IP address for the node.
did	The number of nodes names considered valid and processed by the requests.
nlc	The number of nodes in the node list now (after all processing).

If the **nlist add** request is issued when no node list exists, it is handled as if it were an **nlist new** request.

Example of mmpmon nlist add request:

This topic is an example of the **nlist add** request to add node names to a list of nodes for **mmpmon** processing and the output that displays.

A two- node cluster has nodes **node1** (199.18.1.2), a non-quorum node, and **node2** (199.18.1.5), a quorum node. A remote cluster has node **node3** (199.18.1.8). The **mmpmon** command is run on **node1**.

Assume that **commandFile** contains this line:

```
nlist add n2 199.18.1.2
```

and this command is issued:

```
mmpmon -p -i commandFile
```

Note in this example that an alias name **n2** was used for **node2**, and an IP address was used for **node1**. Notice how the values for **_ni_** and **_nx_** differ in these cases.

The output is similar to this:

```
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ add _rc_ 0 _t_ 1121955894 _tu_ 261881 _c_ 2
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ add _rc_ 0 _t_ 1121955894 _tu_ 261881 _ni_ n2 _nx_
node2 _nxip_ 199.18.1.5
```

```
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ add _rc_ 0 _t_ 1121955894 _tu_ 261881 _ni_
199.18.1.2 _nx_ node1 _nxip_ 199.18.1.2
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ add _rc_ 0 _t_ 1121955894 _tu_ 261881 _did_ 2 _nlc_
2
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1 nlist add
initial status 0
name count 2
timestamp 1121955879/468858
node name n2, OK (name used: node2, IP address 199.18.1.5)
node name 199.18.1.2, OK (name used: node1, IP address 199.18.1.2)
final status 0
node names processed 2
current node list count 2
```

The requests **nlist add** and **nlist sub** behave in a similar way and use the same keyword and response format.

These requests are rejected if issued while quorum has been lost.

Delete a node list

The **nlist del** (node list delete) request deletes a node list if one exists. If no node list exists, the request succeeds and no error code is produced.

Table 9 describes the keywords for the **nlist del** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 9. Keywords and values for the **mmpmon nlist del** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the _**n**_ value).
req	The action requested. In this case, the value is `del`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

Example of mmpmon nlist del request:

This topic is an example of the **nlist del** request to delete a node list and the output that displays.

Assume that **commandFile** contains this line:

```
nlist del
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ del _rc_ 0 _t_ 1121956817 _tu_ 46050
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1 nlist del status OK timestamp 1121956908/396381
```

Create a new node list

The **nlist new** (node list new) request deletes the current node list if one exists, creates a new, empty node list, and then attempts to add the specified node names to the node list. The node names are separated by blanks.

Table 10 describes the keywords for the **nlist new** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 10. Keywords and values for the **mmpmon nlist new** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `new`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

Show the contents of the current node list

The **nlist s** (node list show) request displays the current contents of the node list. If no node list exists, a count of zero is returned and no error is produced.

Table 11 describes the keywords for the **nlist s** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 11. Keywords and values for the **mmpmon nlist s** *response*

Keyword	Description
_n_	IP address of the node processing the request. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `s`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.
_c_	Number of nodes in the node list.
mbr	GPFS preferred node name for the list member.
_ip_	GPFS preferred IP address for the list member.

Example of mmpmon nlist s request:

This topic is an example of the **nlist s** request to show the contents of the current node list and the output that displays.

Assume that **commandFile** contains this line:

```
nlist s
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ s _rc_ 0 _t_ 1121956950 _tu_ 863292 _c_ 2
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ s _rc_ 0 _t_ 1121956950 _tu_ 863292 _mbr_ node1
_ip_ 199.18.1.2
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ s _rc_ 0 _t_ 1121956950 _tu_ 863292 _mbr_
node2 _ip_ 199.18.1.5
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1 nlist s
status 0
name count 2
timestamp 1121957505/165931
node name node1, IP address 199.18.1.2
node name node2, IP address 199.18.1.5
```

If there is no node list, the response looks like:

```
_nlist_ _n_ 199.18.1.2 _nn_ node1 _req_ s _rc_ 0 _t_ 1121957395 _tu_ 910440 _c_ 0
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1 nlist s
status 0
name count 0
timestamp 1121957436/353352
the node list is empty
```

The **nlist s** request is rejected if issued while quorum has been lost. Only one response line is presented.

```
_failed_ _n_ 199.18.1.8 _nn_ node2 _rc_ 668 _t_ 1121957395 _tu_ 910440
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node2: failure status 668 timestamp 1121957395/910440
lost quorum
```

Delete node names from a list of nodes for mmpmon processing

The **nlist sub** (subtract a node from the node list) request removes a node from a list of node names.

These keywords and responses are similar to the **nlist add** request. The **_req_** keyword (action requested) for **nlist sub** is sub.

For more information, see the topic "Add node names to a list of nodes for mmpmon processing" on page 10.

Node list examples and error handling

The **nlist** facility can be used to obtain GPFS performance data from nodes other than the one on which the **mmpmon** command is invoked. This information is useful to see the flow of GPFS I/O from one node to another, and spot potential problems.

A successful fs_io_s request propagated to two nodes:

This topic is an example of a successful **fs_io_s** request to two nodes to display the I/O statistics per mounted file system and the resulting system output.

This command is issued:

```
mmpmon -p -i command_file
```

where **command_file** has this:

```
nlist new node1 node2
fs_io_s
```

The output is similar to this:

```
_fs_io_s_ _n_ 199.18.1.2 _nn_ node1 _rc_ 0 _t_ 1121974197 _tu_ 278619 _cl_
xxx.localdomain _fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0
_dir_ 0 _iu_ 0
_fs_io_s_ _n_ 199.18.1.2 _nn_ node1 _rc_ 0 _t_ 1121974197 _tu_ 278619 _cl_
xxx.localdomain _fs_ gpfs1 _d_ 1 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0
_dir_ 0 _iu_ 0
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974167 _tu_ 116443 _cl_
cl1.xxx.com _fs_ fs3 _d_ 3 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0
_iu_ 3
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974167 _tu_ 116443 _cl_
cl1.xxx.comm _fs_ fs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0
_iu_ 0
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974167 _tu_ 116443 _cl_
xxx.localdomain _fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0
_dir_ 0 _iu_ 0
```

The responses from a propagated request are the same as they are issued on each node separately.

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1 fs_io_s OK
cluster: xxx.localdomain
filesystem: gpfs2
disks: 2
timestamp: 1121974088/463102
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0

mmpmon node 199.18.1.2 name node1 fs_io_s OK
cluster: xxx.localdomain
filesystem: gpfs1
disks: 1
timestamp: 1121974088/463102
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0

mmpmon node 199.18.1.5 name node2 fs_io_s OK
cluster: cl1.xxx.com
filesystem: fs3
disks: 3
timestamp: 1121974058/321741
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 2

mmpmon node 199.18.1.5 name node2 fs_io_s OK
cluster: cl1.xxx.com
filesystem: fs2
disks: 2
timestamp: 1121974058/321741
```

```
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0

mmpmon node 199.18.1.5 name node2 fs_io_s OK
cluster: xxx.localdomain
filesystem: gpfs2
disks: 2
timestamp: 1121974058/321741
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Failure on a node accessed by mmpmon:

This is an example of the system output for a failed request to two nodes to display the I/O statistics per mounted file system.

In this example, the same scenario described in "A successful fs_io_s request propagated to two nodes" on page 13 is run on **node2**, but with a failure on **node1** (a non-quorum node) because **node1** was shutdown:

```
_failed_ _n_ 199.18.1.5 _nn_ node2 _fn_ 199.18.1.2 _fnn_ node1 _rc_ 233
_t_ 1121974459 _tu_ 602231
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974459 _tu_ 616867 _cl_
cl1.xxx.com _fs_ fs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0
_iu_ 0
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974459 _tu_ 616867 _cl_
cl1.xxx.com _fs_ fs3 _d_ 3 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0
_iu_ 0
_fs_io_s_ _n_ 199.18.1.5 _nn_ node2 _rc_ 0 _t_ 1121974459 _tu_ 616867 _cl_
node1.localdomain _fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.5 name node2:
from node 199.18.1.2 from name node1: failure status 233 timestamp 1121974459/602231
node failed (or never started)
mmpmon node 199.18.1.5 name node2 fs_io_s OK
cluster: cl1.xxx.com
filesystem: fs2
disks: 2
timestamp: 1121974544/222514
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0

mmpmon node 199.18.1.5 name node2 fs_io_s OK
```

```
cluster: cl1.xxx.com
filesystem: fs3
disks: 3
timestamp: 1121974544/222514
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0

mmpmon node 199.18.1.5 name node2 fs_io_s OK
cluster: xxx.localdomain
filesystem: gpfs2
disks: 2
timestamp: 1121974544/222514
bytes read: 0
bytes written: 0
opens: 0
closes: 0
reads: 0
writes: 0
readdir: 0
inode updates: 0
```

Node shutdown and quorum loss: In this example, the quorum node (**node2**) is shutdown, causing quorum loss on **node1**. Running the same example on **node2**, the output is similar to:

```
_failed_ _n_ 199.18.1.2 _nn_ node1 _rc_ 668 _t_ 1121974459 _tu_ 616867
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.2 name node1: failure status 668 timestamp 1121974459/616867
lost quorum
```

In this scenario there can be a window where **node2** is down and **node1** has not yet lost quorum. When quorum loss occurs, the **mmpmon** command does not attempt to communicate with any nodes in the node list. The goal with failure handling is to accurately maintain the node list across node failures, so that when nodes come back up they again contribute to the aggregated responses.

Node list failure values:
Table 12 describes the keywords and values produced by the **mmpmon** command on a node list failure:

Table 12. Keywords and values for the **mmpmon nlist** *failures*

Keyword	Description
_n_	IP address of the node processing the node list. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
_fn_	IP address of the node that is no longer responding to **mmpmon** requests.
fnn	The name by which GPFS knows the node that is no longer responding to **mmpmon** requests
_rc_	Indicates the status of the operation. See "Return codes from mmpmon" on page 44.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

Reset statistics to zero

The **reset** request resets the statistics that are displayed with **fs_io_s** and **io_s** requests. The **reset** request *does not* reset the histogram data, which is controlled and displayed with **rhist** requests.

Table 13 describes the keywords for the **reset** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag. The response is a single string.

Table 13. Keywords and values for the **mmpmon reset** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

Example of mmpmon reset request

This topic is an example of how to reset file system I/O and I/O statistics to zero.

Assume that **commandFile** contains this line:

```
reset
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_reset_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1066660148 _tu_ 407431
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 reset OK
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Understanding the request histogram facility

Use the **mmpmon rhist** requests to control the request histogram facility.

The request histogram facility tallies I/O operations using a set of counters. Counters for reads and writes are kept separately. They are categorized according to a pattern that may be customized by the user. A default pattern is also provided. The **size range** and **latency range** input parameters to the **rhist nr** request are used to define the pattern.

The first time that you run the **rhist** requests, assess if there is a noticeable performance degradation. Collecting histogram data may cause performance degradation. This is possible once the histogram facility is enabled, but will probably not be noticed while the commands themselves are running. It is more of a long term issue as the GPFS daemon runs with histograms enabled.

The histogram lock is used to prevent two **rhist** requests from being processed simultaneously. If an **rhist** request fails with an **_rc_** of 16, the lock is in use. Reissue the request.

The histogram data survives file system mounts and unmounts. In order to reset this data, use the **rhist reset** request.

Table 14 on page 18 describes the **rhist** requests:

Table 14. rhist requests for the **mmpmon** *command*

Request	Description
rhist nr	"Changing the request histogram facility request size and latency ranges" on page 19
rhist off	"Disabling the request histogram facility" on page 21. This is the default.
rhist on	"Enabling the request histogram facility" on page 22
rhist p	"Displaying the request histogram facility pattern" on page 23
rhist reset	"Resetting the request histogram facility data to zero" on page 26
rhist s	"Displaying the request histogram facility statistics values" on page 27

Specifying the size ranges for I/O histograms

The I/O histogram size ranges are used to categorize the I/O according to the size, in bytes, of the I/O operation.

The size ranges are specified using a string of positive integers separated by semicolons (;). No white space is allowed within the size range operand. Each number represents the upper bound, in bytes, of the I/O request size for that range. The numbers must be monotonically increasing. Each number may be optionally followed by the letters K or k to denote multiplication by 1024, or by the letters M or m to denote multiplication by 1048576 (1024*1024).

For example, the size range operand:

```
512;1m;4m
```

represents these four size ranges

```
0        to     512 bytes
513      to 1048576 bytes
1048577  to 4194304 bytes
4194305 and greater bytes
```

In this example, a read of size 3 MB would fall in the third size range, a write of size 20 MB would fall in the fourth size range.

A size range operand of = (equal sign) indicates that the current size range is not to be changed. A size range operand of * (asterisk) indicates that the current size range is to be changed to the default size range. A maximum of 15 numbers may be specified, which produces 16 total size ranges.

The default request size ranges are:

```
0        to      255 bytes
256      to      511 bytes
512      to     1023 bytes
1024     to     2047 bytes
2048     to     4095 bytes
4096     to     8191 bytes
8192     to    16383 bytes
16384    to    32767 bytes
32768    to    65535 bytes
65536    to   131071 bytes
131072   to   262143 bytes
262144   to   524287 bytes
524288   to  1048575 bytes
1048576  to  2097151 bytes
2097152  to  4194303 bytes
4194304  and greater bytes
```

The last size range collects all request sizes greater than or equal to 4 MB. The request size ranges can be changed by using the **rhist nr** request.

For more information, see “Processing of rhist nr” on page 20.

Specifying the latency ranges for I/O

The I/O histogram latency ranges are used to categorize the I/O according to the latency time, in milliseconds, of the I/O operation.

A full set of latency ranges are produced for each size range. The latency ranges are the same for each size range.

The latency ranges are changed using a string of positive decimal numbers separated by semicolons (;). No white space is allowed within the latency range operand. Each number represents the upper bound of the I/O latency time (in milliseconds) for that range. The numbers must be monotonically increasing. If decimal places are present, they are truncated to tenths.

For example, the latency range operand:

```
1.3;4.59;10
```

represents these four latency ranges:

```
 0.0   to   1.3   milliseconds
 1.4   to   4.5   milliseconds
 4.6   to  10.0   milliseconds
10.1  and greater milliseconds
```

In this example, a read that completes in 0.85 milliseconds falls into the first latency range. A write that completes in 4.56 milliseconds falls into the second latency range, due to the truncation.

A latency range operand of = (equal sign) indicates that the current latency range is not to be changed. A latency range operand of * (asterisk) indicates that the current latency range is to be changed to the default latency range. If the latency range operand is missing, * (asterisk) is assumed. A maximum of 15 numbers may be specified, which produces 16 total latency ranges.

The latency times are in milliseconds. The default latency ranges are:

```
0.0     to    1.0   milliseconds
1.1     to   10.0   milliseconds
10.1    to   30.0   milliseconds
30.1    to  100.0   milliseconds
100.1   to  200.0   milliseconds
200.1   to  400.0   milliseconds
400.1   to  800.0   milliseconds
800.1   to 1000.0   milliseconds
1000.1 and greater  milliseconds
```

The last latency range collects all latencies greater than or equal to 1000.1 milliseconds. The latency ranges can be changed by using the **rhist nr** request.

For more information, see “Processing of rhist nr” on page 20.

Changing the request histogram facility request size and latency ranges

The **rhist nr** (new range) request allows the user to change the size and latency ranges used in the request histogram facility.

The use of **rhist nr** implies an **rhist reset**. Counters for read and write operations are recorded separately. If there are no mounted file systems at the time **rhist nr** is issued, the request still runs. The size range operand appears first, followed by a blank, and then the latency range operand.

Table 15 on page 20 describes the keywords for the **rhist nr** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 15. Keywords and values for the **mmpmon rhist nr** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `nr`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

Processing of rhist nr:

The **rhist nr** request changes the request histogram facility request size and latency ranges.

Processing of **rhist nr** is as follows:

1. The size range and latency range operands are parsed and checked for validity. If they are not valid, an error is returned and processing terminates.
2. The histogram facility is disabled.
3. The new ranges are created, by defining the following histogram counters:
   a. Two sets, one for read and one for write.
   b. Within each set, one category for each size range.
   c. Within each size range category, one counter for each latency range.

      For example, if the user specifies 11 numbers for the size range operand and 2 numbers for the latency range operand, this produces 12 size ranges, each having 3 latency ranges, because there is one additional range for the top endpoint. The total number of counters is 72: 36 read counters and 36 write counters.
4. The new ranges are made current.
5. The old ranges are discarded. Any accumulated histogram data is lost.

The histogram facility must be explicitly enabled again using **rhist on** to begin collecting histogram data using the new ranges.

The **mmpmon** command does not have the ability to collect data only for read operations, or only for write operations. The **mmpmon** command does not have the ability to specify size or latency ranges that have different values for read and write operations. The **mmpmon** command does not have the ability to specify latency ranges that are unique to a given size range.

For more information, see "Specifying the size ranges for I/O histograms" on page 18 and "Specifying the latency ranges for I/O" on page 19.

Example of mmpmon rhist nr request:

This topic is an example of using **rhist nr** to change the request histogram facility request size and latency changes.

Assume that **commandFile** contains this line:

```
rhist nr 512;1m;4m 1.3;4.5;10
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_rhist_ _n_ 199.18.2.5 _nn_ node1 _req_ nr 512;1m;4m 1.3;4.5;10 _rc_ 0 _t_ 1078929833 _tu_ 765083
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist nr 512;1m;4m 1.3;4.5;10 OK
```

In this case, **mmpmon** has been instructed to keep a total of 32 counters. There are 16 for read and 16 for write. For the reads, there are four size ranges, each of which has four latency ranges. The same is true for the writes. They are as follows:

```
size range            0   to   512    bytes
  latency range       0.0   to   1.3    milliseconds
  latency range       1.4   to   4.5    milliseconds
  latency range       4.6   to  10.0    milliseconds
  latency range       10.1 and greater  milliseconds
size range          513   to  1048576 bytes
  latency range       0.0   to   1.3    milliseconds
  latency range       1.4   to   4.5    milliseconds
  latency range       4.6   to  10.0    milliseconds
  latency range       10.1 and greater  milliseconds
size range       1048577  to  4194304 bytes
  latency range       0.0   to   1.3    milliseconds
  latency range       1.4   to   4.5    milliseconds
  latency range       4.6   to  10.0    milliseconds
  latency range       10.1 and greater  milliseconds
size range       4194305 and greater bytes
  latency range       0.0   to   1.3    milliseconds
  latency range       1.4   to   4.5    milliseconds
  latency range       4.6   to  10.0    milliseconds
  latency range       10.1 and greater  milliseconds
```

In this example, a read of size 15 MB that completes in 17.8 milliseconds would fall in the last latency range listed here. When this read completes, the counter for the last latency range will be increased by one.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

An example of an unsuccessful response is:

```
_rhist_ _n_ 199.18.2.5 _nn_ node1 _req_ nr 512;1m;4m 1;4;8;2 _rc_ 22 _t_ 1078929596 _tu_ 161683
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist nr 512;1m;4m 1;4;8;2 status 22 range error
```

In this case, the last value in the latency range, 2, is out of numerical order.

Note that the request **rhist nr = =** does not make any changes. It is ignored.

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Disabling the request histogram facility

The **rhist off** request disables the request histogram facility. This is the default value.

The data objects remain persistent, and the data they contain is not disturbed. This data is not updated again until **rhist on** is issued. **rhist off** may be combined with **rhist on** as often as desired. If there are no mounted file systems at the time **rhist off** is issued, the facility is still disabled. The response is a single string.

Table 16 describes the keywords for the **rhist off** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 16. Keywords and values for the **mmpmon rhist off** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `off`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

Example of mmpmon rhist off request:

This topic is an example of the **rhist off** request to disable the histogram facility and the output that displays.

Assume that **commandFile** contains this line:

```
rhist off
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ off _rc_ 0 _t_ 1066938820 _tu_ 5755
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist off OK
```

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

```
mmpmon node 199.18.1.8 name node1 rhist off status 16
lock is busy
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Enabling the request histogram facility

The **rhist on** request enables the request histogram facility.

When **rhist on** is invoked the first time, this request creates the necessary data objects to support histogram data gathering. This request may be combined with **rhist off** (or another **rhist on**) as often as desired. If there are no mounted file systems at the time **rhist on** is issued, the facility is still enabled. The response is a single string.

Table 17 describes the keywords for the **rhist on** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 17. Keywords and values for the **mmpmon rhist on** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.

Table 17. Keywords and values for the **mmpmon rhist on** *response (continued)*

Keyword	Description
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `on`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

Example of mmpmon rhist on request:

This topic is an example of the **rhist on** request to enable the request histogram facility and the output that displays.

Assume that **commandFile** contains this line:

```
rhist on
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ on _rc_ 0 _t_ 1066936484 _tu_ 179346
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist on OK
```

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

```
mmpmon node 199.18.1.8 name node1 rhist on status 16
lock is busy
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Displaying the request histogram facility pattern

The **rhist p** request displays the request histogram facility pattern.

The **rhist p** request returns the entire enumeration of the request size and latency ranges. The facility must be enabled for a pattern to be returned. If there are no mounted file systems at the time this request is issued, the request still runs and returns data. The pattern is displayed for both read and write.

Table 18 describes the keywords for the **rhist p** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 18. Keywords and values for the **mmpmon rhist p** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `p`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).

Table 18. Keywords and values for the **mmpmon rhist p** *response (continued)*

Keyword	Description
_tu_	Microseconds part of the current time of day.
_k_	The kind, **r** or **w**, (read or write) depending on what the statistics are for.
_R_	Request size range, minimum and maximum number of bytes.
_L_	Latency range, minimum and maximum, in milliseconds.

The request size ranges are in bytes. The zero value used for the upper limit of the last size range means 'and above'. The request size ranges can be changed by using the **rhist nr** request.

The latency times are in milliseconds The zero value used for the upper limit of the last latency range means 'and above'. The latency ranges can be changed by using the **rhist nr** request.

The **rhist p** request allows an application to query for the entire latency pattern. The application can then configure itself accordingly. Since latency statistics are reported only for ranges with nonzero counts, the statistics responses may be sparse. By querying for the pattern, an application can be certain to learn the complete histogram set. The user may have changed the pattern using the **rhist nr** request. For this reason, an application should query for the pattern and analyze it before requesting statistics.

If the facility has never been enabled, the **_rc_** field will be nonzero. An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

If the facility has been previously enabled, the **rhist p** request will still display the pattern even if **rhist off** is currently in effect.

If there are no mounted file systems at the time **rhist p** is issued, the pattern is still displayed.

Example of mmpmon rhist p request:

This topic is an example of the **rhist p** request to display the request histogram facility pattern and the output that displays.

Assume that **commandFile** contains this line:

```
rhist p
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The response contains all the latency ranges inside each of the request ranges. The data are separate for read and write:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ p _rc_ 0 _t_ 1066939007 _tu_ 386241 _k_ r
... data for reads ...
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ p _rc_ 0 _t_ 1066939007 _tu_ 386241 _k_ w
... data for writes ...
_end_
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist p OK read
... data for reads ...
mmpmon node 199.188.1.8 name node1 rhist p OK write
... data for writes ...
```

Here is an example of data for reads:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ p _rc_ 0 _t_ 1066939007 _tu_ 386241 _k_ r
_R_          0            255
_L_        0.0            1.0
_L_        1.1           10.0
_L_       10.1           30.0
_L_       30.1          100.0
_L_      100.1          200.0
_L_      200.1          400.0
_L_      400.1          800.0
_L_      800.1         1000.0
_L_     1000.1              0
_R_        256            511
_L_        0.0            1.0
_L_        1.1           10.0
_L_       10.1           30.0
_L_       30.1          100.0
_L_      100.1          200.0
_L_      200.1          400.0
_L_      400.1          800.0
_L_      800.1         1000.0
_L_     1000.1              0
_R_        512           1023
_L_        0.0            1.0
_L_        1.1           10.0
_L_       10.1           30.0
_L_       30.1          100.0
_L_      100.1          200.0
_L_      200.1          400.0
_L_      400.1          800.0
_L_      800.1         1000.0
_L_     1000.1              0
...
_R_    4194304              0
_L_        0.0            1.0
_L_        1.1           10.0
_L_       10.1           30.0
_L_       30.1          100.0
_L_      100.1          200.0
_L_      200.1          400.0
_L_      400.1          800.0
_L_      800.1         1000.0
_L_     1000.1              0
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist p OK read
size range         0 to          255
  latency range    0.0 to          1.0
  latency range    1.1 to         10.0
  latency range   10.1 to         30.0
  latency range   30.1 to        100.0
  latency range  100.1 to        200.0
  latency range  200.1 to        400.0
  latency range  400.1 to        800.0
  latency range  800.1 to       1000.0
  latency range 1000.1 to            0
size range       256 to          511
  latency range    0.0 to          1.0
  latency range    1.1 to         10.0
  latency range   10.1 to         30.0
  latency range   30.1 to        100.0
  latency range  100.1 to        200.0
  latency range  200.1 to        400.0
  latency range  400.1 to        800.0
  latency range  800.1 to       1000.0
  latency range 1000.1 to            0
size range       512 to         1023
```

```
  latency range    0.0 to          1.0
  latency range    1.1 to         10.0
  latency range   10.1 to         30.0
  latency range   30.1 to        100.0
  latency range  100.1 to        200.0
  latency range  200.1 to        400.0
  latency range  400.1 to        800.0
  latency range  800.1 to       1000.0
  latency range 1000.1 to            0
...
size range   4194304 to           0
  latency range    0.0 to          1.0
  latency range    1.1 to         10.0
  latency range   10.1 to         30.0
  latency range   30.1 to        100.0
  latency range  100.1 to        200.0
  latency range  200.1 to        400.0
  latency range  400.1 to        800.0
  latency range  800.1 to       1000.0
  latency range 1000.1 to            0
```

If the facility has never been enabled, the **_rc_** field will be nonzero.

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ p _rc_ 1 _t_ 1066939007 _tu_ 386241
```

If the **-p** flag is not specified, the output is similar to this:

```
mmpmon node 199.18.1.8 name node1 rhist p status 1
not yet enabled
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Resetting the request histogram facility data to zero

The **rhist reset** request resets the histogram statistics.

Table 19 describes the keywords for the **rhist reset** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag. The response is a single string.

Table 19. Keywords and values for the **mmpmon rhist reset** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is `reset`.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.

If the facility has been previously enabled, the reset request will still reset the statistics even if **rhist off** is currently in effect. If there are no mounted file systems at the time **rhist reset** is issued, the statistics are still reset.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

Example of mmpmon rhist reset request:

This topic is an example of the **rhist reset** request to reset the histogram facility data to zero and the output that displays.

Assume that **commandFile** contains this line:

```
rhist reset
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ reset _rc_ 0 _t_ 1066939007 _tu_ 386241
```

If the **-p** flag is not specified, the output is similar to:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ reset _rc_ 0 _t_ 1066939007 _tu_ 386241
```

If the facility has never been enabled, the **_rc_** value will be nonzero:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ reset _rc_ 1 _t_ 1066939143 _tu_ 148443
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist reset status 1
not yet enabled
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Displaying the request histogram facility statistics values

The **rhist s** request returns the current values for all latency ranges which have a nonzero count.

Table 20 describes the keywords for the **rhist s** response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 20. Keywords and values for the **mmpmon rhist s** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
req	The action requested. In this case, the value is **s**.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Microseconds part of the current time of day.
_k_	The kind, **r** or **w**, (read or write) depending on what the statistics are for.
_R_	Request size range, minimum and maximum number of bytes.
_NR_	Number of requests that fell in this size range.
_L_	Latency range, minimum and maximum, in milliseconds.
_NL_	Number of requests that fell in this latency range. The sum of all **_NL_** values for a request size range equals the **_NR_** value for that size range.

If the facility has been previously enabled, the **rhist s** request will still display the statistics even if **rhist off** is currently in effect. This allows turning the histogram statistics on and off between known points and reading them later. If there are no mounted file systems at the time **rhist s** is issued, the statistics are still displayed.

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

Example of mmpmon rhist s request:

This topic is an example of the **rhist s** request to display the request histogram facility statistics values and the output that displays.

Assume that **commandFile** contains this line:

```
rhist s
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_rhist_ _n_ 199.18.2.5 _nn_ node1 _req_ s _rc_ 0 _t_ 1066939007 _tu_ 386241 _k_ r
_R_       65536      131071 _NR_       32640
_L_         0.0         1.0 _NL_       25684
_L_         1.1        10.0 _NL_        4826
_L_        10.1        30.0 _NL_        1666
_L_        30.1       100.0 _NL_         464
_R_      262144      524287 _NR_        8160
_L_         0.0         1.0 _NL_        5218
_L_         1.1        10.0 _NL_         871
_L_        10.1        30.0 _NL_        1863
_L_        30.1       100.0 _NL_         208
_R_     1048576     2097151 _NR_        2040
_L_         1.1        10.0 _NL_         558
_L_        10.1        30.0 _NL_         809
_L_        30.1       100.0 _NL_         673
_rhist_ _n_ 199.18.2.5 _nn_ node1 _req_ s _rc_ 0 _t_ 1066939007 _tu_ 386241 _k_ w
_R_      131072      262143 _NR_       12240
_L_         0.0         1.0 _NL_       10022
_L_         1.1        10.0 _NL_        1227
_L_        10.1        30.0 _NL_         783
_L_        30.1       100.0 _NL_         208
_R_      262144      524287 _NR_        6120
_L_         0.0         1.0 _NL_        4419
_L_         1.1        10.0 _NL_         791
_L_        10.1        30.0 _NL_         733
_L_        30.1       100.0 _NL_         177
_R_      524288     1048575 _NR_        3060
_L_         0.0         1.0 _NL_        1589
_L_         1.1        10.0 _NL_         581
_L_        10.1        30.0 _NL_         664
_L_        30.1       100.0 _NL_         226
_R_     2097152     4194303 _NR_         762
_L_         1.1         2.0 _NL_         203
_L_        10.1        30.0 _NL_         393
_L_        30.1       100.0 _NL_         166
_end_
```

This small example shows that the reports for read and write may not present the same number of ranges or even the same ranges. Only those ranges with nonzero counters are represented in the response. This is true for both the request size ranges and the latency ranges within each request size range.

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.2.5 name node1 rhist s OK timestamp 1066933849/93804 read
size range            65536 to    131071 count      32640
  latency range         0.0 to       1.0 count      25684
  latency range         1.1 to      10.0 count       4826
  latency range        10.1 to      30.0 count       1666
  latency range        30.1 to     100.0 count        464
size range           262144 to    524287 count       8160
  latency range         0.0 to       1.0 count       5218
  latency range         1.1 to      10.0 count        871
  latency range        10.1 to      30.0 count       1863
  latency range        30.1 to     100.0 count        208
size range          1048576 to    2097151 count      2040
  latency range         1.1 to       10.0 count       558
  latency range        10.1 to       30.0 count       809
  latency range        30.1 to      100.0 count       673
mmpmon node 199.18.2.5 name node1 rhist s OK timestamp 1066933849/93968 write
size range           131072 to     262143 count     12240
  latency range         0.0 to        1.0 count     10022
  latency range         1.1 to       10.0 count      1227
  latency range        10.1 to       30.0 count       783
  latency range        30.1 to      100.0 count       208
size range           262144 to     524287 count      6120
  latency range         0.0 to        1.0 count      4419
  latency range         1.1 to       10.0 count       791
  latency range        10.1 to       30.0 count       733
  latency range        30.1 to      100.0 count       177
size range           524288 to    1048575 count      3060
  latency range         0.0 to        1.0 count      1589
  latency range         1.1 to       10.0 count       581
  latency range        10.1 to       30.0 count       664
  latency range        30.1 to      100.0 count       226
size range          2097152 to     4194303 count      762
  latency range         1.1 to         2.0 count      203
  latency range        10.1 to        30.0 count      393
  latency range        30.1 to       100.0 count     166
```

If the facility has never been enabled, the **_rc_** value will be nonzero:

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ reset _rc_ 1 _t_ 1066939143 _tu_ 148443
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 rhist reset status 1
not yet enabled
```

An **_rc_** value of 16 indicates that the histogram operations lock is busy. Retry the request.

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Understanding the Remote Procedure Call (RPC) facility

The **mmpmon** requests that start with **rpc_s** display an aggregation of execution time taken by RPCs for a time unit, for example the last 10 seconds. The statistics displayed are the average, minimum, and maximum of RPC execution time over the last 60 seconds, 60 minutes, 24 hours, and 30 days.

Table 21 describes the **rpc_s** requests:

Table 21. rpc_s requests for the **mmpmon** *command*

Request	Description
rpc_s	"Displaying the aggregation of execution time for Remote Procedure Calls (RPCs)" on page 30

Table 21. rpc_s requests for the **mmpmon** *command (continued)*

Request	Description
rpc_s size	"Displaying the Remote Procedure Call (RPC) execution time according to the size of messages" on page 32

The information displayed with **rpc_s** is similar to what is displayed with the **mmdiag --rpc** command.

Displaying the aggregation of execution time for Remote Procedure Calls (RPCs)

The **rpc_s** request returns the aggregation of execution time for RPCs.

Table 22 describes the keywords for the **rpc_s** response, in the order that they appear in the output.

Table 22. Keywords and values for the **mmpmon rpc_s** *response*

Keyword	Description
req	Indicates the action requested. The action can be either **size**, **node**, or **message**. If no action is requested, the default is the **rpc_s** action.
_n_	Indicates the IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	Indicates the hostname that corresponds to the IP address (the **_n_** value).
_rn_	Indicates the IP address of the remote node responding. This is the address by which GPFS knows the node. The statistics displayed are the averages from **_nn_** to this **_rnn_**.
rnn	Indicates the hostname that corresponds to the remote node IP address (the **_rn_** value). The statistics displayed are the averages from **_nn_** to this **_rnn_**.
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Indicates the microseconds part of the current time of day.
rpcObj	Indicates the beginning of the statistics for **_obj_**.
obj	Indicates the RPC object being displayed.
nsecs	Indicates the number of one-second intervals maintained.
nmins	Indicates the number of one-minute intervals maintained.
nhours	Indicates the number of one-hour intervals maintained.
ndays	Indicates the number of one-day intervals maintained.
stats	Indicates the beginning of the RPC statistics.
tmu	Indicates the time unit (seconds, minutes, hours, or days).
_av_	Indicates the average value of execution time for **_cnt_** RPCs during this time unit.
min	Indicates the minimum value of execution time for **_cnt_** RPCs during this time unit.
max	Indicates the maximum value of execution time for **_cnt_** RPCs during this time unit.
cnt	Indicates the count of RPCs that occurred during this time unit.

The values allowed for **_rpcObj_** are the following:

- **AG_STAT_CHANNEL_WAIT**
- **AG_STAT_SEND_TIME_TCP**
- **AG_STAT_SEND_TIME_VERBS**
- **AG_STAT_RECEIVE_TIME_TCP**
- **AG_STAT_RPC_LATENCY_TCP**
- **AG_STAT_RPC_LATENCY_VERBS**

- **AG_STAT_RPC_LATENCY_MIXED**
- **AG_STAT_LAST**

Example of mmpmon rpc_s request:

This topic is an example of the **rpc_s** request to display the aggregation of execution time for remote procedure calls (RPCs).

Assume that the file **commandFile** contains the following line:

```
rpc_s
```

The following command is issued:

```
mmpmon -p  -i commandFile
```

The output is similar to the following example:

```
_response_ begin mmpmon rpc_s
_mmpmon::rpc_s_  _req_ node _n_ 192.168.56.168 _nn_ node3 _rn_ 192.168.56.167 _rnn_ node2 _rc_ 0 _t_ 1388417709  _tu_ 641530
_rpcObj_ _obj_ AG_STAT_CHANNEL_WAIT _nsecs_ 60 _nmins_ 60 _nhours_ 24 _ndays_ 30
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
..............
..............
..............
_rpcObj_ _obj_ AG_STAT_SEND_TIME_TCP _nsecs_ 60 _nmins_ 60 _nhours_ 24 _ndays_ 30
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
..........................
..........................
..........................
_response_ end
```

If the **-p** flag is not specified, the output is similar to the following example:

```
Object: AG_STAT_CHANNEL_WAIT
 nsecs: 60
 nmins: 60
 nhours: 24
 ndays: 30
 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0
```

```
TimeUnit: sec
AverageValue: 0.000
MinValue: 0.000
MaxValue: 0.00
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Displaying the Remote Procedure Call (RPC) execution time according to the size of messages

The **rpc_s size** request returns the cached RPC-related size statistics.

Table 23 describes the keywords for the **rpc_s size** response, in the order that they appear in the output.

Table 23. Keywords and values for the **mmpmon rpc_s size** *response*

Keyword	Description
req	Indicates the action requested. In this case, the value is **rpc_s size**.
_n_	Indicates the IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	Indicates the hostname that corresponds to the IP address (the **_n_** value).
_rc_	Indicates the status of the operation.
_t_	Indicates the current time of day in seconds (absolute seconds since Epoch (1970)).
_tu_	Indicates the microseconds part of the current time of day.
rpcSize	Indicates the beginning of the statistics for this **_size_** group.
size	Indicates the size of the messages for which statistics are collected.
nsecs	Indicates the number of one-second intervals maintained.
nmins	Indicates the number of one-minute intervals maintained.
nhours	Indicates the number of one-hour intervals maintained.
ndays	Indicates the number of one-day intervals maintained.
stats	Indicates the beginning of the RPC-size statistics.
tmu	Indicates the time unit.
_av_	Indicates the average value of execution time for **_cnt_** RPCs during this time unit.
min	Indicates the minimum value of execution time for **_cnt_** RPCs during this time unit.
max	Indicates the maximum value of execution time for **_cnt_** RPCs during this time unit.
cnt	Indicates the count of RPCs that occurred during this time unit.

Example of mmpmon rpc_s size request:

This topic is an example of the **rpc_s size** request to display the RPC execution time according to the size of messages.

Assume that the file `commandFile` contains the following line:

```
rpc_s size
```

The following command is issued:

```
mmpmon -p  -i commandFile
```

The output is similar to the following example:

```
_mmpmon::rpc_s_ _req_ size _n_ 192.168.56.167 _nn_ node2 _rc_ 0 _t_ 1388417852  _tu_ 572950
_rpcSize_ _size_ 64 _nsecs_ 60 _nmins_ 60 _nhours_ 24 _ndays_ 30
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
.....................
.....................
.....................
_rpcSize_ _size_ 256 _nsecs_ 60 _nmins_ 60 _nhours_ 24 _ndays_ 30
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ sec _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
.....................
.....................
_stats_ _tmu_ min _av_ 0.692, _min_ 0.692, _max_ 0.692, _cnt_ 1
_stats_ _tmu_ min _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ min _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_stats_ _tmu_ min _av_ 0.000, _min_ 0.000, _max_ 0.000, _cnt_ 0
_response_ end
```

If the **-p** flag is not specified, the output is similar to the following example:

```
Bucket size: 64
 nsecs: 60
 nmins: 60
 nhours: 24
 ndays: 30
 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0

 TimeUnit: sec
 AverageValue: 0.131
 MinValue: 0.131
 MaxValue: 0.131
 Countvalue: 1

 TimeUnit: sec
 AverageValue: 0.000
 MinValue: 0.000
 MaxValue: 0.000
 Countvalue: 0
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Displaying mmpmon version

The **ver** request returns a string containing version information.

Table 24 Describes the keywords for the **ver** (version) response, in the order that they appear in the output. These keywords are used only when **mmpmon** is invoked with the **-p** flag.

Table 24. Keywords and values for the **mmpmon ver** *response*

Keyword	Description
_n_	IP address of the node responding. This is the address by which GPFS knows the node.
_nn_	The hostname that corresponds to the IP address (the **_n_** value).
_v_	The version of **mmpmon**.
_lv_	The level of **mmpmon**.
_vt_	The fix level variant of **mmpmon**.

Example of mmpmon ver request

This topic is an example of the **ver** request to display the **mmpmon** version and the output that displays.

Assume that **commandFile** contains this line:

```
ver
```

and this command is issued:

```
mmpmon -p -i commandFile
```

The output is similar to this:

```
_ver_ _n_ 199.18.1.8 _nn_ node1 _v_ 3 _lv_ 3 _vt_ 0
```

If the **-p** flag is not specified, the output is similar to:

```
mmpmon node 199.18.1.8 name node1 version 3.3.0
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Example mmpmon scenarios and how to analyze and interpret their results

This topic is an illustration of how **mmpmon** is used to analyze I/O data and draw conclusions based on it.

The **fs_io_s** and **io_s** requests are used to determine a number of GPFS I/O parameters and their implication for overall performance. The **rhist** requests are used to produce histogram data about I/O sizes and latency times for I/O requests. The request *source* and prefix directive *once* allow the user of **mmpmon** to more finely tune its operation.

fs_io_s and io_s output - how to aggregate and analyze the results

The **fs_io_s** and **io_s** requests can be used to determine a number of GPFS I/O parameters and their implication for overall performance.

The output from the **fs_io_s** and **io_s** requests can be used to determine:

1. The I/O service rate of a node, from the application point of view. The **io_s** request presents this as a sum for the entire node, while **fs_io_s** presents the data per file system. A rate can be approximated

by taking the **_br_** (bytes read) or **_bw_** (bytes written) values from two successive invocations of **fs_io_s** (or **io_s_**) and dividing by the difference of the sums of the individual **_t_** and **_tu_** values (seconds and microseconds).

This must be done for a number of samples, with a reasonably small time between samples, in order to get a rate which is reasonably accurate. Since we are sampling the information at a given interval, inaccuracy can exist if the I/O load is not smooth over the sampling time.

For example, here is a set of samples taken approximately one second apart, when it was known that continuous I/O activity was occurring:

```
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862476 _tu_ 634939 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 3737124864 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 3570 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862477 _tu_ 645988 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 3869245440 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 3696 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862478 _tu_ 647477 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 4120903680 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 3936 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862479 _tu_ 649363 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 4309647360 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 4116 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862480 _tu_ 650795 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 4542431232 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 4338 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862481 _tu_ 652515 _cl_ cluster1.ibm.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 4743757824 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 4530 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862482 _tu_ 654025 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 4963958784 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 4740 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862483 _tu_ 655782 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 5177868288 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 4944 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862484 _tu_ 657523 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 5391777792 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 5148 _dir_ 0 _iu_ 5
_fs_io_s_ _n_ 199.18.1.3 _nn_ node1 _rc_ 0 _t_ 1095862485 _tu_ 665909 _cl_ cluster1.xxx.com
_fs_ gpfs1m _d_ 3 _br_ 0 _bw_ 5599395840 _oc_ 4 _cc_ 3 _rdc_ 0 _wc_ 5346 _dir_ 0 _iu_ 5
```

This simple **awk** script performs a basic rate calculation:

```
BEGIN {
  count=0;
  prior_t=0;
  prior_tu=0;
  prior_br=0;
  prior_bw=0;
}

{
  count++;

  t = $9;
  tu = $11;
  br = $19;
  bw = $21;

  if(count > 1)
  {
    delta_t = t-prior_t;
    delta_tu = tu-prior_tu;
    delta_br = br-prior_br;
    delta_bw = bw-prior_bw;
    dt = delta_t + (delta_tu / 1000000.0);
    if(dt > 0) {
      rrate = (delta_br / dt) / 1000000.0;
      wrate = (delta_bw / dt) / 1000000.0;

printf("%5.1f MB/sec read %5.1f MB/sec write\n",rrate,wrate);
    }
  }
```

```
  prior_t=t;
  prior_tu=tu;
  prior_br=br;
  prior_bw=bw;
}
```

The calculated service rates for each adjacent pair of samples is:

```
  0.0 MB/sec read    130.7 MB/sec write
  0.0 MB/sec read    251.3 MB/sec write
  0.0 MB/sec read    188.4 MB/sec write
  0.0 MB/sec read    232.5 MB/sec write
  0.0 MB/sec read    201.0 MB/sec write
  0.0 MB/sec read    219.9 MB/sec write
  0.0 MB/sec read    213.5 MB/sec write
  0.0 MB/sec read    213.5 MB/sec write
  0.0 MB/sec read    205.9 MB/sec write
```

Since these are discrete samples, there can be variations in the individual results. For example, there may be other activity on the node or interconnection fabric. I/O size, file system block size, and buffering also affect results. There can be many reasons why adjacent values differ. This must be taken into account when building analysis tools that read **mmpmon** output and interpreting results.

For example, suppose a file is read for the first time and gives results like this.

```
  0.0 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
 92.1 MB/sec read      0.0 MB/sec write
 89.0 MB/sec read      0.0 MB/sec write
 92.1 MB/sec read      0.0 MB/sec write
 90.0 MB/sec read      0.0 MB/sec write
 96.3 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
```

If most or all of the file remains in the GPFS cache, the second read may give quite different rates:

```
  0.0 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
235.5 MB/sec read      0.0 MB/sec write
287.8 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
  0.0 MB/sec read      0.0 MB/sec write
```

Considerations such as these need to be taken into account when looking at application I/O service rates calculated from sampling **mmpmon** data.

2. Usage patterns, by sampling at set times of the day (perhaps every half hour) and noticing when the largest changes in I/O volume occur. This does not necessarily give a rate (since there are too few samples) but it can be used to detect peak usage periods.
3. If some nodes service significantly more I/O volume than others over a given time span.
4. When a parallel application is split across several nodes, and is the only significant activity in the nodes, how well the I/O activity of the application is distributed.
5. The total I/O demand that applications are placing on the cluster. This is done by obtaining results from **fs_io_s** and **io_s** in aggregate for all nodes in a cluster.
6. The rate data may appear to be erratic. Consider this example:

```
 0.0 MB/sec read    0.0 MB/sec write
 6.1 MB/sec read    0.0 MB/sec write
92.1 MB/sec read    0.0 MB/sec write
89.0 MB/sec read    0.0 MB/sec write
12.6 MB/sec read    0.0 MB/sec write
 0.0 MB/sec read    0.0 MB/sec write
 0.0 MB/sec read    0.0 MB/sec write
 8.9 MB/sec read    0.0 MB/sec write
92.1 MB/sec read    0.0 MB/sec write
```

```
90.0 MB/sec read    0.0 MB/sec write
96.3 MB/sec read    0.0 MB/sec write
 4.8 MB/sec read    0.0 MB/sec write
 0.0 MB/sec read    0.0 MB/sec write
```

The low rates which appear before and after each group of higher rates can be due to the I/O requests occurring late (in the leading sampling period) and ending early (in the trailing sampling period.) This gives an apparently low rate for those sampling periods.

The zero rates in the middle of the example could be caused by reasons such as no I/O requests reaching GPFS during that time period (the application issued none, or requests were satisfied by buffered data at a layer above GPFS), the node becoming busy with other work (causing the application to be undispatched), or other reasons.

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Request histogram (rhist) output - how to aggregate and analyze the results

The **rhist** requests are used to produce histogram data about I/O sizes and latency times for I/O requests.

The output from the **rhist** requests can be used to determine:

1. The number of I/O requests in a given size range. The sizes may vary based on operating system, explicit application buffering, and other considerations. This information can be used to help determine how well an application or set of applications is buffering its I/O. For example, if there are many very small or many very large I/O transactions. A large number of overly small or overly large I/O requests may not perform as well as an equivalent number of requests whose size is tuned to the file system or operating system parameters.
2. The number of I/O requests in a size range that have a given latency time. Many factors can affect the latency time, including but not limited to: system load, interconnection fabric load, file system block size, disk block size, disk hardware characteristics, and the operating system on which the I/O request is issued.

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output" on page 43.

Using request *source* and prefix directive *once*

The request *source* and prefix directive *once* allow **mmpmon** users to more finely tune their operations.

The **source** request causes **mmpmon** to read requests from a file, and when finished return to reading requests from the input stream.

The prefix directive **once** can be placed in front of any **mmpmon** request. The **once** prefix indicates that the request be run only once, irrespective of the setting of the **-r** flag on the **mmpmon** command. It is useful for requests that do not need to be issued more than once, such as to set up the node list or turn on the request histogram facility.

These rules apply when using the **once** prefix directive and **source** request:

1. **once** with nothing after it is an error that terminates **mmpmon** processing.
2. A file invoked with the **source** request may contain **source** requests, causing file nesting of arbitrary depth. No check is done for loops in this situation.
3. The request **once source** *filename* causes the **once** prefix to be applied to all the **mmpmon** requests in *filename*, including any **source** requests in the file.
4. If a *filename* specified with the **source** request cannot be opened for read, an error is returned and **mmpmon** terminates.

5. If the **-r** flag on the **mmpmon** command has any value other than one, and all requests are prefixed with **once**, **mmpmon** runs all the requests once, issues a message, and then terminates.

An example of *once* and *source* usage:

This topic provides and example of the **once** and **source** requests and the output that displays.

This command is issued:

```
mmpmon -p -i command.file -r 0 -d 5000 | tee output.file
```

File **command.file** consists of this:

```
once source mmpmon.header
once rhist nr 512;1024;2048;4096 =
once rhist on
source mmpmon.commands
```

File **mmpmon.header** consists of this:

```
ver
reset
```

File **mmpmon.commands** consists of this:

```
fs_io_s
rhist s
```

The **output.file** is similar to this:

```
_ver_ _n_ 199.18.1.8 _nn_ node1 _v_ 2 _lv_ 4 _vt_ 0
_reset_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770129 _tu_ 511981
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ nr 512;1024;2048;4096 = _rc_ 0 _t_ 1129770131 _tu_ 524674
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ on _rc_ 0 _t_ 1129770131 _tu_ 524921
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770131 _tu_ 525062 _cl_ node1.localdomain
_fs_ gpfs1 _d_ 1 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0 _iu_ 0
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770131 _tu_ 525062 _cl_ node1.localdomain
_fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0 _iu_ 0
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770131 _tu_ 525220 _k_ r
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770131 _tu_ 525228 _k_ w
_end_
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770136 _tu_ 526685 _cl_ node1.localdomain
_fs_ gpfs1 _d_ 1 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0 _iu_ 0
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770136 _tu_ 526685 _cl_ node1.localdomain
_fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 395018 _oc_ 504 _cc_ 252 _rdc_ 0 _wc_ 251 _dir_ 0 _iu_ 147
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770136 _tu_ 526888 _k_ r
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770136 _tu_ 526896 _k_ w
_R_ 0 512 _NR_ 169
_L_ 0.0 1.0 _NL_ 155
_L_ 1.1 10.0 _NL_ 7
_L_ 10.1 30.0 _NL_ 1
_L_ 30.1 100.0 _NL_ 4
_L_ 100.1 200.0 _NL_ 2
_R_ 513 1024 _NR_ 16
_L_ 0.0 1.0 _NL_ 15
_L_ 1.1 10.0 _NL_ 1
_R_ 1025 2048 _NR_ 3
_L_ 0.0 1.0 _NL_ 32
_R_ 2049 4096 _NR_ 18
_L_ 0.0 1.0 _NL_ 18
_R_ 4097 0 _NR_ 16
_L_ 0.0 1.0 _NL_ 16
_end_
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770141 _tu_ 528613 _cl_ node1.localdomain
_fs_ gpfs1 _d_ 1 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0 _iu_ 0
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770141 _tu_ 528613 _cl_ node1.localdomain
_fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 823282 _oc_ 952 _cc_ 476 _rdc_ 0 _wc_ 474 _dir_ 0 _iu_ 459
```

```
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770141 _tu_ 528812 _k_ r
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770141 _tu_ 528820 _k_ w
_R_ 0 512 _NR_ 255
_L_ 0.0 1.0 _NL_ 241
_L_ 1.1 10.0 _NL_ 7
_L_ 10.1 30.0 _NL_ 1
_L_ 30.1 100.0 _NL_ 4
_L_ 100.1 200.0 _NL_ 2
_R_ 513 1024 _NR_ 36
_L_ 0.0 1.0 _NL_ 35
_L_ 1.1 10.0 _NL_ 1
_R_ 1025 2048 _NR_ 90
_L_ 0.0 1.0 _NL_ 90
_R_ 2049 4096 _NR_ 55
_L_ 0.0 1.0 _NL_ 55
_R_ 4097 0 _NR_ 38
_L_ 0.0 1.0 _NL_ 37
_L_ 1.1 10.0 _NL_ 1
_end_
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770146 _tu_ 530570 _cl_ node1.localdomain
_fs_ gpfs1 _d_ 1 _br_ 0 _bw_ 0 _oc_ 0 _cc_ 0 _rdc_ 0 _wc_ 0 _dir_ 0 _iu_ 1
_fs_io_s_ _n_ 199.18.1.8 _nn_ node1 _rc_ 0 _t_ 1129770146 _tu_ 530570 _cl_ node1.localdomain
_fs_ gpfs2 _d_ 2 _br_ 0 _bw_ 3069915 _oc_ 1830 _cc_ 914 _rdc_ 0 _wc_ 901 _dir_ 0 _iu_ 1070
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770146 _tu_ 530769 _k_ r
_rhist_ _n_ 199.18.1.8 _nn_ node1 _req_ s _rc_ 0 _t_ 1129770146 _tu_ 530778 _k_ w
_R_ 0 512 _NR_ 526
_L_ 0.0 1.0 _NL_ 501
_L_ 1.1 10.0 _NL_ 14
_L_ 10.1 30.0 _NL_ 2
_L_ 30.1 100.0 _NL_ 6
_L_ 100.1 200.0 _NL_ 3
_R_ 513 1024 _NR_ 74
_L_ 0.0 1.0 _NL_ 70
_L_ 1.1 10.0 _NL_ 4
_R_ 1025 2048 _NR_ 123
_L_ 0.0 1.0 _NL_ 117
_L_ 1.1 10.0 _NL_ 6
_R_ 2049 4096 _NR_ 91
_L_ 0.0 1.0 _NL_ 84
_L_ 1.1 10.0 _NL_ 7
_R_ 4097 0 _NR_ 87
_L_ 0.0 1.0 _NL_ 81
_L_ 1.1 10.0 _NL_ 6
_end_
.............. and so forth ......................
```

If this command is issued with the same file contents:

```
mmpmon -i command.file -r 0 -d 5000 | tee output.file.english
```

The file **output.file.english** is similar to this:

```
mmpmon node 199.18.1.8 name node1 version 3.1.0
mmpmon node 199.18.1.8 name node1 reset OK
mmpmon node 199.18.1.8 name node1 rhist nr 512;1024;2048;4096 = OK
mmpmon node 199.18.1.8 name node1 rhist on OK
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs1
disks:                  1
timestamp:      1129770175/950895
bytes read:             0
bytes written:          0
opens:                  0
closes:                 0
reads:                  0
writes:                 0
readdir:                0
```

```
inode updates:           0

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs2
disks:                   2
timestamp:      1129770175/950895
bytes read:              0
bytes written:

opens:                   0
closes:                  0
reads:                   0
writes:                  0
readdir:                 0
inode updates:           0
mmpmon node 199.18.1.8 name node1 rhist s OK read timestamp 1129770175/951117
mmpmon node 199.18.1.8 name node1 rhist s OK write timestamp 1129770175/951125
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs1
disks:                   1
timestamp:      1129770180/952462
bytes read:              0
bytes written:           0
opens:                   0
closes:                  0
reads:                   0
writes:                  0
readdir:                 0
inode updates:           0

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs2
disks:                   2
timestamp:      1129770180/952462
bytes read:              0
bytes written:      491310
opens:                 659
closes:                329
reads:                   0
writes:                327
readdir:                 0
inode updates:          74
mmpmon node 199.18.1.8 name node1 rhist s OK read timestamp 1129770180/952711
mmpmon node 199.18.1.8 name node1 rhist s OK write timestamp 1129770180/952720
size range               0 to        512 count        214
  latency range        0.0 to        1.0 count        187
  latency range        1.1 to       10.0 count         15
  latency range       10.1 to       30.0 count          6
  latency range       30.1 to      100.0 count          5
  latency range      100.1 to      200.0 count          1
size range             513 to       1024 count         27
  latency range        0.0 to        1.0 count         26
  latency range      100.1 to      200.0 count          1
size range            1025 to       2048 count         32
  latency range        0.0 to        1.0 count         29
  latency range        1.1 to       10.0 count          1
  latency range       30.1 to      100.0 count          2
size range            2049 to       4096 count         31
  latency range        0.0 to        1.0 count         30
  latency range       30.1 to      100.0 count          1
size range            4097 to          0 count         23
  latency range        0.0 to        1.0 count         23
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
```

```
filesystem:     gpfs1
disks:                  1
timestamp:      1129770185/954401
bytes read:             0
bytes written:          0
opens:                  0
closes:                 0
reads:                  0
writes:                 0
readdir:                0
inode updates:          0

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs2
disks:                  2
timestamp:      1129770185/954401
bytes read:             0
bytes written:     1641935
opens:                1062
closes:                531
reads:                   0
writes:                529
readdir:                 0
inode updates:         523
mmpmon node 199.18.1.8 name node1 rhist s OK read timestamp 1129770185/954658
mmpmon node 199.18.1.8 name node1 rhist s OK write timestamp 1129770185/954667
size range              0 to        512 count        305
  latency range       0.0 to        1.0 count        270
  latency range       1.1 to       10.0 count         21
  latency range      10.1 to       30.0 count          6
  latency range      30.1 to      100.0 count          6
  latency range     100.1 to      200.0 count          2
size range            513 to       1024 count         39
  latency range       0.0 to        1.0 count         36
  latency range       1.1 to       10.0 count          1
  latency range      30.1 to      100.0 count          1
  latency range     100.1 to      200.0 count          1
size range           1025 to       2048 count         89
  latency range       0.0 to        1.0 count         84
  latency range       1.1 to       10.0 count          2
  latency range      30.1 to      100.0 count          3
size range           2049 to       4096 count         56
  latency range       0.0 to        1.0 count         54
  latency range       1.1 to       10.0 count          1
  latency range      30.1 to      100.0 count          1
size range           4097 to          0 count         40
  latency range       0.0 to        1.0 count         39
  latency range       1.1 to       10.0 count          1
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs1
disks:                  1
timestamp:      1129770190/956480
bytes read:             0
bytes written:          0
opens:                  0
closes:                 0
reads:                  0
writes:                 0
readdir:                0
inode updates:          0

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs2
disks:                  2
```

```
timestamp:      1129770190/956480
bytes read:               0
bytes written:     3357414
opens:                1940
closes:                969
reads:                   0
writes:                952
readdir:                 0
inode updates:        1101
mmpmon node 199.18.1.8 name node1 rhist s OK read timestamp 1129770190/956723
mmpmon node 199.18.1.8 name node1 rhist s OK write timestamp 1129770190/956732
size range               0 to        512 count        539
  latency range        0.0 to        1.0 count        494
  latency range        1.1 to       10.0 count         29
  latency range       10.1 to       30.0 count          6
  latency range       30.1 to      100.0 count          8
  latency range      100.1 to      200.0 count          2
size range             513 to       1024 count         85
  latency range        0.0 to        1.0 count         81
  latency range        1.1 to       10.0 count          2
  latency range       30.1 to      100.0 count          1
  latency range      100.1 to      200.0 count          1
size range            1025 to       2048 count        133
  latency range        0.0 to        1.0 count        124
  latency range        1.1 to       10.0 count          5
  latency range       10.1 to       30.0 count          1
  latency range       30.1 to      100.0 count          3
size range            2049 to       4096 count         99
  latency range        0.0 to        1.0 count         91
  latency range        1.1 to       10.0 count          6
  latency range       10.1 to       30.0 count          1
  latency range       30.1 to      100.0 count          1
size range            4097 to          0 count         95
  latency range        0.0 to        1.0 count         90
  latency range        1.1 to       10.0 count          4
  latency range       10.1 to       30.0 count          1
mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs1
disks:                  1
timestamp:      1129770195/958310
bytes read:             0
bytes written:          0
opens:                  0
closes:                 0
reads:                  0
writes:                 0
readdir:                0
inode updates:          0

mmpmon node 199.18.1.8 name node1 fs_io_s OK
cluster:        node1.localdomain
filesystem:     gpfs2
disks:                  2
timestamp:      1129770195/958310
bytes read:             0
bytes written:    3428107
opens:               2046
closes:              1023
reads:                  0
writes:               997
readdir:                0
inode updates:       1321
mmpmon node 199.18.1.8 name node1 rhist s OK read timestamp 1129770195/958568
mmpmon node 199.18.1.8 name node1 rhist s OK write timestamp 1129770195/958577
size range               0 to        512 count        555
  latency range        0.0 to        1.0 count        509
```

```
  latency range        1.1 to       10.0 count         30
  latency range       10.1 to       30.0 count          6
  latency range       30.1 to      100.0 count          8
  latency range      100.1 to      200.0 count          2
size range             513 to       1024 count         96
  latency range        0.0 to        1.0 count         92
  latency range        1.1 to       10.0 count          2
  latency range       30.1 to      100.0 count          1
  latency range      100.1 to      200.0 count          1
size range            1025 to       2048 count        143
  latency range        0.0 to        1.0 count        134
  latency range        1.1 to       10.0 count          5
  latency range       10.1 to       30.0 count          1
  latency range       30.1 to      100.0 count          3
size range            2049 to       4096 count        103
  latency range        0.0 to        1.0 count         95
  latency range        1.1 to       10.0 count          6
  latency range       10.1 to       30.0 count          1
  latency range       30.1 to      100.0 count          1
size range            4097 to          0 count        100
  latency range        0.0 to        1.0 count         95
latency range         1.1 to         10.0 count          4
latency range        10.1 to         30.0 count          1
.............. and so forth ......................
```

For information on interpreting **mmpmon** output results, see "Other information about mmpmon output."

Other information about mmpmon output

When interpreting the results from the **mmpmon** output there are several points to consider.

Consider these important points:

- On a node acting as a server of a GPFS file system to NFS clients, NFS I/O is accounted for in the statistics. However, the I/O is that which goes between GPFS and NFS. If NFS caches data, in order to achieve better performance, this activity is not recorded.
- I/O requests made at the application level may not be exactly what is reflected to GPFS. This is dependent on the operating system, and other factors. For example, an application read of 100 bytes may result in obtaining, and caching, a 1 MB block of data at a code level above GPFS (such as the libc I/O layer.) . Subsequent reads within this block result in no additional requests to GPFS.
- The counters kept by **mmpmon** are not atomic and may not be exact in cases of high parallelism or heavy system load. This design minimizes the performance impact associated with gathering statistical data.
- Reads from data cached by GPFS will be reflected in statistics and histogram data. Reads and writes to data cached in software layers above GPFS will be reflected in statistics and histogram data when those layers actually call GPFS for I/O.
- Activity from snapshots affects statistics. I/O activity necessary to maintain a snapshot is counted in the file system statistics.
- Some (generally minor) amount of activity in the root directory of a file system is reflected in the statistics of the file system manager node, and not the node which is running the activity.
- The open count also includes **creat()** call counts.

Counter sizes and counter wrapping

The **mmpmon** command may be run continuously for extended periods of time. The user must be aware that counters may wrap.

This information applies to the counters involved:

- The statistical counters used for the **io_s** and **fs_io_s** requests are maintained by GPFS at all times, even when **mmpmon** has not been invoked. It is suggested that you use the **reset** request prior to starting a sequence of **io_s** or **fs_io_s** requests.
- The bytes read and bytes written counters are unsigned 64-bit integers. They are used in the **fs_io_s** and **io_s** requests, as the **_br_** and **_bw_** fields.
- The counters associated with the **rhist** requests are updated only when the request histogram facility has been enabled.
- The counters used in the **rhist** requests are unsigned 64-bit integers.
- All other counters are unsigned 32-bit integers.

For more information, see "fs_io_s and io_s output - how to aggregate and analyze the results" on page 34 and "Request histogram (rhist) output - how to aggregate and analyze the results" on page 37.

Return codes from mmpmon

This topic provides the **mmpmon** return codes and explanations for the codes.

These are the return codes that can appear in the **_rc_** field:

0 Successful completion.

1 One of these has occurred:

1. For the **fs_io_s** request, no file systems are mounted.
2. For an **rhist** request, a request was issued that requires the request histogram facility to be enabled, but it is not. The facility is not enabled if:
   - Since the last **mmstartup** was issued, **rhist on** was never issued.
   - **rhist nr** was issued and **rhist on** was not issued afterwards.

2 For one of the **nlist** requests, the node name is not recognized.

13 For one of the **nlist** requests, the node name is a remote node, which is not allowed.

16 For one of the **rhist** requests, the histogram operations lock is busy. Retry the request.

17 For one of the **nlist** requests, the node name is already in the node list.

22 For one of the **rhist** requests, the size or latency range parameters were not in ascending order or were otherwise incorrect.

233 For one of the **nlist** requests, the specified node is not joined to the cluster.

668 For one of the **nlist** requests, quorum has been lost in the cluster.

Performance monitoring tool overview

The performance monitoring tool collects metrics from GPFS and protocols and provides performance information.

The performance monitoring system is started by default and consists of three parts: Collectors, Sensors, and Proxies.

Collector

In the previous release of IBM Spectrum Scale, the performance monitoring tool could be configured with a single collector only. From version 4.2, the performance monitoring tool can be configured with multiple collectors to increase scalability and fault-tolerance. This latter configuration is referred to as federation.

In a multi-collector federated configuration, the collectors need to be aware of each other, otherwise a collector would only return the data stored in its own measurement database. Once the collectors are aware of their peer collectors, they can collaborate with each other to collate measurement data for a given measurement query. All collectors that are part of the federation are specified in the peers configuration option in the collector's configuration file as shown in the following example:

```
peers = { host = "collector1.mydomain.com" port = "9085" },
    { host = "collector2.mydomain.com" port = "9085" }
```

The port number is the one specified by the `federationport` configuration option, typically set to 9085. You can also list the current host so that the same configuration file can be used for all the collector machines.

Once the peers have been specified, any query for measurement data might be directed to any of the collectors listed in the peers section and the collector collects and assembles a response based on all relevant data from all collectors. Hence, clients need to only contact a single collector instead of all of them in order to get all the measurements available in the system.

To distribute the measurement data reported by sensors over multiple collectors, multiple collectors might be specified when configuring the sensors.

If multiple collectors are specified, the sensors pick one to report their measurement data to. The sensors use stable hashes to pick the collector such that the sensor-collector relationship does not change too much if new collectors are added or if a collector is removed.

Additionally, sensors and collectors can be configured for high availability. In this setting, sensors report their measurement data to more than one collector such that the failure of a single collector would not lead to any data loss. For instance, if the collector redundancy is increased to two, every sensor reports to two collectors. As a side-effect of increasing the redundancy to two, the bandwidth consumed for reporting measurement data is duplicated. The collector redundancy has to be configured before the sensor configuration is stored in IBM Spectrum Scale by changing the **`colRedundancy`** option in `/opt/IBM/zimon/ZIMonSensors.cfg`.

Sensor

A sensor is a component that collects performance data from a node. Typically there are multiple sensors run on any node that is required to collect metrics. By default, the sensors are started on every node.

Sensors identify the collector from the information present in the sensor configuration. The sensor configuration is managed by IBM Spectrum Scale, and can be retrieved and changed using the **`mmperfmon`** command. A copy is stored in `/opt/IBM/zimon/ZIMonSensors.cfg`. However, this copy must not be edited by users.

Proxy

A proxy is run for each of the protocols to collect the metrics for that protocol.

By default, the NFS and SMB proxies are started automatically with those protocols. They do not need to be started or stopped. However, to retrieve metrics for SMB, NFS or Object, these protocols have to be active on the specific node.

For information on enabling Object metrics, see the "Enabling protocol metrics" on page 76 topic.

For information on enabling Transparent cloud tiering metrics, see *Integrating Transparent Cloud Tiering metrics with performance monitoring tool* in *IBM Spectrum Scale: Administration Guide*.

Configuring the performance monitoring tool

The performance monitoring tool, collector, sensors, and proxies, are a part of the IBM Spectrum Scale distribution. The tool is installed with the GPFS core packages on all nodes. The tools packages are small, approximately 400 KB for the sensors and 1200 KB for the collector.

Note: The tool is supported on Linux nodes only.

For information on the usage of ports for the performance monitoring tool, see the *Firewall recommendations for Performance Monitoring tool* in *IBM Spectrum Scale: Administration Guide*.

Configuring the sensor

Performance monitoring sensors can either be managed manually as individual files on each node or managed automatically by IBM Spectrum Scale. Starting with IBM Spectrum Scale version 5.0.0, only IBM Spectrum Scale-managed sensor configuration is actively supported.

Identifying the type of configuration in use:

If the performance monitoring infrastructure was installed previously, you might need to identify the type of configuration the system is currently using.

If the sensor configuration is managed automatically, the configuration is stored within IBM Spectrum Scale. If it is managed automatically, it can be viewed with the **mmperfmon config show** command. The set of nodes where this configuration is enabled can be identified through the **mmlscluster** command. Those nodes where performance monitoring metrics collection is enabled are marked with the **perfmon** designation as shown in the following sample:

```
prompt# mmlscluster
GPFS cluster information
========================
GPFS cluster name:         s1.zimon.zc2.ibm.com
GPFS cluster id:           13860500485217864948
GPFS UID domain:           s1.zimon.zc2.ibm.com
Remote shell command:      /usr/bin/ssh
Remote file copy command: /usr/bin/scp
Repository type:           CCR
Node Daemon node name         IP address     Admin node name        Designation
--------------------------------------------------------------------------------
1     s1.zimon.zc2.ibm.com   9.4.134.196   s1.zimon.zc2.ibm.com   quorum-perfmon
2     s2.zimon.zc2.ibm.com   9.4.134.197   s2.zimon.zc2.ibm.com   quorum-perfmon
3     s3.zimon.zc2.ibm.com   9.4.134.198   s3.zimon.zc2.ibm.com   quorum-perfmon
4     s4.zimon.zc2.ibm.com   9.4.134.199   s4.zimon.zc2.ibm.com   quorum-perfmon
5     s5.zimon.zc2.ibm.com   9.4.134.2     s5.zimon.zc2.ibm.com   quorum-perfmon
```

If **mmperfmon config show** does not show any configuration and no nodes are designated perfmon, the configuration can be managed manually.

Automated configuration:

Starting with version 4.2 of the performance monitoring tool, sensors can be configured on nodes that are part of an IBM Spectrum Scale cluster through an IBM Spectrum Scale based configuration mechanism. However, this requires the installation of IBM Spectrum Scale 4.2 or later versions on all the nodes where a sensor is running and where the sensors are to be configured. It also requires the entire cluster to be at least running IBM Spectrum Scale 4.1.1 or later version, and the execution of the `mmchconfig release=LATEST` command.

The automated configuration method allows the sensor configuration to be stored as part of the IBM Spectrum Scale configuration. Automated configuration is only available for the sensor configuration files (`/opt/IBM/zimon/ZIMonSensors.cfg`) but not for the collector configuration files (`/opt/IBM/zimon/`

ZIMonCollector.cfg). In this setup, the /opt/IBM/zimon/ZIMonSensors.cfg configuration file on each IBM Spectrum Scale node is maintained by IBM Spectrum Scale. As a result, the file must not be edited manually because whenever IBM Spectrum Scale needs to update a configuration parameter, the file is regenerated and any manual modifications are overwritten. Before using the automated configuration, an initial configuration needs to be stored within IBM Spectrum Scale. You can store this initial configuration by using the **mmperfmon config generate** command as shown:

```
prompt# mmperfmon config generate \
--collectors collector1.domain.com,collector2.domain.com,...
```

The **mmperfmon config generate** command uses a template configuration file for generating the automated configuration. The default location for that template configuration is /opt/IBM/zimon/defaults/ZIMonSensors.cfg.

The template configuration includes the initial settings for all the sensors and may be modified prior to invoking the **mmperfmon config generate** command. This file also includes a parameter called **colCandidates**. This parameter specifies the number of collectors that each sensor must report its data to. This may be of interest for high-availability setups, where each metric must be sent to two collectors in case one collector becomes unavailable.

Once the configuration file is stored within IBM Spectrum Scale, it can be activated as follows:

```
prompt# mmchnode --perfmon –N nodeclass1,nodeclass2,...
```

Note: Any previously existing configuration file is overwritten. Configuration changes result in a new version of the configuration file, which is then propagated through the IBM Spectrum Scale cluster at the file level.

To deactivate the performance monitoring tool, the same command is used but with the --noperfmon switch supplied instead. Configuration parameters can be changed with the following command where *parami* is of the form sensorname.sensorattribute:

```
prompt# mmperfmon config update param1=value1 param2=value2 ...
```

Sensors that collect per cluster metrics such as GPFSDiskCap, GPFSFilesetQuota, GPFSFileset, and GPFSPool must only run on a single node in the cluster for the following reasons:

1. They typically impose some overhead.
2. The data reported is the same, independent of the node the sensor is running on

Other sensors such, as the cluster export services sensors, must also only run on a specific set of nodes. For all these sensors, the restrict function is especially intended.

Some sensors, such as VFS, are not enabled by default even though they have associated predefined queries with the **mmperfmon query** command. To enable VFS sensors, use the **mmfsadm vfsstats enable** command on the node. To enable a sensor, set the period value to an integer greater than 0 and restart the sensors on that node by using the **systemctl restart pmsensors** command.

Removing an automated configuration

When upgrading the performance monitoring tool, it is important to note how the previous version was configured and if the configuration mechanism is to be changed. Before IBM Spectrum Scale 4.2, the system was configured using a file-based configuration where the configuration files were manually edited and propagated to the requisite nodes. If the configuration mechanism is to be changed, it is important to verify that the installed versions of both IBM Spectrum Scale and the performance monitoring tool support the new configuration method. However, if you want to use the manual configuration method, then take care of the following:

1. None of the nodes in the cluster must be designated **perfmon** nodes. If the nodes in the cluster are designated as **perfmon** nodes then run mmchnode --noperfmon –N all command.

2. Delete the centrally stored configuration information by issuing `mmperfmon config delete --all` command.
3. Starting from IBM Spectrum Scale version 5.0.0, manual configuration is no longer actively supported.

The `/opt/IBM/zimon/ZIMonSensors.cfg` file is then maintained manually. This mode is useful if sensors are to be installed on non-Spectrum Scale nodes or if you want to have a cluster with multiple levels of IBM Spectrum Scale running.

Manual configuration:

Performance monitoring tools can also be configured manually by the user.

Important: If you are using IBM Spectrum Scale 4.1.1 or later version, the performance monitoring tool gets automatically configured. This will automatically override any manual changes you try to make to the configuration. If you wish to change an automated configuration to a manual one, follow the steps given in *Removing an automated configuration* in the *Automated configuration* section in the *IBM Spectrum Scale: Administration Guide*.

When configuring the performance monitoring tool manually, the installation toolkit sets up a default set of sensors to monitor on each node. You can modify the sensors on each individual node.

The configuration file of the sensors, `ZimonSensors.cfg`, is located on each node in the `/opt/IBM/zimon` folder. The file lists all groups of sensors in it. The configuration file includes the parameter setting of the sensors, such as the reporting frequency, and controls the sensors that are active within the cluster. The file also contains the host name of the node where the collector is running that the sensor must be reporting to.

For example:

```
sensors =
{
        name = "CPU"
        period = 1
},
{       name = "Load"
        period = 1
},
{
        name = "Memory"
        period = 1
},
{
        name = "Network"
        period = 1
        filter = "eth*"
        # filters are currently ignored.
},
{
        name = "Netstat"
        period = 1
},
```

The period in the example specifies the interval size in number of seconds when a sensor group gathers data. `0` means that the sensor group is disabled and `1` runs the sensor group every second. You can specify a higher value to decrease the frequency at which the data is collected.

Whenever the configuration file is changed, you must stop and restart the pmsensor daemon by using the following commands:

1. Issue the **systemctl stop pmsensors** command to stop (deactivate) the sensor.
2. Issue the **systemctl start pmsensors** command to restart (activate) the sensor.

Some sensors such as the cluster export services sensors run on a specific set of nodes. Other sensors such as the GPFSDiskCap sensor must run on a single node in the cluster since the data reported is the same, independent of the node the sensor is running on. For these types of sensors, the restrict function is especially intended. For example, to restrict a **NFSIO** sensor to a node class and change the reporting period to once every 10 hours, you can specify NFSIO.period=36000 NFSIO.restrict=nodeclass1 as attribute value pairs in the update command.

Some sensors, such as VFS, are not enabled by default even though they have associated predefined queries with the **mmperfmon query** command. This is so because the collector might display performance issues of its own if it is required to collect more than 1000000 metrics per second. To enable VFS sensors, use the **mmfsadm vfsstats enable** command on the node. To enable a sensor, set the period value to an integer greater than 0 and restart the sensors on that node by using the **systemctl restart pmsensors** command.

Adding or removing a sensor from an existing automated configuration:

The performance monitoring system can be configured manually or through an automated process. To add a set of sensors for an automatic configuration, generate a file containing the sensors and the configuration parameters to be used.

The following example shows the file /tmp/new-pmsensors.conf that is used to add the following sensors:

- A new sensor NFSIO, which is not activated yet (period=0).
- Another sensor SMBStats, whose metrics are reported every second (period=1).

The restrict field is set to cesNodes so that these sensors only run on nodes from the cesNodes node class:

```
{
        name = "NFSIO"
        period = 0
        restrict = "cesNodes"
        type = "Generic"
},
{
        name = "SMBStats"
        period = 1
        restrict = "cesNodes"
        type = "Generic"
}
```

Ensure that the sensors are added and listed as part of the performance monitoring configuration. If any of the sensors mentioned in the file exist already, they are mentioned in the output for the command and those sensors are ignored, and the existing sensor configuration is kept. After the sensor is added to the configuration file, its configuration settings can be updated using **mmperfmon config update** command.

Run the following command to delete a sensor from the configuration:

```
prompt# mmperfmon config delete --sensors Sensor[,Sensor...]
```

Note: IBM Spectrum Scale version 4.2.2 has two new sensors: GPFSPool and GPFSFileset for the pmsensor service. If an older version of the IBM Spectrum Scale performance monitoring system is upgraded, these sensors are not automatically enabled. This is because automatically enabling the sensors might cause the collectors to consume more main memory than what was set aside for monitoring. Changing the memory footprint of the collector database might cause issues for the users if the collectors are tightly configured. For information on how to manually configure the performance monitoring system (file-managed configuration), see the *Manual configuration* section in the *IBM Spectrum Scale: Administration Guide*.

Related reference:

"List of performance metrics" on page 53
The performance monitoring tool can report the following metrics:

Automatic assignment of single node sensors:

A single node is automatically selected by the system to run the GPFSFilesetQuota, GPFSFileset, GPFSPool, and GPFSDiskCap sensors. If the selected node becomes unresponsive, the system reconfigures a healthy node to act as the singleton sensor node.

The GPFSFilesetQuota, GPFSFileset, GPFSPool, and GPFSDiskCap sensors are restricted to a single node in the cluster. The sensors can be restricted to a single node in the cluster by assigning a new restricted value, @CLUSTER_PERF_SENSOR, to the sensors' restrict field in the ZIMonSensor.cfg file. The restricted value allows the newly installed cluster to be run without being reconfigured by the user.

The sensor node is selected automatically based on following criteria:

- The node has the perfmon designation.
- The PERFMON component of the node is HEALTHY.
- The GPFS component of the node is HEALTHY.

Note: You can use the **mmhealth node show** command to find out if the PERFMON and GPFS components of a node are in the HEALTHY state.

By default, this node is selected from all nodes in the cluster. However, if you want to restrict the pool of nodes from which the sensor node is chosen, you can create a node class CLUSTER_PERF_SENSOR_CANDIDATES, using the **mmcrnodeclass** command. After the CLUSTER_PERF_SENSOR_CANDIDATES node class is created, only the nodes in this class can be selected.

The configuration file that is created using the **mmperfmon config generate –collectors CollectorNode[,CollectorNode...],** command includes @CLUSTER_PERF_SENSOR in the restrict fields of the GPFSFilesetQuota, GPFSFileset, GPFSPool, and GPFSDiskCap sensors.

The configuration file of an updated cluster is not configured with this feature automatically, and must be reconfigured by the administrator. You can use the mmperfmon config update SensorName.restrict=@CLUSTER_PERF_SENSOR command, where SensorName is the GPFSFilesetQuota, GPFSFileset, GPFSPool, or GPFSDiskCap, to update the configuration file.

CAUTION: The sensor update works only if the cluster has a minRelelaseLevel of 5.0.1-0 or higher. If you have 4.2.3.-x or 5.0.0.-x nodes in your cluster, this function will not work.

Configuring the collector

The following section describes how to configure the collector in a performance monitoring tool.

The most important configuration options are the domains and the peers configuration options. All other configuration options are best left at their defaults and are explained within the default configuration file shipped with ZIMon.

The configuration file of the collector, ZIMonCollector.cfg, is located in the /opt/IBM/zimon/ folder.

Metric Domain Configuration

The domains configuration indicates the number of metrics to be collected and how long they must be retained and in what granularity. Multiple domains might be specified. If data no longer fits into the current domain, data is spilled over into the next domain and re-sampled.

A simple configuration is:

```
domains = {
# this is the raw domain, aggregation factor for the raw domain is always 0
aggregation = 0
ram = "500m" # amount of RAM to be used
```

```
duration = "12h"
filesize = "1g" # maximum file size
files = 16 # number of files.
}

,
{
# this is the second domain that aggregates to 60 seconds
aggregation = 60
ram = "500m" # amount of RAM to be used
duration = "4w"
filesize = "500m" # maximum file size
files = 4 # number of files.
}

,
{
# this is the third domain that aggregates to 30*60 seconds == 30 minutes
aggregation = 30
ram = "500m" # amount of RAM to be used
duration = "1y"
filesize = "500m" # maximum file size
files = 4 # number of files.
}
```

The configuration file lists several data domains. At least one domain must be present and the first domain represents the raw data collection as the data is collected by sensors. The aggregation parameter for this first domain must be set to 0.

Each domain specifies the following parameters:

- The **duration** parameter indicates the time period until the collected metrics are pushed into the next (coarser-grained) domain. If this option is left out, no limit on the duration is imposed. Permitted units are seconds, hours, days, weeks, months and years { s, h, d, w, m, y }.
- The **ram** parameter indicates the amount of RAM to be allocated for the domain. Once that amount of RAM is filled up, collected metrics are pushed into the next (coarser-grained) domain. If this option is left out, no limit on the amount of RAM available is imposed.
- The **filesize** and **files** parameter indicates how much space is allocated on disk for a given domain. While storing metrics in memory, there is a persistence mechanism in place that also stores the metrics on disk in files of size **filesize**. Once the number of files is reached and a new file is to be allocated, the oldest file is removed from the disk. The persistent storage must be at least as large as the amount of main memory to be allocated for a domain because when the collector is restarted, the in-memory database is re-created from these files.

  If both the ram and the duration parameters are specified, both constraints are active at the same time. As soon as one of the constraints is hit, the collected metrics are pushed into the next (coarser-grained) domain.

The aggregation value, which is used for the second and following domains, indicates the resampling to be performed. Once data is spilled into this domain, the data is resampled to be no better than indicated by the aggregation factor. The value for the second domain is in seconds, the value for domain n (n>2) is the value of domain n-1 multiplied by the aggregation value of domain n.

CAUTION:
Changing the domain ram and duration parameters after data collection has started might lead to the loss of data that is already collected. It it therefore recommended to carefully estimate the collector size based on the monitored installation, and to set these parameters accordingly from the start.

The collector collects the metrics from the sensors. For example, in a five-node cluster where only the load values (load1, load5, load15) are reported, the collector will maintain 15 metrics (3 metrics times 5

nodes). Depending on the number of metrics that are collected, the collector requires a different amount of main memory to store the collected metrics in memory. Assuming 500000 metrics are collected, the following configurations are possible. Depending on the amount of data to be collected, 500000 metrics corresponds to about 1000 nodes.

Configuration 1 (4GB of RAM). Domain one configured at one second granularity for a period of six hours, domain 2 configured at 30 seconds granularity for the next two days, domain 3 configured at 15 minutes granularity for the next two weeks and domain 4 configured at 6-hour granularity for the next 2 months.

Configuration 2 (16GB of RAM). Domain one configured at 1 second granularity for a period of one day, domain 2 configured at 30 sec granularity for the next week, domain 3 configured at 15 minute granularity for the next two months and domain 4 configured at 6-hour granularity for the next year.

Note: The above computation only gives the memory required for the in-memory database, not including the indices necessary for the persistent storage or for the collector program itself.

The collectors can be stopped (deactivated) using the **systemctl stop pmcollector** command.

The collectors can be started (activated) using the **systemctl start pmcollector** command.

Configuring multiple collectors:

The performance monitoring tool installation can have a single collector, or can consist of multiple collectors to increase the scalability or the fault-tolerance of the performance monitoring system. This latter configuration is referred to as “federation”.

Note: For federation to work, all the collectors need to have the same version number.

In a multi-collector federated configuration, the collectors need to know about each other, else a collector would only return the data stored in its own measurement database. Once the collectors know the peer collectors, they will collaborate with each other to collect data for a given measurement query. All collectors that are part of the federation are specified in the `peers` configuration option in the collector’s configuration file as shown below:

```
peers = {
host = "collector1.mydomain.com"
port = "9085"
}, {
host = "collector2.mydomain.com"
port = "9085"
}
```

The port number is the one specified by the `federationport` configuration option, typically set to 9085. It is acceptable to list the current host as well so that the same configuration file can be used for all the collector machines.

Once the peers have been specified, a query for measurement data can be directed to any of the collectors listed in the peers section, and the collector will collect and assemble a response based on all relevant data from all collectors. Hence, clients only need to contact a single collector in order to get all the measurements available in the system.

To distribute the measurement data reported by sensors over multiple collectors, multiple collectors may be specified when automatically configuring the sensors, as shown in the following sample:

```
prompt# mmperfmon config generate \
--collectors collector1.domain.com,collector2.domain.com,...
```

If multiple collectors are specified, the sensors will pick one of the many collectors to report their measurement data to. The sensors use stable hashes to pick the collector such that the sensor-collector relationship does not change too much if new collectors are added or if a collector is removed.

Additionally, sensors and collectors can be configured for high availability. To maintain high availability each metric should be sent to two collectors in case one collector becomes unavailable. In this setting, sensors report their measurement data to more than one collector, so that the failure of a single collector would not lead to any data loss. For instance, if the collector redundancy is increased to two, every sensor will report to two collectors. As a side-effect of increasing the redundancy to two, the bandwidth consumed for reporting measurement data will be duplicated. The collector redundancy has to be configured before the sensor configuration is stored in GPFS by changing the `colRedundancy` option in `/opt/IBM/zimon/defaults/ZIMonSensors.cfg` as explained in the "Configuring the sensor" on page 46 section.

List of performance metrics

The performance monitoring tool can report the following metrics:

Linux metrics:

The following section lists all the Linux metrics::

Linux

All network and general metrics are native. There are no computed metrics in this section.

CPU

This section lists information about CPU in the system. For example, `myMachine|CPU|cpu_user`.

- **cpu_contexts**: Number of context switches across all CPU cores.
- **cpu_guest**: Percentage of total CPU spent running a guest OS. Included in cpu_user.
- **cpu_guest_nice**: Percentage of total CPU spent running as nice guest OS. Included in cpu_nice.
- **cpu_hiq**: Percentage of total CPU spent serving hardware interrupts.
- **cpu_idle**: Percentage of total CPU spent idling.
- **cpu_interrupts**: Number of interrupts serviced.
- **cpu_iowait**: Percentage of total CPU spent waiting for I/O to complete.
- **cpu_nice**: Percentage of total CPU time spent in lowest-priority user processes.
- **cpu_siq**: Percentage of total CPU spent serving software interrupts.
- **cpu_steal**: Percentage of total CPU spent waiting for other OS when running in a virtualized environment.
- **cpu_system**: Percentage of total CPU time spent in kernel mode.
- **cpu_user**: Percentage of total CPU time spent in normal priority user processes.

DiskFree

Gives details about the free disk. Each mounted directory will have a separate section. For example, `myMachine|DiskFree|myMount|df_free`.

- **df_free**: Amount of free disk space on the file system
- **df_total**: Amount of total disk space on the file system
- **df_used**: Amount of used disk space on the file system

Diskstat

Gives details about the Disk status for each of the disks. For example, `myMachine|Diskstat|myDisk|disk_active_ios`.

- **disk_active_ios**: Number of I/O operations currently in progress.
- **disk_aveq**: Weighted number of milliseconds spent doing I/Os.
- **disk_io_time**: Number of milliseconds the system spent doing I/O operation.
- **disk_read_ios**: Total number of read operations completed successfully.
- **disk_read_merged**: Number of (small) read operations that have been merged into a larger read.
- **disk_read_sect**: Number of sectors read.
- **disk_read_time**: Amount of time in milliseconds spent reading.
- **disk_write_ios**: Number of write operations completed successfully.
- **disk_write_merged**: Number of (small) write operations that have been merged into a larger write.
- **disk_write_sect**: Number of sectors written.
- **disk_write_time**: Amount of time in milliseconds spent writing.

Load

Gives details about the load statistics for a particular node. For example, `myMachine|Load|jobs`.

- **jobs**: The total number of jobs that currently exist in the system.
- **load1**: The average load (number of jobs in the run queue) over the last minute.
- **load15**: The average load (number of jobs in the run queue) over the last 15 minutes.
- **load5**: The average load (number of jobs in the run queue) over the five minutes.

Memory

Gives details about the memory statistics for a particular node. For example, `myMachine|Memory|mem_active`.

- **mem_active**: Active memory that was recently accessed.
- **mem_active_anon**: Active memory with no file association, that is, heap and stack memory.
- **mem_active_file**: Active memory that is associated with a file, for example, page cache memory.
- **mem_buffers**: Temporary storage used for raw disk blocks.
- **mem_cached**: In-memory cache for files read from disk (the page cache). Does not include mem_swapcached.
- **mem_dirty**: Memory which is waiting to get written back to the disk.
- **mem_inactive**: Inactive memory that hasn't been accessed recently.
- **mem_inactive_anon**: Inactive memory with no file association, that is, inactive heap and stack memory.
- **mem_inactive_file**: Inactive memory that is associated with a file, for example, page cache memory.
- **mem_memfree**: Total free RAM.
- **mem_memtotal**: Total usable RAM.
- **mem_mlocked**: Memory that is locked.
- **mem_swapcached**: In-memory cache for pages that are swapped back in.
- **mem_swapfree**: Amount of swap space that is currently unused.
- **mem_swaptotal**: Total amount of swap space available.
- **mem_unevictable**: Memory that cannot be paged out.

Netstat

Gives details about the network status for a particular node. For example, `myMachine|Netstat|ns_remote_bytes_r`.

- **ns_closewait**: Number of connections in state TCP_CLOSE_WAIT
- **ns_established**: Number of connections in state TCP_ESTABLISHED
- **ns_listen**: Number of connections in state TCP_LISTEN
- **ns_local_bytes_r**: Number of bytes received (local -> local)
- **ns_local_bytes_s**: Number of bytes sent (local -> local)
- **ns_localconn**: Number of local connections (local -> local)
- **ns_remote_bytes_r**: Number of bytes sent (local -> remote)
- **ns_remote_bytes_s**: Number of bytes sent (remote -> local)
- **ns_remoteconn**: Number of remote connections (local -> remote)
- **ns_timewait**: Number of connections in state TCP_TIME_WAIT

Network

Gives details about the network statistics per interface for a particular node. For example, `myMachine|Network|myInterface|netdev_bytes_r`.

- **netdev_bytes_r**: Number of bytes received.
- **netdev_bytes_s**: Number of bytes sent.
- **netdev_carrier**: Number of carrier loss events.
- **netdev_collisions**: Number of collisions.
- **netdev_compressed_r**: Number of compressed frames received.
- **netdev_compressed_s**: Number of compressed packets sent.
- **netdev_drops_r**: Number of packets dropped while receiving.
- **netdev_drops_s**: Number of packets dropped while sending.
- **netdev_errors_r**: Number of read errors.
- **netdev_errors_s**: Number of write errors.
- **netdev_fifo_r**: Number of FIFO buffer errors.
- **netdev_fifo_s**: Number of FIFO buffer errors while sending.
- **netdev_frames_r**: Number of frame errors while receiving.
- **netdev_multicast_r**: Number of multicast packets received.
- **netdev_packets_r**: Number of packets received.
- **netdev_packets_s**: Number of packets sent.

GPFS metrics:

The following section lists all the GPFS metrics:

GPFSDisk

For each NSD in the system, for example `myMachine|GPFSDisk|myCluster|myFilesystem|myNSD|gpfs_ds_bytes_read`

- **gpfs_ds_bytes_read**: Number of bytes read.
- **gpfs_ds_bytes_written**: Number of bytes written.
- **gpfs_ds_max_disk_wait_rd**: The longest time spent waiting for a disk read operation.
- **gpfs_ds_max_disk_wait_wr**: The longest time spent waiting for a disk write operation.
- **gpfs_ds_max_queue_wait_rd**: The longest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_ds_max_queue_wait_wr**: The longest time between being enqueued for a disk write operation and the completion of that operation.

- **gpfs_ds_min_disk_wait_rd**: The shortest time spent waiting for a disk read operation.
- **gpfs_ds_min_disk_wait_wr**: The shortest time spent waiting for a disk write operation.
- **gpfs_ds_min_queue_wait_rd**: The shortest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_ds_min_queue_wait_wr**: The shortest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_ds_read_ops**: Number of read operations.
- **gpfs_ds_tot_disk_wait_rd**: The total time in seconds spent waiting for disk read operations.
- **gpfs_ds_tot_disk_wait_wr**: The total time in seconds spent waiting for disk write operations.
- **gpfs_ds_tot_queue_wait_rd**: The total time spent between being enqueued for a read operation and the completion of that operation.
- **gpfs_ds_tot_queue_wait_wr**: The total time spent between being enqueued for a write operation and the completion of that operation.
- **gpfs_ds_write_ops**: Number of write operations.

GPFSFileset

For each independent fileset in the file system: *Cluster name* - **GPFSFileset** - *filesystem name* - *fileset name*. For example: `myCluster|GPFSFileset|myFilesystem|myFileset|gpfs_fset_maxInodes`.

- **gpfs_fset_maxInodes**: Maximum number of inodes for this independent fileset.
- **gpfs_fset_freeInodes**: Number of free inodes available for this independent fileset.
- **gpfs_fset_allocInodes**: Number of inodes allocated for this independent fileset.

GPFSFileSystem

For each file system, for example
`myMachine|GPFSFilesystem|myCluster|myFilesystem|gpfs_fs_bytes_read`

- **gpfs_fs_bytes_read**: Number of bytes read.
- **gpfs_fs_bytes_written**: Number of bytes written.
- **gpfs_fs_disks**: Number of disks in the file system.
- **gpfs_fs_max_disk_wait_rd**: The longest time spent waiting for a disk read operation.
- **gpfs_fs_max_disk_wait_wr**: The longest time spent waiting for a disk write operation.
- **gpfs_fs_max_queue_wait_rd**: The longest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_fs_max_queue_wait_wr**: The longest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_fs_min_disk_wait_rd**: The shortest time spent waiting for a disk read operation.
- **gpfs_fs_min_disk_wait_wr**: The shortest time spent waiting for a disk write operation.
- **gpfs_fs_min_queue_wait_rd**: The shortest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_fs_min_queue_wait_wr**: The shortest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_fs_read_ops**: Number of read operations
- **gpfs_fs_tot_disk_wait_rd**: The total time in seconds spent waiting for disk read operations.
- **gpfs_fs_tot_disk_wait_wr**: The total time in seconds spent waiting for disk write operations.
- **gpfs_fs_tot_queue_wait_rd**: The total time spent between being enqueued for a read operation and the completion of that operation.
- **gpfs_fs_tot_queue_wait_wr**: The total time spent between being enqueued for a write operation and the completion of that operation.

- **gpfs_fs_write_ops**: Number of write operations.

GPFSFileSystemAPI

These metrics gives the following information for each file system (application view). For example: `myMachine|GPFSFilesystemAPI|myCluster|myFilesystem|gpfs_fis_bytes_read`.

- **gpfs_fis_bytes_read**: Number of bytes read.
- **gpfs_fis_bytes_written**: Number of bytes written.
- **gpfs_fis_close_calls**: Number of close calls.
- **gpfs_fis_disks**: Number of disks in the file system.
- **gpfs_fis_inodes_written**: Number of inode updates to disk.
- **gpfs_fis_open_calls**: Number of open calls.
- **gpfs_fis_read_calls**: Number of read calls.
- **gpfs_fis_readdir_calls**: Number of readdir calls.
- **gpfs_fis_write_calls**: Number of write calls.

GPFSNSDDisk

These metrics gives the following information about each NSD disk on the NSD server. For example: `myMachine|GPFSNSDDisk|myNSDDisk|gpfs_nsdds_bytes_read`.

- **gpfs_nsdds_bytes_read**: Number of bytes read.
- **gpfs_nsdds_bytes_written**: Number of bytes written.
- **gpfs_nsdds_max_disk_wait_rd**: The longest time spent waiting for a disk read operation.
- **gpfs_nsdds_max_disk_wait_wr**: The longest time spent waiting for a disk write operation.
- **gpfs_nsdds_max_queue_wait_rd**: The longest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_nsdds_max_queue_wait_wr**: The longest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_nsdds_min_disk_wait_rd**: The shortest time spent waiting for a disk read operation.
- **gpfs_nsdds_min_disk_wait_wr**: The shortest time spent waiting for a disk write operation.
- **gpfs_nsdds_min_queue_wait_rd**: The shortest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_nsdds_min_queue_wait_wr**: The shortest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_nsdds_read_ops**: Number of read operations.
- **gpfs_nsdds_tot_disk_wait_rd**: The total time in seconds spent waiting for disk read operations.
- **gpfs_nsdds_tot_disk_wait_wr**: The total time in seconds spent waiting for disk write operations.
- **gpfs_nsdds_tot_queue_wait_rd**: The total time spent between being enqueued for a read operation and the completion of that operation.
- **gpfs_nsdds_tot_queue_wait_wr**: The total time spent between being enqueued for a write operation and the completion of that operation.
- **gpfs_nsdds_write_ops**: Number of write operations.

GPFSNSDFS

These metrics gives the following information for each filesystem served by a specific NSD server. For example: `myMachine|GPFSNSDFS|myFilesystem|gpfs_nsdfs_bytes_read`.

- **gpfs_nsdfs_bytes_read**: Number of NSD bytes read, aggregated to the file system.
- **gpfs_nsdfs_bytes_written**: Number of NSD bytes written, aggregated to the file system.

- **gpfs_nsdfs_read_ops**: Number of NSD read operations, aggregated to the file system.
- **gpfs_nsdfs_write_ops**: Number of NSD write operations, aggregated to the file system.

GPFSNSDPool

These metrics gives the following information for each filesystem and pool served by a specific NSD server. For example: `myMachine|GPFSNSDPool|myFilesystem|myPool|gpfs_nsdpool_bytes_read`.

- **gpfs_nsdpool_bytes_read**: Number of NSD bytes read, aggregated to the file system.
- **gpfs_nsdpool_bytes_written**: Number of NSD bytes written, aggregated to the file system.
- **gpfs_nsdpool_read_ops**: Number of NSD read operations, aggregated to the file system.
- **gpfs_nsdpool_write_ops**: Number of NSD write operations, aggregated to the file system.

GPFSNode

These metrics gives the following information for a particular node. For example: `myNode|GPFSNode|gpfs_ns_bytes_read`.

- **gpfs_ns_bytes_read**: Number of bytes read.
- **gpfs_ns_bytes_written**: Number of bytes written.
- **gpfs_ns_clusters**: Number of clusters participating
- **gpfs_ns_disks**: Number of disks in all mounted file systems
- **gpfs_ns_filesys**: Number of mounted file systems
- **gpfs_ns_max_disk_wait_rd**: The longest time spent waiting for a disk read operation.
- **gpfs_ns_max_disk_wait_wr**: The longest time spent waiting for a disk write operation.
- **gpfs_ns_max_queue_wait_rd**: The longest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_ns_max_queue_wait_wr**: The longest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_ns_min_disk_wait_rd**: The shortest time spent waiting for a disk read operation.
- **gpfs_ns_min_disk_wait_wr**: The shortest time spent waiting for a disk write operation.
- **gpfs_ns_min_queue_wait_rd**: The shortest time between being enqueued for a disk read operation and the completion of that operation.
- **gpfs_ns_min_queue_wait_wr**: The shortest time between being enqueued for a disk write operation and the completion of that operation.
- **gpfs_ns_read_ops**: Number of read operations.
- **gpfs_ns_tot_disk_wait_rd**: The total time in seconds spent waiting for disk read operations.
- **gpfs_ns_tot_disk_wait_wr**: The total time in seconds spent waiting for disk write operations.
- **gpfs_ns_tot_queue_wait_rd**: The total time spent between being enqueued for a read operation and the completion of that operation.
- **gpfs_ns_tot_queue_wait_wr**: The total time spent between being enqueued for a write operation and the completion of that operation.
- **gpfs_ns_write_ops**: Number of write operations.

GPFSNodeAPI

These metrics gives the following information for a particular node from its application point of view. For example: `myMachine|GPFSNodeAPI|gpfs_is_bytes_read`.

- **gpfs_is_bytes_read**: Number of bytes read.
- **gpfs_is_bytes_written**: Number of bytes written.
- **gpfs_is_close_calls**: Number of close calls.

- **gpfs_is_inodes_written**: Number of inode updates to disk.
- **gpfs_is_open_calls**: Number of open calls.
- **gpfs_is_readDir_calls**: Number of readdir calls.
- **gpfs_is_read_calls**: Number of read calls.
- **gpfs_is_write_calls**: Number of write calls.

GPFSPool

For each pool in each file system: *Cluster name* - **GPFSPool** - *filesystem name -pool name*.
For example: `myCluster|GPFSPool|myFilesystem|myPool|gpfs_pool_free_dataKBvalid*`.

- **gpfs_pool_total_dataKB**: Total capacity for data (in KB) in this pool.
- **gpfs_pool_free_dataKB**: Free capacity for data (in KB) in this pool.
- **gpfs_pool_total_metaKB**: Total capacity for metadata (in KB) in this pool.
- **gpfs_pool_free_metaKB**: Free capacity for metadata (in KB) in this pool.

GPFSPoolIO

These metrics give the details about each cluster, filesystem and pool in the system, from the point of view of a specific node. For example:
`myMachine|GPFSPoolIO|myCluster|myFilesystem|myPool|gpfs_pool_bytes_rd`

- **gpfs_pool_bytes_rd**: Total size of all disks for this usage type.
- **gpfs_pool_bytes_wr**: Total available disk space in full blocks for this usage type.
- **gpfs_pool_free_fragkb**: Total available space in fragments for this usage type.

GPFSVFS

Some sensors, such as VFS, are not enabled by default even though they have associated predefined queries with the **mmperfmon query** command. This is so because the collector might display performance issues of its own if it is required to collect more than 1000000 metrics per second. To enable VFS sensors, use the **mmfsadm vfsstats enable** command on the node. To enable a sensor, set the period value to an integer greater than **0** and restart the sensors on that node by using the **systemctl restart pmsensors** command.

These metrics gives the following information about the virtual file operation statistics (count and time) for each node. For example, `myMachine|GPFSVFS|gpfs_vfs_clear`

- **gpfs_vfs_accesses**: Number of accesses operations.
- **gpfs_vfs_accesses_t**: Amount of time in seconds spent in accesses operations.
- **gpfs_vfs_aioread**: Number of aioread operations.
- **gpfs_vfs_aioread_t**: Amount of time in seconds spent in aioread operations.
- **gpfs_vfs_aiowrite**: Number of aiowrite operations.
- **gpfs_vfs_aiowrite_t**: Amount of time in seconds spent in aiowrite operations.
- **gpfs_vfs_clear**: Number of clear operations.
- **gpfs_vfs_clear_t**: Amount of time in seconds spent in clear operations.
- **gpfs_vfs_close**: Number of close operations.
- **gpfs_vfs_close_t**: Amount of time in seconds spent in close operations.
- **gpfs_vfs_create**: Number of create operations.
- **gpfs_vfs_create_t**: Amount of time in seconds spent in create operations.
- **gpfs_vfs_decodeFh**: Number of decodeFh operations.
- **gpfs_vfs_decodeFh_t**: Amount of time in seconds spent in decodeFh operations.

- **gpfs_vfs_detDentry**: Number of detDentry operations.
- **gpfs_vfs_encodeFh**: Number of encodeFh operations.
- **gpfs_vfs_encodeFh_t**: Amount of time in seconds spent in encodeFh operations.
- **gpfs_vfs_flock**: Number of flock operations.
- **gpfs_vfs_flock_t**: Amount of time in seconds spent in flock operations.
- **gpfs_vfs_fsync**: Number of fsync operations.
- **gpfs_vfs_fsyncRange**: Number of fsyncRange operations.
- **gpfs_vfs_fsyncRange_t**: Amount of time in seconds spent in fsyncRange operations.
- **gpfs_vfs_fsync_t**: Amount of time in seconds spent in fsync operations.
- **gpfs_vfs_ftrunc**: Number of ftrunc operations.
- **gpfs_vfs_ftrunc_t**: Amount of time in seconds spent in ftrunc operations.
- **gpfs_vfs_getDentry_t**: Amount of time in seconds spent in getDentry operations.
- **gpfs_vfs_getParent**: Number of getParent operations.
- **gpfs_vfs_getParent_t**: Amount of time in seconds spent in getParent operations.
- **gpfs_vfs_getattr**: Number of getattr operations.
- **gpfs_vfs_getattr_t**: Amount of time in seconds spent in getattr operations.
- **gpfs_vfs_getxattr**: Number of getxattr operations.
- **gpfs_vfs_getxattr_t**: Amount of time in seconds spent in getxattr operations.
- **gpfs_vfs_link**: Number of link operations.
- **gpfs_vfs_link_t**: Amount of time in seconds spent in link operations.
- **gpfs_vfs_listxattr**: Number of listxattr operations.
- **gpfs_vfs_listxattr_t**: Amount of time in seconds spent in listxattr operations.
- **gpfs_vfs_lockctl**: Number of lockctl operations.
- **gpfs_vfs_lockctl_t**: Amount of time in seconds spent in lockctl operations.
- **gpfs_vfs_lookup**: Number of lookup operations.
- **gpfs_vfs_lookup_t**: Amount of time in seconds spent in lookup operations.
- **gpfs_vfs_mapLloff**: Number of mapLloff operations.
- **gpfs_vfs_mapLloff_t**: Amount of time in seconds spent in mapLloff operations.
- **gpfs_vfs_mkdir**: Number of mkdir operations.
- **gpfs_vfs_mkdir_t**: Amount of time in seconds spent in mkdir operations.
- **gpfs_vfs_mknod**: Number of mknod operations.
- **gpfs_vfs_mknod_t**: Amount of time in seconds spent in mknod operations.
- **gpfs_vfs_mmapread**: Number of mmapread operations.
- **gpfs_vfs_mmapread_t**: Amount of time in seconds spent in mmapread operations.
- **gpfs_vfs_mmapwrite**: Number of mmapwrite operations.
- **gpfs_vfs_mmapwrite_t**: Amount of time in seconds spent in mmapwrite operation.
- **gpfs_vfs_mount**: Number of mount operations.
- **gpfs_vfs_mount_t**: Amount of time in seconds spent in mount operations.
- **gpfs_vfs_open**: Number of open operations.
- **gpfs_vfs_open_t**: Amount of time in seconds spent in open operations.
- **gpfs_vfs_read**: Number of read operations.
- **gpfs_vfs_read_t**: Amount of time in seconds spent in read operations.
- **gpfs_vfs_readdir**: Number of readdir operations.
- **gpfs_vfs_readdir_t**: Amount of time in seconds spent in readdir operations.
- **gpfs_vfs_readlink**: Number of readlink operations.

- **gpfs_vfs_readlink_t**: Amount of time in seconds spent in readlink operations
- **gpfs_vfs_readpage**: Number of readpage operations.
- **gpfs_vfs_readpage_t**: Amount of time in seconds spent in readpage operations.
- **gpfs_vfs_remove**: Number of remove operations.
- **gpfs_vfs_remove_t**: Amount of time in seconds spent in remove operations.
- **gpfs_vfs_removexattr**: Number of removexattr operations.
- **gpfs_vfs_removexattr_t**: Amount of time in seconds spent in removexattr operations.
- **gpfs_vfs_rename**: Number of rename operations.
- **gpfs_vfs_rename_t**: Amount of time in seconds spent in rename operations.
- **gpfs_vfs_rmdir**: Number of rmdir operations.
- **gpfs_vfs_rmdir_t**: Amount of time in seconds spent in rmdir operations.
- **gpfs_vfs_setacl**: Number of setacl operations.
- **gpfs_vfs_setacl_t**: Amount of time in seconds spent in setacl operations.
- **gpfs_vfs_setattr**: Number of setattr operations.
- **gpfs_vfs_setattr_t**: Amount of time in seconds spent in setattr operations.
- **gpfs_vfs_setxattr**: Number of setxattr operations.
- **gpfs_vfs_setxattr_t**: Amount of time in seconds spent in setxattr operations.
- **gpfs_vfs_statfs**: Number of statfs operations.
- **gpfs_vfs_statfs_t**: Amount of time in seconds spent in statfs operations.
- **gpfs_vfs_symlink**: Number of symlink operations.
- **gpfs_vfs_symlink_t**: Amount of time in seconds spent in symlink operations.
- **gpfs_vfs_sync**: Number of sync operations.
- **gpfs_vfs_sync_t**: Amount of time in seconds spent in sync operations.
- **gpfs_vfs_tsfattr**: Number of tsfsattr operation.
- **gpfs_vfs_tsfattr_t**: Amount of time in seconds spent in tsfattr operations.
- **gpfs_vfs_tsfsattr**: Number of tsfattr operations.
- **gpfs_vfs_tsfsattr_t**: Amount of time in seconds spent in tsfsattr operations.
- **gpfs_vfs_unmap**: Number of unmap operations.
- **gpfs_vfs_unmap_t**: Amount of time in seconds spent in unmap operations.
- **gpfs_vfs_vget**: Number of vget operations.
- **gpfs_vfs_vget_t**: Amount of time in seconds spent in vget operations.
- **gpfs_vfs_write**: Number of write operations.
- **gpfs_vfs_write_t**: Amount of time in seconds spent in write operations.
- **gpfs_vfs_writepage**: Number of writepage operations.
- **gpfs_vfs_writepage_t**: Amount of time in seconds spent in writepage operations.

GPFSVIO64

These metrics provide details of the Virtual I/O server (VIOS) operations, where VIOS is supported.

Note: GPFSVIO64 is a replacement for GPFSVIO sensor and uses 64-bit counters.
- **gpfs_vio64_readOps**: Number of VIO read operations.
- **gpfs_vio64_shortWriteOps**: Number of VIO short write operations.
- **gpfs_vio64_medWriteOps**: Number of VIO medium write operations.
- **gpfs_vio64_promFTWOps**: Number of VIO promoted full-track write operations.
- **gpfs_vio64_FTWOps**: Number of VIO full-track write operations.

- **gpfs_vio64_flushUpWrOps**: Number of VIO flush update operations.
- **gpfs_vio64_flushPFTWOps**: Number of VIO flush promoted full-track write operations.
- **gpfs_vio64_migratedOps**: Number of VIO strip migration operations.
- **gpfs_vio64_scrubOps**: Number of VIO scrub operations.
- **gpfs_vio64_logWriteOps**: Number of VIO log write operations.
- **gpfs_vio64_RGDWriteOps**: Number of VIO recovery group disk write operations.
- **gpfs_vio64_metaWriteOps**: Number of VIO metadata write operations.

GPFSWaiters

For each independent fileset in the file system: *Node*- **GPFSWaiters** - *waiters_time_threshold (all, 0.1s, 0.2s, 0.5s, 1.0s, 30.0s, 60.0s)*.

Note: Here 'all' implies a waiting time greater than or equal to 0 seconds.

For example: `myNode|GPFSWaiters|all|gpfs_wt_count_all`.

- **gpfs_wt_count_all** : Count of all threads with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_local_io**: Count of threads waiting for local I/O with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_network_io**: Count of threads waiting for network I/O with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_thcond**: Count of threads waiting for a GPFS condition variable to be signaled with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_thmutex**: Count of threads waiting to lock a GPFS mutex with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_delay**: Count of threads waiting for delay interval expiration with waiting time greater than or equal to *waiters_time_threshold* seconds.
- **gpfs_wt_count_syscall**: Count of threads waiting for system call completion with waiting time greater than or equal to *waiters_time_threshold* seconds.

Computed Metrics

These metrics can only be used only through the **mmperfmon query** command. The following metrics are computed for GPFS:

- **gpfs_write_avg_lat (latency)**: gpfs_vfs_write_t / gpfs_vfs_write
- **gpfs_read_avg_lat (latency)**: gpfs_vfs_read_t / gpfs_vfs_read
- **gpfs_create_avg_lat (latency)**: gpfs_vfs_create_t / gpfs_vfs_create
- **gpfs_remove_avg_lat (latency)**: gpfs_vfs_remove_t / gpfs_vfs_remove

List of AFM metrics:

You can only use AFM metric if your system has GPFS configured. The following section lists all the AFM metrics:

GPFSAFM

- **gpfs_afm_bytes_read**: Total number of bytes read from remote system as a result of cache miss.
- **gpfs_afm_bytes_written**: Total number of bytes written to the remote system as a result of cache updates.
- **gpfs_afm_ops_expired**: Number of operations that were sent to remote system because they were expired, i.e. waited the configured async timeout in the gateway queue.

- **gpfs_afm_ops_forced**: Number of operations that were sent to remote system because they were forced out of the gateway queue before the configured async timeout, perhaps due to a dependent operation.
- **gpfs_afm_ops_sync**: Number of synchronous operations that were sent to remote system.
- **gpfs_afm_ops_revoked**: Number of operations that were sent to the remote system because a conflicting token acquired from another GPFS node resulted in a revoke.
- **gpfs_afm_bytes_pending**: Total number of bytes pending, i.e. not yet written to the remote system.
- **gpfs_afm_ops_sent**: Total number of operations sent over the communication protocol to the remote system.
- **gpfs_afm_shortest_time**: Shortest time in seconds that a pending operation waited in the gateway queue before being sent to remote system.
- **gpfs_afm_longest_time**: Longest time in seconds that a pending operation waited in the gateway queue before being sent to remote system.
- **gpfs_afm_avg_time**: Average time in seconds that a pending operation waited in the gateway queue before being sent to remote system.
- **gpfs_afm_tot_read_time**: Total time in seconds to perform read operations from the remote system.
- **gpfs_afm_tot_write_time**: Total time in seconds to perform write operations to the remote system.
- **gpfs_afm_conn_esta**: Total number of times a connection was established with the remote system.
- **gpfs_afm_conn_broken**: Total number of times the connection to the remote system was broken.
- **gpfs_afm_fset_expired**: Total number of times the fileset was marked expired due to a disconnection with remote system and expiry of the configured timeout.
- **gpfs_afm_used_q_memory**: Used memory in bytes by the messages enqueued.
- **gpfs_afm_num_queued_msgs**: Number of messages that are currently enqueued.

GPFSAFMFS

- **gpfs_afm_fs_bytes_read**: Total number of bytes read from remote system as a result of cache miss for this filesystem.
- **gpfs_afm_fs_bytes_written**: Total number of bytes written to the remote system as a result of cache updates for this filesystem.
- **gpfs_afm_fs_ops_expired**: Number of operations that were sent to remote system because they were expired, i.e. waited the configured async timeout in the gateway queue for this filesystem.
- **gpfs_afm_fs_ops_forced**: Number of operations that were sent to remote system because they were forced out of the gateway queue before the configured async timeout, perhaps due to a dependent operation for this filesystem.
- **gpfs_afm_fs_ops_sync**: Number of synchronous operations that were sent to remote system for this filesystem.
- **gpfs_afm_fs_ops_revoked**: Number of operations that were sent to the remote system because a conflicting token acquired from another GPFS node resulted in a revoke for this filesystem.
- **gpfs_afm_fs_bytes_pending**: Total number of bytes pending, i.e. not yet written to the remote system for this filesystem.
- **gpfs_afm_fs_ops_sent**: Total number of operations sent over the communication protocol to the remote system for this filesystem.
- **gpfs_afm_fs_shortest_time**: Shortest time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this filesystem.
- **gpfs_afm_fs_longest_time**: Longest time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this filesystem.
- **gpfs_afm_fs_avg_time**: Average time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this filesystem.
- **gpfs_afm_fs_tot_read_time**: Total time in seconds to perform read operations from the remote system for this filesystem.

- **gpfs_afm_fs_tot_write_time**: Total time in seconds to perform write operations to the remote system for this filesystem.
- **gpfs_afm_fs_conn_esta**: Total number of times a connection was established with the remote system for this filesystem.
- **gpfs_afm_fs_conn_broken**: Total number of times the connection to the remote system was broken for this filesystem.
- **gpfs_afm_fs_fset_expired**: Total number of times the fileset was marked expired due to a disconnection with remote system and expiry of the configured timeout for this filesystem.
- **gpfs_afm_fs_used_q_memory**: Used memory in bytes by the messages queued for this filesystem.
- **gpfs_afm_fs_num_queued_msgs**: Number of messages that are currently queued for this filesystem.

GPFSAFMFSET

- **gpfs_afm_fset_bytes_read**: Total number of bytes read from remote system as a result of cache miss for this fileset.
- **gpfs_afm_fset_bytes_written**: Total number of bytes written to the remote system as a result of cache updates for this fileset.
- **gpfs_afm_fset_ops_expired**: Number of operations that were sent to remote system because they were expired, i.e. waited the configured async timeout in the gateway queue for this fileset.
- **gpfs_afm_fset_ops_forced**: Number of operations that were sent to remote system because they were forced out of the gateway queue before the configured async timeout, perhaps due to a dependent operation for this fileset.
- **gpfs_afm_fset_ops_sync**: Number of synchronous operations that were sent to remote system for this fileset.
- **gpfs_afm_fset_ops_revoked**: Number of operations that were sent to the remote system because a conflicting token acquired from another GPFS node resulted in a revoke for this fileset.
- **gpfs_afm_fset_bytes_pending**: Total number of bytes pending, i.e. not yet written to the remote system for this fileset.
- **gpfs_afm_fset_ops_sent**: Total number of operations sent over the communication protocol to the remote system for this fileset.
- **gpfs_afm_fset_shortest_time**: Shortest time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this fileset.
- **gpfs_afm_fset_longest_time**: Longest time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this fileset.
- **gpfs_afm_fset_avg_time**: Average time in seconds that a pending operation waited in the gateway queue before being sent to remote system for this fileset.
- **gpfs_afm_fset_tot_read_time**: Total time in seconds to perform read operations from the remote system for this fileset.
- **gpfs_afm_fset_tot_write_time**: Total time in seconds to perform write operations to the remote system for this fileset.
- **gpfs_afm_fset_conn_esta**: Total number of times a connection was established with the remote system for this fileset.
- **gpfs_afm_fset_conn_broken**: Total number of times the connection to the remote system was broken for this fileset.
- **gpfs_afm_fset_fset_expired**: Total number of times the fileset was marked expired due to a disconnection with remote system and expiry of the configured timeout for this fileset.
- **gpfs_afm_fset_used_q_memory**: Used memory in bytes by the messages queued for this fileset.
- **gpfs_afm_fset_num_queued_msgs**: Number of messages that are currently queued for this filesystem.

Note: GPFSAFM, GPFSAFMFS, and GPFSAFMFSET also have other metrics which indicate the statistics on the state of remote filesystem operations. These metrics appear in the following format:

- For GPFSAFM: `gpfs_afm_operation_state`
- For GPFSAFMFS: `gpfs_afm_fs_operation_state`
- For GPFSAFMFSET: `gpfs_afm_fset_operation_state`

The operation can be one of the following:

- lookup
- getattr
- readdir
- readlink
- create
- mkdir
- mknod
- remove
- rmdir
- rename
- chmod
- trunc
- stime
- link
- symlink
- setsttr
- setxattr
- open
- close
- read
- readsplit
- writesplit
- write

Each of these options can in turn have one of the following five states:

- queued
- inflight
- complete
- errors
- filter

For example, the following metrics are also available: `gpfs_afm_write_filter`, `gpfs_afm_fs_create_queued`, `gpfs_afm_fset_rmdir_inflight` etc.

Protocol metrics:

The following section lists all the protocol metrics for IBM Spectrum Scale:

NFS metrics:

The following section lists all the NFS metrics::

NFS

NFSIO

- **nfs_read_req**: Number of bytes requested for reading.
- **nfs_write_req**: Number of bytes requested for writing.
- **nfs_read**: Number of bytes transferred for reading.
- **nfs_write**: Number of bytes transferred for writing.
- **nfs_read_ops**: Number of total read operations.
- **nfs_write_ops**: Number of total write operations.
- **nfs_read_err**: Number of erroneous read operations.
- **nfs_write_err**: Number of erroneous write operations.
- **nfs_read_lat**: Time consumed by read operations (in ns).
- **nfs_write_lat**: Time consumed by write operations (in ns).
- **nfs_read_queue**: Time spent in the rpc wait queue.
- **nfs_write_queue**: Time spent in the rpc wait queue.

Computed Metrics

The following metrics are computed for NFS. These metrics can only be used only through the mmperfmon query command.

- **nfs_total_ops**: nfs_read_ops + nfs_write_ops
- **nfsIOlatencyRead**: (nfs_read_lat + nfs_read_queue) / nfs_read_ops
- **nfsIOlatencyWrite**: (nfs_write_lat + nfs_write_queue) / nfs_write_ops
- **nfsReadOpThroughput**: nfs_read/nfs_read_ops
- **nfsWriteOpThroughput**: nfs_write/nfs_write_ops

Object metrics:

The following section lists all the object metrics:

SwiftAccount

- **account_auditor_time**: Timing data for individual account database audits.
- **account_reaper_time**: Timing data for each reap_account() call.
- **account_replicator_time**: Timing data for each database replication attempt not resulting in a failure.
- **account_DEL_time**: Timing data for each DELETE request not resulting in an error.
- **account_DEL_err_time**: Timing data for each DELETE request resulting in an error: bad request, not mounted, missing timestamp.
- **account_GET_time**: Timing data for each GET request not resulting in an error.
- **account_GET_err_time**: Timing data for each GET request resulting in an error: bad request, not mounted, bad delimiter, account listing limit too high, bad accept header.
- **account_HEAD_time**: Timing data for each HEAD request not resulting in an error.
- **account_HEAD_err_time**: Timing data for each HEAD request resulting in an error: bad request, not mounted.
- **account_POST_time**: Timing data for each POST request not resulting in an error.
- **account_POST_err_time**: Timing data for each POST request resulting in an error: bad request, bad or missing timestamp, not mounted.
- **account_PUT_time**: Timing data for each PUT request not resulting in an error.
- **account_PUT_err_time**: Timing data for each PUT request resulting in an error: bad request, not mounted, conflict, recently-deleted.

- **account_REPLICATE_time**: Timing data for each REPLICATE request not resulting in an error.
- **account_REPLICATE_err_time**: Timing data for each REPLICATE request resulting in an error: bad request, not mounted.

SwiftContainer

- **container_auditor_time**: Timing data for each container audit.
- **container_replicator_time**: Timing data for each database replication attempt not resulting in a failure.
- **container_DEL_time**: Timing data for each DELETE request not resulting in an error.
- **container_DEL_err_time**: Timing data for DELETE request errors: bad request, not mounted, missing timestamp, conflict.
- **container_GET_time**: Timing data for each GET request not resulting in an error.
- **container_GET_err_time**: Timing data for GET request errors: bad request, not mounted, parameters not utf8, bad accept header.
- **container_HEAD_time**: Timing data for each HEAD request not resulting in an error.
- **container_HEAD_err_time**: Timing data for HEAD request errors: bad request, not mounted.
- **container_POST_time**: Timing data for each POST request not resulting in an error.
- **container_POST_err_time**: Timing data for POST request errors: bad request, bad x-container-sync-to, not mounted.
- **container_PUT_time**: Timing data for each PUT request not resulting in an error.
- **container_PUT_err_time**: Timing data for PUT request errors: bad request, missing timestamp, not mounted, conflict.
- **container_REPLICATE_time**: Timing data for each REPLICATE request not resulting in an error.
- **container_REPLICATE_err_time**: Timing data for REPLICATE request errors: bad request, not mounted.
- **container_sync_deletes_time**: Timing data for each container database row synchronization via deletion.
- **container_sync_puts_time**: Timing data for each container database row synchronization via PUTing.
- **container_updater_time**: Timing data for processing a container; only includes timing for containers which needed to update their accounts.

SwiftObject

- **object_auditor_time**: Timing data for each object audit (does not include any rate-limiting sleep time for max_files_per_second, but does include rate-limiting sleep time for max_bytes_per_second).
- **object_expirer_time**: Timing data for each object expiration attempt, including ones resulting in an error.
- **object_replicator_partition_delete_time**: Timing data for partitions replicated to another node because they didn't belong on this node. This metric is not tracked per device.
- **object_replicator_partition_update_time**: Timing data for partitions replicated which also belong on this node. This metric is not tracked per-device.
- **object_DEL_time**: Timing data for each DELETE request not resulting in an error.
- **object_DEL_err_time**: Timing data for DELETE request errors: bad request, missing timestamp, not mounted, precondition failed. Includes requests which couldn't find or match the object.
- **object_GET_time**: Timing data for each GET request not resulting in an error. Includes requests which couldn't find the object (including disk errors resulting in file quarantine).
- **object_GET_err_time**: Timing data for GET request errors: bad request, not mounted, header timestamps before the epoch, precondition failed. File errors resulting in a quarantine are not counted here.
- **object_HEAD_time**: Timing data for each HEAD request not resulting in an error. Includes requests which couldn't find the object (including disk errors resulting in file quarantine).

- **object_HEAD_err_time**: Timing data for HEAD request errors: bad request, not mounted.
- **object_POST_time**: Timing data for each POST request not resulting in an error.
- **object_POST_err_time**: Timing data for POST request errors: bad request, missing timestamp, delete-at in past, not mounted.
- **object_PUT_time**: Timing data for each PUT request not resulting in an error.
- **object_PUT_err_time**: Timing data for PUT request errors: bad request, not mounted, missing timestamp, object creation constraint violation, delete-at in past.
- **object_REPLICATE_time**: Timing data for each REPLICATE request not resulting in an error.
- **object_REPLICATE_err_time**: Timing data for REPLICATE request errors: bad request, not mounted.
- **object_updater_time**: Timing data for object sweeps to flush async_pending container updates. Does not include object sweeps which did not find an existing async_pending storage directory.

SwiftProxy

- **proxy_account_latency**: Timing data up to completion of sending the response headers, 200: standard response for successful HTTP requests.
- **proxy_container_latency**: Timing data up to completion of sending the response headers, 200: standard response for successful HTTP requests.
- **proxy_object_latency**: Timing data up to completion of sending the response headers, 200: standard response for successful HTTP requests.
- **proxy_account_GET_time**: Timing data for GET request, start to finish, 200: standard response for successful HTTP requests
- **proxy_account_GET_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 200: standard response for successful HTTP requests.
- **proxy_account_HEAD_time**: Timing data for HEAD request, start to finish, 204: request processed, no content returned.
- **proxy_account_HEAD_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 204: request processed, no content returned.
- **proxy_container_DEL_time**: Timing data for DELETE request, start to finish, 204: request processed, no content returned.
- **proxy_container_DEL_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 204: request processed, no content returned.
- **proxy_container_GET_time**: Timing data for GET request, start to finish, 200: standard response for successful HTTP requests.
- **proxy_container_GET_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 200: standard response for successful HTTP requests.
- **proxy_container_HEAD_time**: Timing data for HEAD request, start to finish, 204: request processed, no content returned.
- **proxy_container_HEAD_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 204: request processed, no content returned. 1
- **proxy_container_PUT_time**: Timing data for each PUT request not resulting in an error, 201: request has been fulfilled; new resource created.
- **proxy_container_PUT_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 201: request has been fulfilled; new resource created.
- **proxy_object_DEL_time**: Timing data for DELETE request, start to finish, 204: request processed, no content returned.
- **proxy_object_DEL_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 204: request processed, no content returned.
- **proxy_object_GET_time**: Timing data for GET request, start to finish, 200: standard response for successful HTTP requests.

- **proxy_object_GET_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 200: standard response for successful HTTP requests.
- **proxy_object_HEAD_time**: Timing data for HEAD request, start to finish, 200: request processed, no content returned.
- **proxy_object_HEAD_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, , 200: request processed, no content returned.
- **proxy_object_PUT_time**: Timing data for each PUT request not resulting in an error, 201: request has been fulfilled; new resource created.
- **proxy_object_PUT_bytes**: The sum of bytes transferred in (from clients) and out (to clients) for requests, 201: request has been fulfilled; new resource created.

Note: For information about computed metrics for object, see "Performance monitoring for object metrics" on page 74.

SMB metrics:

The following section lists all the SMB metrics::

SMBGlobalStats
- **connect count**: Number of connections since startup of parent smbd process
- **disconnect count**: Number of connections closed since startup
- **idle**: Describes idling behavior of smbds
  - **count**: Number of times the smbd processes are waiting for events in epoll
  - **time**: Times the smbd process spend in epoll waiting for events
- **cpu_user time**: The user time determined by the get_rusage system call in seconds
- **cpu_system time**: The system time determined by the get_rusage system call in seconds
- **request count**: Number of SMB requests since startup
- **push_sec_ctx**: Smbds switch between the user and the root security context; push allows to put the current context onto a stack
  - **count**: Number of time the current security context is pushed onto the stack
  - **time**: The time it takes to put the current security context; this includes all syscalls required to save the current context on the stack
- **pop_sec_ctx**: Getting the last security context from the stack and restore it
  - **count**: Number of times the current security context is restored from the stack
  - **time**: The time it takes to put the restore the security context from the stack; this includes all syscalls required to get restore the security context from the stack
- **set_sec_ctx**:
  - **count**: Number of times the security context is set for user
  - **time**: The time it takes to set the security context for user
- **set_root_sec_ctx**:
  - **count**: Number of times the security context is set for user
  - **time**: The time it takes to set the security context for user

SMB2 metrics

These metrics are available for all of the following areas:
- **op_count**: Number of times the corresponding SMB request has been called.
- **op_idle**
  - **for notify**: Time between notification request and a corresponding notification being sent

  - **for oplock breaks**: Time waiting until an oplock is broken
  - for all others the value is always zero
- **op_inbytes**: Number of bytes received for the corresponding request including protocol headers
- **op_outbytes**: Number of bytes sent for the corresponding request including protocol headers.
- **op_time**: The total amount of time spent for all corresponding SMB2 requests.

CTDB metrics:

The following section lists all the CTDB metrics::
- **CTDB version**: Version of the CTDB protocol used by the node.
- **Current time of statistics**: Time when the statistics are generated. This is useful when collecting statistics output periodically for post-processing.
- **Statistics collected since**: Time when CTDB was started or the last time statistics was reset. The output shows the duration and the timestamp.
- **num_clients**: Number of processes currently connected to CTDB's UNIX socket. This includes recovery daemon, CTDB tool and SMB processes (smbd, winbindd).
- **frozen**: 1 if the databases are currently frozen, 0 if otherwise.
- **recovering**: 1 if recovery is active, 0 if otherwise.
- **num_recoveries**: Number of recoveries since the start of CTDB or since the last statistics reset.
- **client_packets_sent**: Number of packets sent to client processes via UNIX domain socket.
- **client_packets_recv**: Number of packets received from client processes via UNIX domain socket.
- **node_packets_sent**: Number of packets sent to the other nodes in the cluster via TCP.
- **node_packets_recv**: Number of packets received from the other nodes in the cluster via TCP.
- **keepalive_packets_sent**: Number of keepalive messages sent to other nodes. CTDB periodically sends keepalive messages to other nodes. For more information, see the KeepAliveInterval tunable in CTDB-tunables(7) on the CTDB documentation website.
- **keepalive_packets_recv**: Number of keepalive messages received from other nodes.
- node: This section lists various types of messages processed which originated from other nodes via TCP.
  - **req_call**: Number of REQ_CALL messages from the other nodes.
  - **reply_call**: Number of REPLY_CALL messages from the other nodes.
  - **req_dmaster**: Number of REQ_DMASTER messages from the other nodes.
  - **reply_dmaster**: Number of REPLY_DMASTER messages from the other nodes.
  - **reply_error**: Number of REPLY_ERROR messages from the other nodes.
  - **req_message**: Number of REQ_MESSAGE messages from the other nodes.
  - **req_control**: Number of REQ_CONTROL messages from the other nodes.
  - **reply_control**: Number of REPLY_CONTROL messages from the other nodes.
- **client**: This section lists various types of messages processed which originated from clients via UNIX domain socket.
  - **req_call**: Number of REQ_CALL messages from the clients.
  - **req_message**: Number of REQ_MESSAGE messages from the clients.
  - **req_control**: Number of REQ_CONTROL messages from the clients.
- **timeouts**: This section lists timeouts occurred when sending various messages.
  - **call**: Number of timeouts for REQ_CALL messages.
  - **control**: Number of timeouts for REQ_CONTROL messages.
  - **traverse**: Number of timeouts for database traverse operations.
- **locks**: This section lists locking statistics.

  - **num_calls**: Number of completed lock calls. This includes database locks and record locks.
  - **num_current**: Number of scheduled lock calls. This includes database locks and record locks.
  - **num_pending**: Number of queued lock calls. This includes database locks and record locks.
  - **num_failed**: Number of failed lock calls. This includes database locks and record locks.
- **total_calls**: Number of req_call messages processed from clients. This number should be same as client --> req_call.
- **pending_calls**: Number of req_call messages which are currently being processed. This number indicates the number of record migrations in flight.
- **childwrite_calls**: Number of record update calls. Record update calls are used to update a record under a transaction.
- **pending_childwrite_calls**: Number of record update calls currently active.
- **memory_used**: The amount of memory in bytes currently used by CTDB using talloc. This includes all the memory used for CTDBÂ´s internal data structures. This does not include the memory mapped TDB databases.
- **max_hop_count**: The maximum number of hops required for a record migration request to obtain the record. High numbers indicate record contention.
- **total_ro_delegations**: Number of read-only delegations created.
- **total_ro_revokes**: Number of read-only delegations that were revoked. The difference between total_ro_revokes and total_ro_delegations gives the number of currently active read-only delegations.
- **hop_count_buckets**: Distribution of migration requests based on hop counts values.
- **lock_buckets**: Distribution of record lock requests based on time required to obtain locks. Buckets are < 1ms, < 10ms, < 100ms, < 1s, < 2s, < 4s, < 8s, < 16s, < 32s, < 64s, > 64s.
- **locks_latency**: The minimum, the average and the maximum time (in seconds) required to obtain record locks.
- **reclock_ctdbd**: The minimum, the average and the maximum time (in seconds) required to check if recovery lock is still held by recovery daemon when recovery mode is changed. This check is done in ctdb daemon.
- **reclock_recd**: The minimum, the average and the maximum time (in seconds) required to check if recovery lock is still held by recovery daemon during recovery. This check is done in recovery daemon.
- **call_latency**: The minimum, the average and the maximum time (in seconds) required to process a REQ_CALL message from client. This includes the time required to migrate a record from remote node, if the record is not available on the local node.
- **childwrite_latency**: The minimum, the average and the maximum time (in seconds) required to update records under a transaction.

Cross protocol metrics:

The following section lists all the cross protocol metrics::

- **nfs_iorate_read_perc**: nfs_read_ops/(op_count+nfs_read_ops)
- **nfs_iorate_read_perc_exports**: 1.0*nfs_read_ops/(op_count+nfs_read_ops)
- **nfs_iorate_write_perc**: nfs_write_ops/(write | op_count+nfs_write_ops)
- **nfs_iorate_write_perc_exports**: 1.0*nfs_write_ops/(op_count+nfs_write_ops)
- **nfs_read_throughput_perc**: nfs_read/(read | op_outbytes+nfs_read)
- **nfs_write_throughput_perc**: nfs_write/(write | op_outbytes+nfs_write)
- **smb_iorate_read_perc**: op_count/(op_count+nfs_read_ops)
- **smb_iorate_write_perc**: op_count/(op_count+nfs_write_ops)
- **smb_latency_read**: read | op_time/read | op_count
- **smb_latency_write**: write | op_time/write | op_count
- **smb_read_throughput_perc**: read | op_outbytes/(read | op_outbytes+nfs_read)

- **smb_total_cnt**: write | op_count+close | op_count
- **smb_tp**: op_inbytes+op_outbytes
- **smb_write_throughput_perc**: write | op_outbytes/(write | op_outbytes+nfs_write)
- **total_read_throughput**: nfs_read+read | op_outbytes
- **total_write_throughput**: nfs_write+write | op_inbytes

Cloud services metrics:

The following section lists all the metrics for Cloud services:

Cloud services
- **mcs_total_bytes**: Total number of bytes uploaded to or downloaded from the cloud storage tier.
- **mcs_total_requests**: Total number of migration, recall, or remove requests.
- **mcs_total_request_time**: Time (in second) taken for all migration, recall, or remove requests.
- **mcs_total_failed_requests**: Total number of failed migration, recall, or remove requests.
- **mcs_total_failed_requests_time**: The total time (msec) spent in failed migration, recall, or remove requests.
- **mcs_total_persisted_bytes**: The total number of transferred bytes that are successfully persisted on the cloud provider. This is used for both migrate and recall operations.
- **mcs_total_retried_operations**: The total number of retry PUT operations. This is used for both migrate and recall operations.
- **mcs_total_operation_errors**: The total number of erroneous PUT/GET operations based on the operation specified in the mcs_operation key.
- **mcs_total_successful_operations**: The total number of successful PUT/GET operations for both data and metadata.
- **mcs_total_operation_time**: The total time taken (msec) for PUT /GET operations for both data and metadata.
- **mcs_total_persisted_time**: For PUT, the total time taken (msec) for transferring and persisting the bytes on the cloud provider. For GET, the total time taken (msec) for downloading and persisting the bytes on the file system.
- **mcs_total_failed_operations**: The total number of failed PUT/GET operations.
- **mcs_total_operation_errors_time**: The total time taken (msec) for erroneous PUT /GET operations.
- **mcs_total_persisted_parts**:The total number of transferred parts persisted successfully on the cloud provider in case of multipart upload.
- **mcs_total_parts**: The total number of parts transferred to the cloud provider in case of multipart upload.
- **tct_fset_total_bytes**: Total number of bytes uploaded to or downloaded from the cloud storage tier with respect to a fileset.
- **tct_fset_total_successful_operations**: The total number of successful PUT/GET operations for both data and metadata with respect to a fileset.
- **tct_fset_total_operation_time**: The total time taken (msec) for PUT /GET operations for both data and metadata with respect to a fileset.
- **tct_fset_total_persisted_bytes**: The total number of transferred bytes from a fileset that are successfully persisted on the cloud provider. This is used for both migrate and recall operations.
- **tct_fset_total_persisted_time**:For PUT, the total time taken (msec) for transferring and persisting the bytes on the cloud provider. For GET, the total time taken (msec) for downloading and persisting the bytes on the fileset.
- **tct_fset_total_retried_operations**: The total number of retry PUT operations with respect to a fileset. This is used for both migrate and recall operations.

- **tct_fset_total_failed_operations**: The total number of failed PUT/GET operations with respect to a fileset.
- **tct_fset_total_operation_errors**: The total number of erroneous PUT/GET operations with respect to a fileset based on the operation specified in the mcs_operation key
- **tct_fset_total_operation_errors_time**: The total time taken (msec) for erroneous PUT /GET operations with respect to a fileset.
- **tct_fset_total_persisted_parts**: The total number of transferred parts (from a fileset) persisted successfully on the cloud provider in case of multipart upload.
- **tct_fset_total_parts**: The total number of parts transferred to the cloud provider from a fileset in case of a multipart upload.
- **tct_fset_csap_used**: Total number of bytes used by a fileset for a specific CSAP.
- **tct_fset_total_requests**: Total number of migration, recall, or remove requests with respect to a fileset.
- **tct_fset_total_request_time**: Time (in second) taken for all migration, recall, or remove requests with respect to a fileset.
- **tct_fset_total_failed_requests**: Total number of failed migration, recall, or remove requests with respect to a fileset.
- **tct_fset_total_failed_requests_time**: The total time (msec) spent in failed migration, recall, or remove requests with respect to a fileset.
- **tct_fset_total_blob_time**: The total blob time on the fileset.
- **tct_fs_total_successful_operations**: The total number of successful PUT/GET operations for both data and metadata with respect to a file system.
- **tct_fs_total_operation_time**: The total time taken (msec) for PUT /GET operations for both data and metadata with respect to a file system.
- **tct_fs_total_persisted_bytes**: The total number of transferred bytes from a file system that are successfully persisted on the cloud provider. This is used for both migrate and recall operations.
- **tct_fs_total_persisted_time**:For PUT, the total time taken (msec) for transferring and persisting the bytes on the cloud provider. For GET, the total time taken (msec) for downloading and persisting the bytes on the file system.
- **tct_fs_total_retried_operations**: The total number of retry PUT operations with respect to a file system. This is used for both migrate and recall operations.
- **tct_fs_total_failed_operations**: The total number of failed PUT/GET operations with respect to a file system.
- **tct_fs_total_operation_errors**: The total number of erroneous PUT/GET operations with respect to a file system based on the operation specified in the mcs_operation key
- **tct_fs_total_operation_errors_time**: The total time taken (msec) for erroneous PUT /GET operations with respect to a file system.
- **tct_fs_total_persisted_parts**: The total number of transferred parts (from a file system) persisted successfully on the cloud provider in case of multipart upload.
- **tct_fs_total_parts**: The total number of parts transferred to the cloud provider from a file system in case of a multipart upload.
- **tct_fs_csap_used**: Total number of bytes used by a file system for a specific CSAP.
- **tct_fs_total_requests**: Total number of migration, recall, or remove requests with respect to a file system.
- **tct_fs_total_request_time**: Time (in second) taken for all migration, recall, or remove requests with respect to a file system.
- **tct_fs_total_failed_requests**: Total number of failed migration, recall, or remove requests with respect to a file system.
- **tct_fs_total_failed_requests_time**: The total time (msec) spent in failed migration, recall, or remove requests with respect to a file system.

- **tct_fs_total_blob_time**: The total blob time on the file system.

Performance monitoring for object metrics

The **mmperfmon** command can be used to obtain object metrics information. Ensure that `pmswift` is configured and the object sensors are added to measure the object metrics.

The **mmperfmon** command is enhanced to calculate and print the sum, average, count, minimum, and maximum of metric data for object queries. The following command can be used to display metric data for object queries:

```
mmperfmon query NamedQuery [StartTime EndTime]
```

Currently, the calculation of the sum, average, count, minimum, and maximum is only applicable for the following object metrics:

- **account_HEAD_time**
- **account_GET_time**
- **account_PUT_time**
- **account_POST_time**
- **account_DEL_time**
- **container_HEAD_time**
- **container_GET_time**
- **container_PUT_time**
- **container_POST_time**
- **container_DEL_time**
- **object_HEAD_time**
- **object_GET_time**
- **object_PUT_time**
- **object_POST_time**
- **object_DEL_time**
- **proxy_account_latency**
- **proxy_container_latency**
- **proxy_object_latency**
- **proxy_account_GET_time**
- **proxy_account_GET_bytes**
- **proxy_account_HEAD_time**
- **proxy_account_HEAD_bytes**
- **proxy_account_POST_time**
- **proxy_account_POST_bytes**
- **proxy_container_GET_time**
- **proxy_container_GET_bytes**
- **proxy_container_HEAD_time**
- **proxy_container_HEAD_bytes**
- **proxy_container_POST_time**
- **proxy_container_POST_bytes**
- **proxy_container_PUT_time**
- **poxy_container_PUT_bytes**
- **proxy_container_PUT_time**
- **proxy_container_PUT_bytes**

- **proxy_container_DEL_time**
- **proxy_container_DEL_bytes**
- **proxy_object_GET_time**
- **proxy_object_GET_bytes**
- **proxy_object_HEAD_time**
- **proxy_object_HEAD_bytes**
- **proxy_object_POST_time**
- **proxy_object_POST_bytes**
- **proxy_object_PUT_time**
- **proxy_object_PUT_bytes**
- **proxy_object_PUT_time**
- **proxy_object_PUT_bytes**
- **proxy_object_DEL_time**
- **proxy_object_DEL_bytes**
- **proxy_object_POST_time**
- **proxy_object_POST_bytes**

To run a `objObj` query for object metrics, issue the following command. This command calculates and prints the sum, average, count, minimum, and maximum of metric data for the object `objObj` for all the metrics mentioned above.

```
mmperfmon query objObj 2016-09-28-09:56:39 2016-09-28-09:56:43

1: cluster1.ibm.com|SwiftObject|object_auditor_time
2:  cluster1.ibm.com|SwiftObject|object_expirer_time
3:  cluster1.ibm.com|SwiftObject|object_replication_partition_delete_time
4:  cluster1.ibm.com|SwiftObject|object_replication_partition_update_time
5:  cluster1.ibm.com|SwiftObject|object_DEL_time
6:  cluster1.ibm.com|SwiftObject|object_DEL_err_time
7:  cluster1.ibm.com|SwiftObject|object_GET_time
8:  cluster1.ibm.com|SwiftObject|object_GET_err_time
9:  cluster1.ibm.com|SwiftObject|object_HEAD_time
10: cluster1.ibm.com|SwiftObject|object_HEAD_err_time
11: cluster1.ibm.com|SwiftObject|object_POST_time
12: cluster1.ibm.com|SwiftObject|object_POST_err_time
13: cluster1.ibm.com|SwiftObject|object_PUT_time
14: cluster1.ibm.com|SwiftObject|object_PUT_err_time
15: cluster1.ibm.com|SwiftObject|object_REPLICATE_time
16: cluster1.ibm.com|SwiftObject|object_REPLICATE_err_time
17: cluster1.ibm.com|SwiftObject|object_updater_time

Row object_auditor_time object_expirer_time  object_replication_partition_delete_time
object_replication_partition_update_time  object_DEL_time  object_DEL_err_time
object_GET_time  object_GET_err_time  object_HEAD_time object_HEAD_err_time object_POST_time
object_POST_err_time  object_PUT_time  object_PUT_err_time  object_REPLICATE_time
object_REPLICATE_err_time  object_updater_time
1 2016-09-28 09:56:39 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
  0.855923 0.000000 0.000000 0.000000 45.337915 0.000000 0.000000 0.000000 0.000000
2 2016-09-28 09:56:40 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
  0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
3 2016-09-28 09:56:41 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
  0.931925 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
4 2016-09-28 09:56:42 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
  0.855923 0.000000 0.000000 0.000000 516.280890 0.000000 0.000000 0.000000 0.000000

object_DEL_total_time   = 0.0           object_PUT_total_time = 561.618805
object_GET_total_time   = 0.0           object_POST_total_time = 0.0
object_HEAD_total_time  = 1.786948      object_PUT_max_time = 516.28089
object_POST_max_time    = 0.0           object_GET_max_time = 0.0
object_HEAD_max_time    = 0.931025      object_DEL_max_time = 0.0
object_GET_avg_time     = 0.0           object_DEL_avg_time = 0.0
```

```
object_PUT_avg_time     = 280.809402    object_POST_avg_time = 0.0
object_HEAD_avg_time    = 0.893474      object_DEL_time_count = 0.0
object_POST_time_count  = 0             object_PUT_time_count = 2
object_HEAD_time_count  = 2             object_GET_time_count = 0
object_DEL_min_time     = 0.0           object_PUT_min_time = 45.337915
object_GET_min_time     = 0.0           object_POST_min_time = 0.0
object_HEAD_min_time    = 0.855923
```

Enabling protocol metrics

The type of information that is collected for NFS, SMB and Object protocols are configurable. This section describes the location of the configuration data for these protocols.

Configuration information for SMB and NFS in the `ZimonSensors.cfg` file references the sensor definition files in the `/opt/IBM/zimon` folder. For example:

- The `CTDBDBStats.cfg` file is referred in:

```
{       name = "CTDBDBStats"
        period = 1
        type = "Generic"
},
```

- The `CTDBStats.cfg` file is referred in:

```
{       name = "CTDBStats"
        period = 1
        type = "Generic"
},
```

- The `NFSIO.cfg` file is referred in:

```
{
        # NFS Ganesha statistics
        name = "NFSIO"
        period = 1
        type = "Generic"
},
```

- The `SMBGlobalStats.cfg` file is referred in:

```
{       name = "SMBGlobalStats"
        period = 1
        type = "Generic"
},
```

- The `SMBStats.cfg` file is referred in:

```
{       name = "SMBStats"
        period = 1
        type = "Generic"
},
```

At the time of installation, the object metrics proxy is configured to start by default on each Object protocol node.

The object metrics proxy server, **pmswiftd** is controlled by the corresponding service script called `pmswiftd`, located at `/etc/rc.d/init.d/pmswiftd.service`. You can start and stop the pmswiftd service script using the **systemctl start pmswiftd** and **systemctl stop pmswiftd** commands respectively. You can also view the status of the pmswiftd service script by using the **systemctl status pmswiftd** command.

In a system restart, the object metrics proxy server restarts automatically. In case of a failover, the server will start automatically. If for some reason this does not occur, the server must be started manually using the **systemctl start pmswiftd** command.

Starting and stopping the performance monitoring tool

You can start and stop the performance monitoring tool using the following commands:

Starting the performance monitoring tool

Use the **systemctl start pmsensors** command to start performance monitoring on a node.

Use the **systemctl start pmcollector** command on a node that has the collector.

Stopping the performance monitoring tool

Use the **systemctl stop pmsensors** command to stop sensor service on all nodes where active.

Use the **systemctl stop pmcollector** command to stop collector service on nodes where GUI is installed.

Note:

The **systemctl** commands only work for systems that use **systemd** scripts. On systems that use **sysv** initialization scripts, you must use the **service pmsensors** and **service pmcollector** commands instead of the **systemctl** commands.

Restarting the performance monitoring tool

If the pmsensors or pmcollector package is upgraded, the corresponding daemon is stopped and needs to be started again.

To start the sensor on a particular node, use the **systemctl start pmsensors** command. To start the collector, use the **systemctl start pmcollector** command.

If the `ZIMonCollector.cfg` file is changed, the pmsensors service on that node needs to be restarted with **systemctl restart pmcollector** command.

With manual configuration, if the `ZIMonSensors.cfg` file is changed, the pmsensors service on that node needs to be restarted using the **systemctl restart pmsensors** command. No action is necessary for IBM Spectrum Scale managed sensor configuration.

To restart the collector, use the **systemctl restart pmcollector** command.

Note:

This command only works for systems that use systemd scripts. On systems that use sysv initialization scripts, you must use the **service pmsensors** and **service pmcollector** command instead of the **systemctl** command.

For information on restarting the sensors and collectors for Transparent cloud tiering, see *Integrating Transparent Cloud Tiering metrics with performance monitoring tool* in *IBM Spectrum Scale: Administration Guide*.

Configuring the metrics to collect performance data

For performance reasons, the performance monitoring tool by default does not collect all the available metrics. You can add other metrics to focus on particular performance problems.

For the available metrics, see “List of performance metrics” on page 53.

For information on sensor configuration, see “Configuring the sensor” on page 46.

Viewing and analyzing the performance data

The performance monitoring tool allows you to view the metrics associated with GPFS and the associated protocols, get a graphical representation of the status and trends of the key performance indicators, and analyze IBM Spectrum Scale performance problems.

You can view and analyze the performance monitoring data using the following methods:

- Using the **mmperfmon** command.
- Using an open source visualization tool called Grafana.

Note: You may also monitor the performance through IBM Spectrum Scale GUI. For more information on using the IBM Spectrum Scale GUI for performance monitoring, see "Performance monitoring using IBM Spectrum Scale GUI" on page 87. The performance data that is available with **mmperfmon** query, GUI or any other visualization tool depends on the which sensors are installed and enabled. This can be determined by looking at the sensor configuration. For more information on sensor configuration, see the *Configuring the sensor* section in the *IBM Spectrum Scale: Problem Determination Guide*.

Viewing performance data with mmperfmon

To view the metrics associated with GPFS and the associated protocols, run the **mmperfmon** command with the **query** option. You can also use the **mmperfmon** command with the **query** option to detect performance issues and problems. You can collect metrics for all nodes or for a particular node.

- Problem: System slowing down

  Use **mmperfmon query compareNodes cpu_user** or **mmperfmon query compareNodes cpu_system** command to compare CPU metrics for all the nodes in your system.

  1. Check if there is a node that has a significantly higher CPU utilization for the entire time period. If so, see if this trend continues. You might need to investigate further on this node.
  2. Check if there is a node that has significantly lower CPU utilization over the entire period. If so, check if that node has a health problem.
  3. Use **mmperfmon query compareNodes protocolThroughput** to look at the throughput for each of the nodes for the different protocols.

     Note: Note that the metrics of each individual protocol cannot always include exact I/O figures.
  4. Use **mmperfmon query compareNodes protocolIORate** to look at the I/O performance for each of the nodes in your system.
- Problem: A particular node is causing problems

  Use **mmperfmon query usage** to show the CPU, memory, storage, and network usage.
- Problem: A particular protocol is causing problems

  Use **mmperfmon query** to investigate problems with your specific protocol. You can compare cross-node metrics using **mmperfmon query compareNodes**.

  For example, **mmperfmon query compareNodes nfs_read_ops.**

  Compare the NFS read operations on all the nodes that are using NFS. By comparing the different NFS metrics, you can identify which node is causing the problems. The problem might either manifest itself as running with much higher values than the other nodes, or much lower (depending on the issue) when considered over several buckets of time.
- Problem: A particular protocol is causing problems on a particular node.

  Use **mmperfmon query** on the particular node to look deeper into the protocol performance on that node.

  For example, if there is a problem with NFS:

  – **mmperfmon query nfsIOlatency** - To get details of the nfsIOlatency.
  – **mmperfmon query nfsIOrate** - To get details of the NFS I/O rate.
  – **mmperfmon query nfsThroughput** - To get details of the NFS throughput.

For more information on **mmperfmon**, see *mmperform* in *IBM Spectrum Scale: Command and Programming Reference*

List of queries:

You can make the following predefined queries with **query** option of the **mmperfmon** command.

General and network

- **usage**: Retrieves details about the CPU, memory, storage and network usage
- **cpu**: Retrieves details of the CPU utilization in system and user space, and context switches.
- **netDetails**: Retrieves details about the network.
- **NetErrors**: Retrieves details about network problems, such as collisions, drops, and errors, for all available networks.
- **compareNodes**: Compares a single metric across all nodes running sensors

GPFS

GPFS metric queries gives an overall view of the GPFS without considering the protocols.

- **gpfsCRUDopsLatency**: Retrieves information about the GPFS CRUD operations latency
- **gpfsFSWaits**: Retrieves information on the maximum waits for read and write operations for all file systems.
- **gpfsNSDWaits**: Retrieves information on the maximum waits for read and write operations for all disks.
- **gpfsNumberOperations**: Retrieves the number of operations to the GPFS file system.
- **gpfsVFSOpCounts**: Retrieves VFS operation counts.

Cross protocol

These queries retrieve information after comparing metrics between different protocols on a particular node.

- **protocolIOLatency**: Compares latency per protocol (SMB, NFS, Object).
- **protocolIORate**: Retrieves the percentage of total I/O rate per protocol (SMB, NFS, Object).
- **protocolThroughput**: Retrieves the percentage of total throughput per protocol (SMB, NFS, Object).

NFS

These queries retrieve metrics associated with the NFS protocol.

- **nfsIOLatency**: Retrieves the NFS I/O Latency in nanoseconds.
- **nfsIORate**: Retrieves the NFS I/O operations per second (NFS IOPS).
- **nfsThroughput**: Retrieves the NFS Throughput in bytes per second.
- **nfsErrors**: Retrieves the NFS error count for read and write operations.
- **nfsQueue**: Retrieves the NFS read and write queue latency in nanoseconds.
- **nfsThroughputPerOp**: Retrieves the NFS read and write throughput per op in bytes

Object

- **objAcc**: Details on the Object Account performance
  Retrieved metrics:
  - **account_auditor_time**
  - **account_reaper_time**
  - **account_replicator_time**
  - **account_DEL_time**

  - **account_DEL_err_time**
  - **account_GET_time**
  - **account_GET_err_time**
  - **account_HEAD_time**
  - **account_HEAD_err_time**
  - **account_POST_time**
  - **account_POST_err_time**
  - **account_PUT_time**
  - **account_PUT_err_time**
  - **account_REPLICATE_time**
  - **account_REPLICATE_err_time**
- **objCon**: Details on the Object Container performance
  Retrieved metrics:
  - **container_auditor_time**
  - **container_replicator_time**
  - **container_DEL_time**
  - **container_DEL_err_time**
  - **container_GET_time**
  - **container_GET_err_time**
  - **container_HEAD_time**
  - **container_HEAD_err_time**
  - **container_POST_time**
  - **container_POST_err_time**
  - **container_PUT_time**
  - **container_PUT_err_time**
  - **container_REPLICATE_time**
  - **container_REPLICATE_err_time**
  - **container_sync_deletes_time**
  - **container_sync_puts_time**
  - **container_updater_time**
- **objObj**: Details on the Object performance
  Retrieved metrics:
  - **object_auditor_time**
  - **object_expirer_time**
  - **object_replicator_partition_delete_time**
  - **object_replicator_partition_update_time**
  - **object_DEL_time**
  - **object_DEL_err_time**
  - **object_GET_time**
  - **object_GET_err_time**
  - **object_HEAD_time**
  - **object_HEAD_err_time**
  - **object_POST_time**
  - **object_POST_err_time**
  - **object_PUT_time**

  - **object_PUT_err_time**
  - **object_REPLICATE_err_time**
  - **object_REPLICATE_time**
  - **object_updater_time**
- **objPro**: Details on the Object Proxy performance
  Retrieved metrics:
  - **proxy_account_latency**
  - **proxy_container_latency**
  - **proxy_object_latency**
  - **proxy_account_GET_time**
  - **proxy_account_GET_bytes**
  - **proxy_account_HEAD_time**
  - **proxy_account_HEAD_bytes**
  - **proxy_account_POST_time**
  - **proxy_account_POST_bytes**
  - **proxy_container_DEL_time**
  - **proxy_container_DEL_bytes**
  - **proxy_container_GET_time**
  - **proxy_container_GET_bytes**
  - **proxy_container_HEAD_time**
  - **proxy_container_HEAD_bytes**
  - **proxy_container_POST_time**
  - **proxy_container_POST_bytes**
  - **proxy_container_PUT_time**
  - **proxy_container_PUT_bytes**
  - **proxy_object_DEL_time**
  - **proxy_object_DEL_bytes**
  - **proxy_object_GET_time**
  - **proxy_object_GET_bytes**
  - **proxy_object_HEAD_time**
  - **proxy_object_HEAD_bytes**
  - **proxy_object_POST_time**
  - **proxy_object_POST_bytes**
  - **proxy_object_PUT_time**
  - **proxy_object_PUT_bytes**
- **objAccIO**: Information on the Object Account IO rate
  Retrieved metrics:
  - **account_GET_time**
  - **account_GET_err_time**
  - **account_HEAD_time**
  - **account_HEAD_err_time**
  - **account_POST_time**
  - **account_POST_err_time**
  - **account_PUT_time**
  - **account_PUT_err_time**

- **objConIO**: Information on the Object Container IO rate
  Retrieved metrics:
  - **container_GET_time**
  - **container_GET_err_time**
  - **container_HEAD_time**
  - **container_HEAD_err_time**
  - **container_POST_time**
  - **container_POST_err_time**
  - **container_PUT_time**
  - **container_PUT_err_time**
- **objObjIO**: Information on the Object Object IO rate
  Retrieved metrics:
  - **object_GET_time**
  - **object_GET_err_time**
  - **object_HEAD_time**
  - **object_HEAD_err_time**
  - **object_POST_time**
  - **object_POST_err_time**
  - **object_PUT_time**
  - **object_PUT_err_time**
- **objProIO**: Information on the Object Proxy IO rate
  Retrieved metrics:
  - **proxy_account_GET_time**
  - **proxy_account_GET_bytes**
  - **proxy_container_GET_time**
  - **proxy_container_GET_bytes**
  - **proxy_container_PUT_time**
  - **proxy_container_PUT_bytes**
  - **proxy_object_GET_time**
  - **proxy_object_GET_bytes**
  - **proxy_object_PUT_time**
  - **proxy_object_PUT_bytes**
- **objAccThroughput**: Information on the Object Account Throughput
  Retrieved metrics:
  - **account_GET_time**
  - **account_PUT_time**
- **objConThroughput**: Information on the Object Container Throughput
  Retrieved metrics:
  - **container_GET_time**
  - **container_PUT_time**
- **objObjThroughput**: Information on the Object Throughput
  Retrieved metrics:
  - **object_GET_time**
  - **object_PUT_time**
- **objProThroughput**: Information on the Object Proxy Throughput

Retrieved metrics:
  - **proxy_account_GET_time**
  - **proxy_account_GET_bytes**
  - **proxy_container_GET_time**
  - **proxy_container_GET_bytes**
  - **proxy_container_PUT_time**
  - **proxy_container_PUT_bytes**
  - **proxy_object_GET_time**
  - **proxy_object_GET_bytes**
  - **proxy_object_PUT_time**
  - **proxy_object_PUT_bytes**
- **objAccLatency**: Information on the Object Account Latency
  Retrieved metric:
  - **proxy_account_latency**
- **objConLatency**: Information on the Object Container Latency
  Retrieved metric:
  - **proxy_container_latency**
- **objObjLatency**: Information on the Object Latency
  Retrieved metric:
  - **proxy_object_latency**

SMB

These queries retrieve metrics associated with SMB.
- **smb2IOLatency**: Retrieves the SMB2 I/O latencies per bucket size (default 1 sec).
- **smb2IORate**: Retrieves the SMB2 I/O rate in number of operations per bucket size (default 1 sec).
- **smb2Throughput**: Retrieves the SMB2 Throughput in bytes per bucket size (default 1 sec).
- **smb2Writes** : Retrieves count, # of idle calls, bytes in and out, and operation time for SMB2 writes.
- **smbConnections**: - Retrieves the number of SMB connections.

CTDB

These queries retrieve metrics associated with CTDB.
- **ctdbCallLatency**: Retrieves information on the CTDB call latency.
- **ctdbHopCountDetails**: Retrieves information on the CTDB hop count buckets 0 to 5 for one database.
- **ctdbHopCounts** :Retrieves information on the CTDB hop counts (bucket 00 = 1-3 hops) for all databases.

Using IBM Spectrum Scale performance monitoring bridge with Grafana

The IBM Spectrum Scale performance monitoring bridge is a stand-alone Python application, which uses Grafana to display performance data. Grafana is an open source tool for visualizing time series and application metrics. It provides a powerful platform to create, explore, and share dashboards and data. .

IBM Spectrum Scale performance monitoring bridge could be used for exploring the performance data on Grafana dashboards. The IBM Spectrum Scale performance monitoring bridge emulates an openTSDB API, which is used by Grafana to set up and populate the graphs. The metadata received from IBM Spectrum Scale is used to create the Grafana graphs, and the data from IBM Spectrum Scale is used to populate these graphs. Two version of the IBM Spectrum Scale performance monitoring bridge are now available. You can download the latest version of the IBM Spectrum Scale performance monitoring bridge from the Prerequisite and Download page. For more information on the new version of the bridge, see

"New features of the IBM Spectrum Scale performance monitoring bridge version 2"

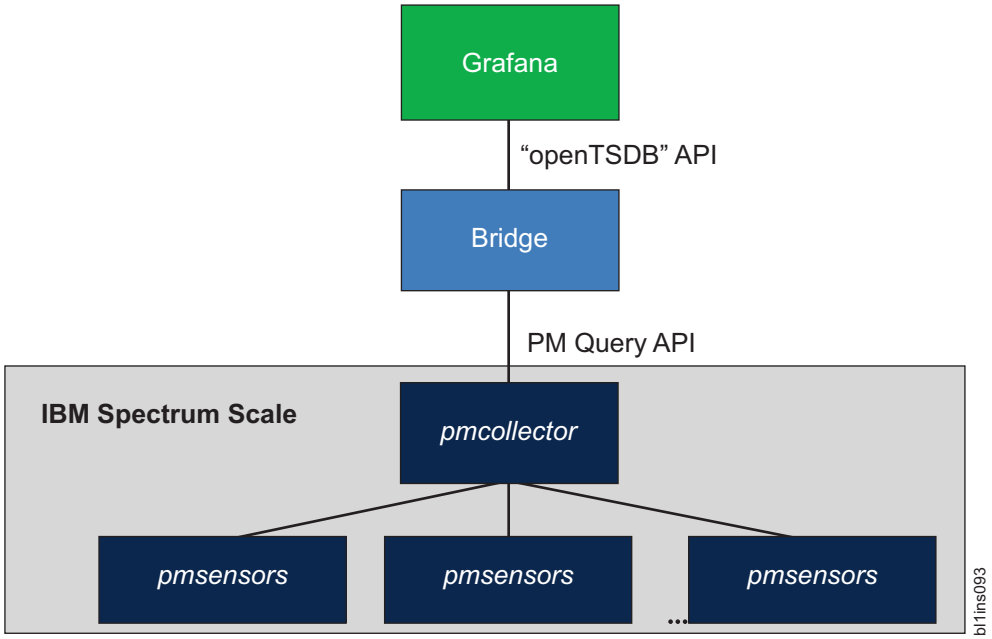

Figure 2. IBM Spectrum Scale integration framework for Grafana

Attention: The IBM Spectrum Scale performance monitoring bridge is a separate component and not a part of the IBM Spectrum Scale standard package. It can be downloaded from IBM developerWorks® Wiki. For more information on the Grafana software, see Grafana.

New features of the IBM Spectrum Scale performance monitoring bridge version 2

The IBM Spectrum Scale performance monitoring bridge version 2 has the following features:

- The IBM Spectrum Scale performance monitoring bridge version 2 is Python3 compatible.
- The bridge query format is now compatible with the OpenTSDB API versions 2.2 and 2.3 that are necessary for using Grafana's Nested Templating feature. For more information about nested templating, see Grafana's Nested Templating.
- The IBM Spectrum Scale performance monitoring bridge version 2 supports Grafana version 4.2.0 and above.
- HTTPS(SSL) connection support via port 8443 is now available. For more information, see How to setup HTTPS(SSL) connection for the IBM Spectrum Scale Performance Monitoring Bridge.
- The IBM Spectrum Scale performance monitoring bridge version 2 has a built-in logging mechanism. For more information on the built in logging mechanism, see Deep-Dive Error Analysis. For information on other troubleshooting tips and the Chrome Dev Tools options, see the Problem Determination Guide.
- New dashboard examples can be downloaded and imported from the Advanced Dashboards set package.

Note: Check the What's new page, or the `README.txt` file from the download package for a complete list of new features, changes, and bug fixes. For information about prerequisites to download the IBM Spectrum Scale performance monitoring bridge, see the Prerequisite and Download page.

Setting up IBM Spectrum Scale performance monitoring bridge for Grafana:

Follow these steps to set up the IBM Spectrum Scale performance monitoring bridge for Grafana.

The IBM Spectrum Scale system must run version 4.2.2 or above. Run the **mmlsconfig** command to view the current configuration of a GPFS cluster.

All the graphical charts that are displayed in Grafana are developed based on the performance data collected by the IBM Spectrum Scale performance monitoring tool. The performance monitoring tool packages are included in the IBM Spectrum Scale self-extracting package and get installed automatically during the IBM Spectrum Scale installation with the installation toolkit.

If you did not use the installation toolkit or disabled the performance monitoring installation during your system setup, install the performance monitoring tool manually. For more information on manually installing the performance monitoring tool, see *Manually installing the Performance Monitoring tool* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*

1. Verify that Python and CherryPy are installed on the IBM Spectrum Scale system.

   IBM Spectrum Scale Performance Monitoring Bridge is a stand-alone Python application and requires Python 2.7 or above to function properly. CherryPy is an object-oriented HTTP framework in Python, with flexible configurations.

   In order to work, the bridge needs constant access to a pmcollector. To prevent the additional network traffic, install and run the bridge code directly on a pmcollector node. In a multi-collector environment, there is no need to run the bridge on each pmcollector node separately, if they are configured in federated mode. The federation mode allows collectors to connect and collaborate with their peer collectors. If the peers are specified, any query for measurement data must be directed to any of the collectors listed in the peer definition. The chosen collector collects and assembles a response based on all relevant data from all the collectors. For more information on the performance monitoring tool, see *Performance Monitoring tool overview* in *IBM Spectrum Scale: Administration Guide*

   Note: Python and CherryPy must be downloaded for the bridge to work properly. CherryPy is not installed on any GPFS™ cluster node by default. The easiest way to set up CherryPy is described in the `ReadMe` file available with any CherryPy installation package. The IBM Spectrum Scale performance monitoring bridge version 1 and version 2 require different versions of Python and CherryPy to work properly. For information on the versions of Python and CherryPy needed for the bridge to work, see the Prerequisite and Download page.
2. Set up IBM Spectrum Scale performance monitoring bridge:
   a. Issue the following command on the pmcollector node to download and unpack the zimonGrafanaInt.tar file. The zimonGrafanaInt.tar file can be downloaded from here.

      ```
      # tar xf zimonGrafanaIntf.tar
      ```

   b. Issue the following command to run the bridge application from the directory `zimonGrafanaIntf start`:

      ```
      # python zimonGrafanaIntf.py –s < pmcollector host>
      ```

   c. If the bridge did establish the connection to the specified pmcollector and the initialization of the metadata was performed successfully, the following message is displayed at the end of line: `server starting`.
      Otherwise, check the `zserver.log` stored in the `zimonGrafanaIntf` directory. Additionally, issue the following command to check that the pmcollector service is running properly:
      `# systemctl status pmcollector`
3. Install Grafana version 2.6.1 or later.

   Note:

It is recommended to deploy Grafana 3.0.4 or later version. Download the Grafana source package from Grafana and install according to given instructions. Before you start Grafana for the first time, check the configuration options in Grafana configuration for port settings. Start the Grafana server as described on the Grafana configuration pages.

If you want to use an earlier version of Grafana (earlier than 3.0.4), the dashboard configuration described in the next step cannot be used.

4. Add the IBM Spectrum Scale bridge as a `Data Source` option to Grafana.

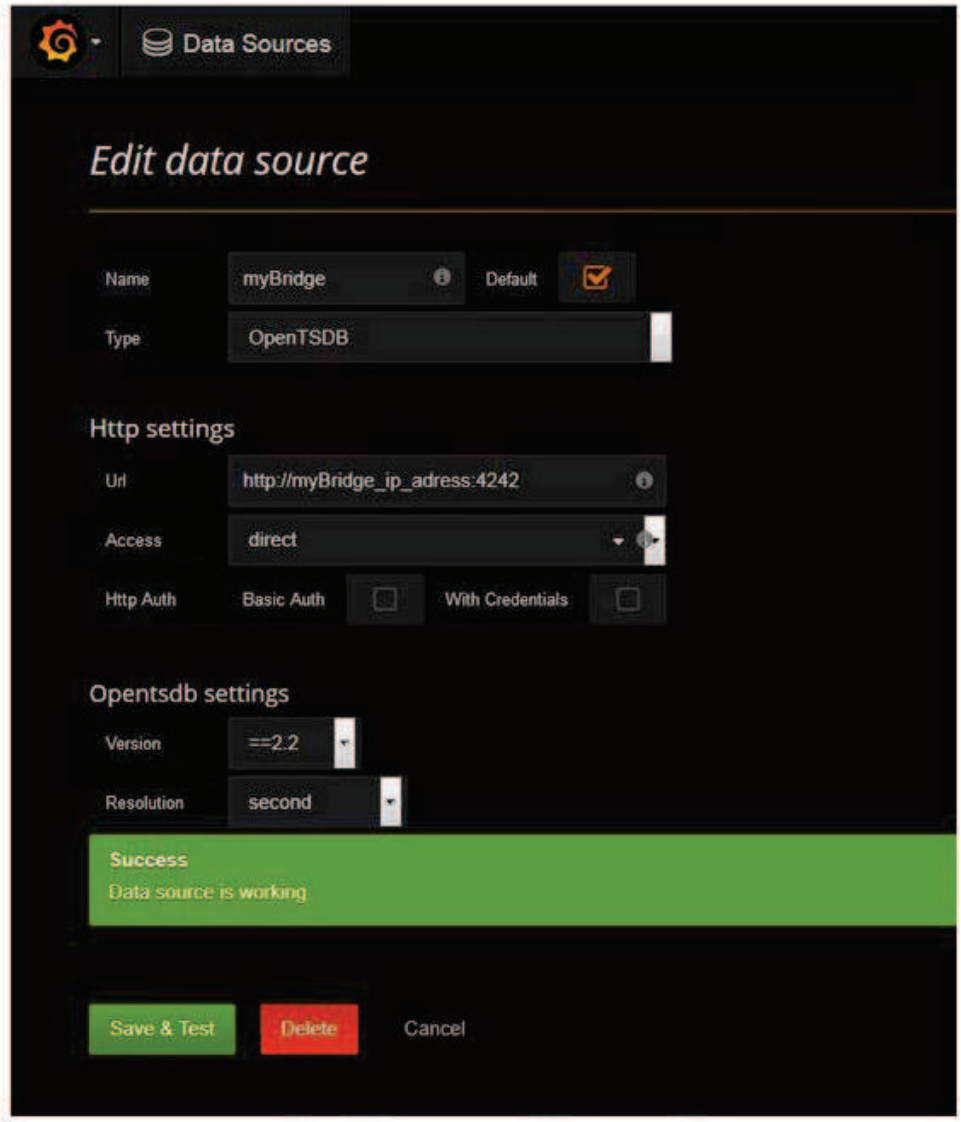

Figure 3. Adding IBM Spectrum Scale monitoring bridge as a data source

a. Click the Grafana icon on the upper left corner to view the main menu.
b. Select **Data Sources** to navigate to the data source list page.
c. Click **Add New** in the navigation bar.
d. Complete the configuration details for the **OpenTSDB** data source.

   Note:

IBM Spectrum Scale bridge listens on port 4242, and the millisecond option is not supported for **Resolution**.

e. Click **Save & Test** to ensure that the system is configured correctly.

Note: IBM Spectrum Scale performance monitoring bridge version 2 includes HTTPS(SSL) connection support via port 8443. For more information, see How to setup HTTPS(SSL) connection for the IBM Spectrum Scale Performance Monitoring Bridge.

Performance monitoring using IBM Spectrum Scale GUI

The IBM Spectrum Scale GUI provides a graphical representation of the status and historical trends of the key performance indicators. This helps the users to make decisions easily without wasting time.

The following table lists the performance monitoring options that are available in the IBM Spectrum Scale GUI.

Table 25. Performance monitoring options available in IBM Spectrum Scale GUI

Option	Function
Monitoring > **Statistics**	Displays performance of system resources and file and object storage in various performance charts. You can select the required charts and monitor the performance based on the filter criteria. The pre-defined performance widgets and metrics help in investigating every node or any particular node that is collecting the metrics.
Monitoring > **Dashboards**	Provides an easy to read and real-time user interface that shows a graphical representation of the status and historical trends of key performance indicators. This helps the users to make decisions easily without wasting time.
Nodes	Provides an easy way to monitor the performance, health status, and configuration aspects of all available nodes in the IBM Spectrum Scale cluster.
Network	Provides an easy way to monitor the performance and health status of various types of networks and network adapters.
Monitoring > **Thresholds**	Provides an option to configure and various thresholds based on the performance monitoring metrics. You can also monitor the threshold rules and the events that are associated with each rule.
Files > **File Systems**	Provides a detailed view of the performance and health aspects of file systems.
Files > **Filesets**	Provides a detailed view of the fileset performance.
Storage > **Pools**	Provides a detailed view of the performance and health aspects of storage pools.
Storage > **NSDs**	Provides a detailed view of the performance and health aspects of individual NSDs.
Files > **Transparent Cloud Tiering**	Provides insight into health, performance and configuration of the transparent cloud tiering service.
Files > **Active File Management**	Provides a detailed view of the configuration, performance, and health status of AFM cache relationship, AFM disaster recovery (AFMDR) relationship, and gateway nodes.

The **Statistics** page is used for selecting the attributes based on which the performance of the system needs to be monitored and comparing the performance based on the selected metrics. You can also mark charts as favorite charts and these charts become available for selection when you add widgets in the dashboard. You can display only two charts at a time in the **Statistics** page.

Favorite charts that are defined in the **Statistics** page and the predefined charts are available for selection in the **Dashboard**.

You can configure the system to monitor the performance of the following functional areas in the system:

- Network
- System resources
- NSD server
- IBM Spectrum Scale client
- NFS
- SMB
- Object
- CTDB
- Transparent cloud tiering. This option is available only when the cluster is configured to work with the transparent cloud tiering service.
- Waiters
- AFM

Note: The functional areas such as NFS, SMB, Object, CTDB, and Transparent cloud tiering are available only if the feature is enabled in the system.

The performance and capacity data are collected with the help of the following two components:

- **Sensor:** The sensors are placed on all the nodes and they share the data with the collector. The sensors run on any node that is required to collect metrics. Sensors are started by default only on the protocol nodes.
- **Collector:** Collects data from the sensors. The metric collector runs on a single node and gathers metrics from all the nodes that are running the associated sensors. The metrics are stored in a database on the collector node. The collector ensures aggregation of data once data gets older. The collector can run on any node in the system. By default, the collector runs on the management node. You can configure multiple collectors in the system. To configure performance monitoring through GUI, it is mandatory to configure a collector on each GUI node.

The following picture provides a graphical representation of the performance monitoring configuration for GUI.

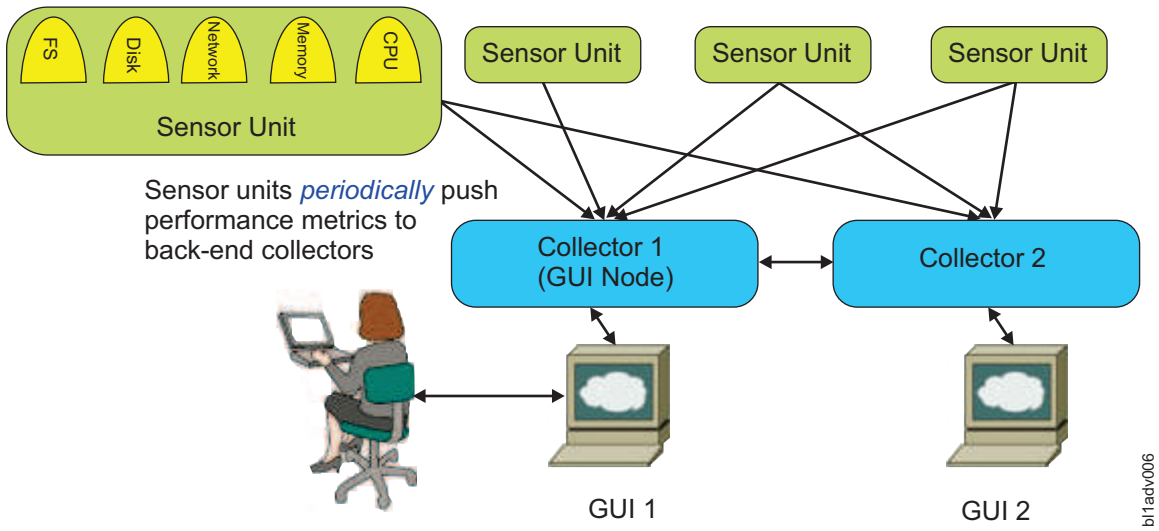

Figure 4. Performance monitoring configuration for GUI

The **mmperfmon** command can be used to query performance data through CLI, and configure the performance data collection. The GUI displays a subset of the available metrics.

Configuring performance monitoring options in GUI

You need to configure and enable the performance monitoring for GUI to view the performance data in the GUI.

Enabling performance tools in management GUI

You need to enable performance tools in the management GUI to display performance data in the management GUI. For more information on how to enable performance tools in GUI, see *Enabling performance tools in management GUI* section in the *IBM Spectrum Scale: Administration Guide*.

Verifying sensor and collector configurations

Do the following to verify whether collectors are working properly:

1. Issue **systemctl status pmcollector** on the GUI node to confirm that the collector is running. Start collector it if it is not started already.
2. If you cannot start the service, verify the log file that is located at the following location to fix the issue: `/var/log/zimon/ZIMonCollector.log`.
3. Use a sample CLI query to test if data collection works properly. For example:

   ```
   mmperfmon query cpu_user
   ```

Do the following to verify whether sensors are working properly:

1. Confirm that the sensor is configured correctly by issuing the **mmperfmon config show** command. This command lists the content of the sensor configuration that is located at the following location: `/opt/IBM/zimon/ZIMonSensors.cfg`. The configuration must point to the node where the collector is running and all the expected sensors must be enabled. An enabled sensor has a period greater than `0` in the same config file.
2. Issue **systemctl status pmsensors** to verify the status of the sensors.

Configuring performance metrics and display options in the Statistics page of the GUI

Use the **Monitoring** > **Statistics** page to monitor the performance of system resources and file and object storage. Performance of the system can be monitored by using various pre-defined charts. You can select the required charts and monitor the performance based on the filter criteria.

The pre-defined performance charts and metrics help in investigating every node or any particular node that is collecting the metrics. The following figure shows various configuration options that are available in the Statistics page of the management GUI.

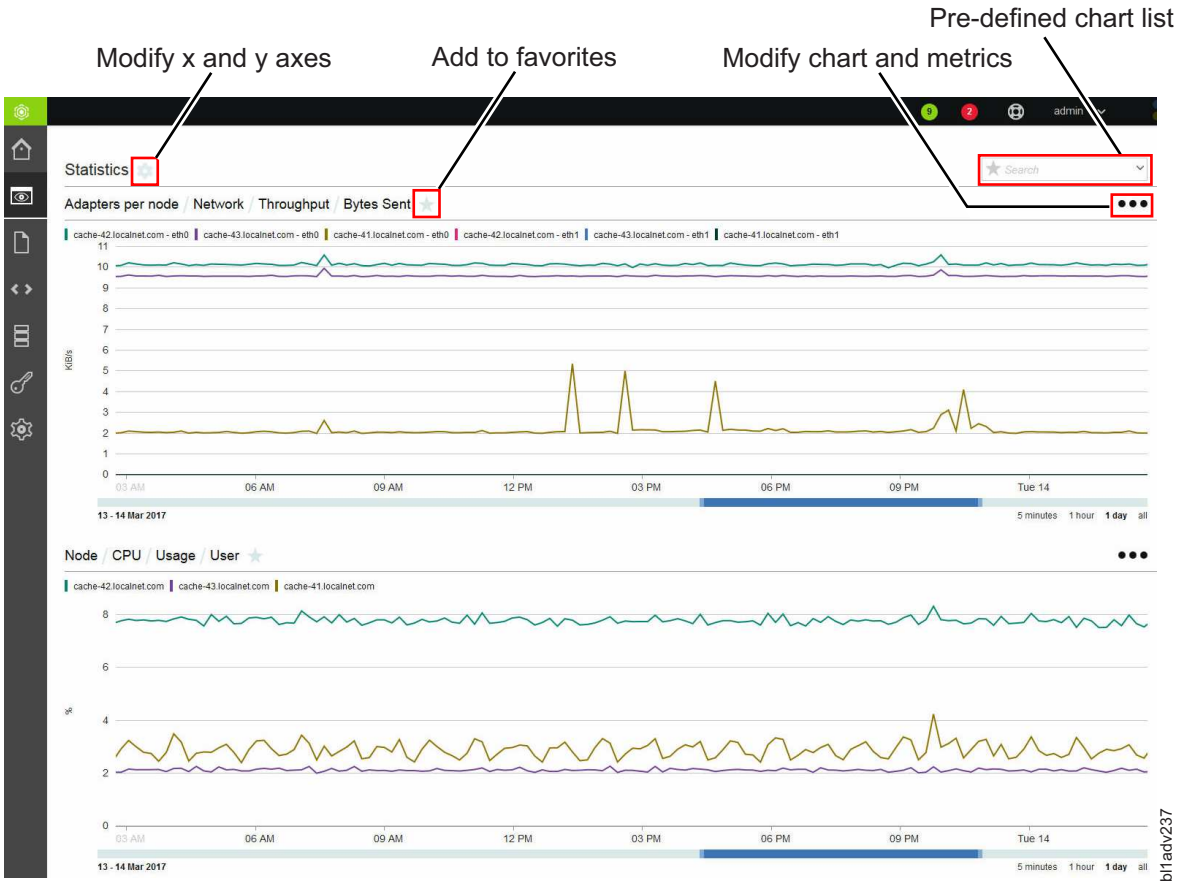

Figure 5. Statistics page in the IBM Spectrum Scale management GUI

You can select pre-defined charts that are available for selection from pre-defined chart list. You can display up to two charts at a time.

Display options in performance charts

The charting section displays the performance details based on various aspects. The GUI provides a rich set of controls to view performance charts. You can use these controls to perform the following actions on the charts that are displayed on the page:

- Zoom the chart by using the mouse wheel or resizing the timeline control. Y-axis can be automatically adjusted during zooming.
- Click and drag the chart or the timeline control at the bottom. Y-axis can be automatically adjusted during panning.

- Compare charts side by side. You can synchronize y-axis and bind x-axis. To modify the x and y axes of the chart, click the configuration symbol next to the title *Statistics* and select the required options.
- Link the timelines of the two charts together by using the display options that are available.
- The Dashboard helps to access all single graph charts, which are either predefined or custom created favorites.

Selecting performance and capacity metrics

To monitor the performance of the system, you need to select the appropriate metrics to be displayed in the performance charts. Metrics are grouped under the combination of resource types and aggregation levels. The resource types determine the area from which the data is taken to create the performance analysis and aggregation level determines the level at which the data is aggregated. The aggregation levels that are available for selection varies based on the resource type.

Sensors are configured against each resource type. The following table provides a mapping between resource types and sensors under the Performance category.

Table 26. Sensors available for each resource type

Resource type	Sensor name	Candidate nodes
Network	Network	All
System Resources	CPU	All
	Load	
	Memory	
NSD Server	GPFSNSDDisk	NSD Server nodes
IBM Spectrum Scale Client	GPFSFilesystem	IBM Spectrum Scale Client nodes
	GPFSVFS	
	GPFSFilesystemAPI	
NFS	NFSIO	Protocol nodes running NFS service
SMB	SMBStats	Protocol nodes running SMB service
	SMBGlobalStats	
Waiters	GPFSWaiters	All nodes
CTDB	CTDBStats	Protocol nodes running SMB service
Object	SwiftAccount	Protocol nodes running Object service
	SwiftContainer	
	SwiftObject	
	SwiftProxy	
AFM	GPFSAFM	All nodes
	GPFSAFMFS	
	GPFSAFMFSET	
Transparent Cloud Tiering	MCStoreGPFSStats	Cloud gateway nodes
	MCStoreIcstoreStats	
	MCStoreLWEStats	

The resource type *Waiters* are used to monitor the long running file system threads. Waiters are characterized by the purpose of the corresponding file system threads. For example, an RPC call waiter that is waiting for Network I/O threads or a waiter that is waiting for a local disk I/O file system

operation. Each waiter has a wait time associated with it and it defines how long the waiter is already waiting. With some exceptions, long waiters typically indicate that something in the system is not healthy.

The *Waiters* performance chart shows the aggregation of the total count of waiters of all nodes in the cluster above a certain threshold. Different thresholds from 100 milliseconds to 60 seconds can be selected in the list below the aggregation level. By default, the value shown in the graph is the sum of the number of waiters that exceed threshold in all nodes of the cluster at that point in time. The filter functionality can be used to display waiters data only for some selected nodes or file systems. Furthermore, there are separate metrics for different waiter types such as Local Disk I/O, Network I/O, ThCond, ThMutex, Delay, and Syscall.

You can also monitor the capacity details that are aggregated at the following levels:

- NSD
- Node
- File system
- Pool
- Fileset
- Cluster

The following table lists the sensors that are used for capturing the capacity details.

Table 27. Sensors available to capture capacity details

Sensor name	Candidate nodes
DiskFree	All nodes
GPFSFilesetQuota	Only a single node
GPFSDiskCap	Only a single node
GPFSPool	Only a single node where all GPFS file systems are mounted. The GUI does not display any values based on this sensor but it displays warnings or errors due to thresholds based on this sensor.
GPFSFileset	Only a single node. The GUI does not display any values based on this sensor but it displays warnings or errors due to thresholds based on this sensor.

You can edit an existing chart by clicking the icon that is available on the upper right corner of the performance chart and select Edit to modify the metrics selections. Do the following to drill down to the metric you are interested in:

1. Select the cluster to be monitored from the **Cluster** field. You can either select the local cluster or the remote cluster.
2. Select **Resource type**. This is the area from which the data is taken to create the performance analysis.
3. Select **Aggregation level**. The aggregation level determines the level at which the data is aggregated. The aggregation levels that are available for selection varies based on the resource type.
4. Select the entities that need to be graphed. The table lists all entities that are available for the chosen resource type and aggregation level. When a metric is selected, you can also see the selected metrics in the same grid and use methods like sorting, filtering, or adjusting the time frame to select the entities that you want to select.
5. Select **Metrics**. Metrics is the type of data that need to be included in the performance chart. The list of metrics that is available for selection varies based on the resource type and aggregation type.
6. Use the filter option to further narrow down in addition to the objects and metrics selection by using filters. Depending on the selected object category and aggregation level, the "Filter" section can be

displayed underneath the aggregation level, allowing one or more filters to be set. Filters are specified as regular expressions as shown in the following examples:

- As a single entity:
  node1
  eth0
- Filter metrics applicable to multiple nodes as shown in the following examples:
  - To select a range of nodes such as node1, node2 and node3:
    node1|node2|node3
    node[1-3]
  - To filter based on a string of text. For example, all nodes starting with 'nod' or ending with 'int':
    nod.+|.+int
  - To filter network interfaces eth0 through eth6, bond0 and eno0 through eno6:
    eth[0-6]|bond0|eno[0-6]
  - To filter nodes starting with 'strg' or 'int' and ending with 'nx':
    (strg)|(int).+nx

Creating favorite charts

Favorite charts are nothing but customized predefined charts. Favorite charts along with the predefined charts are available for selection when you add widgets in the Dashboard page.

To create favorite charts, click the ‘star’ symbol that is placed next to the chart title and enter the label.

Configuring the dashboard to view performance charts

The **Monitoring** > **Dashboard** page provides an easy to read, single page, and real-time user interface that provides a quick overview of the system performance.

The dashboard consists of several dashboard widgets and the associated favorite charts that can be displayed within a chosen layout. Currently, the following important widget types are available in the dashboard:

- Performance
- File system capacity by fileset
- System health events
- System overview
- Filesets with the largest growth rate in last week
- Timeline

The following picture highlights the configuration options that are available in the edit mode of the dashboard.

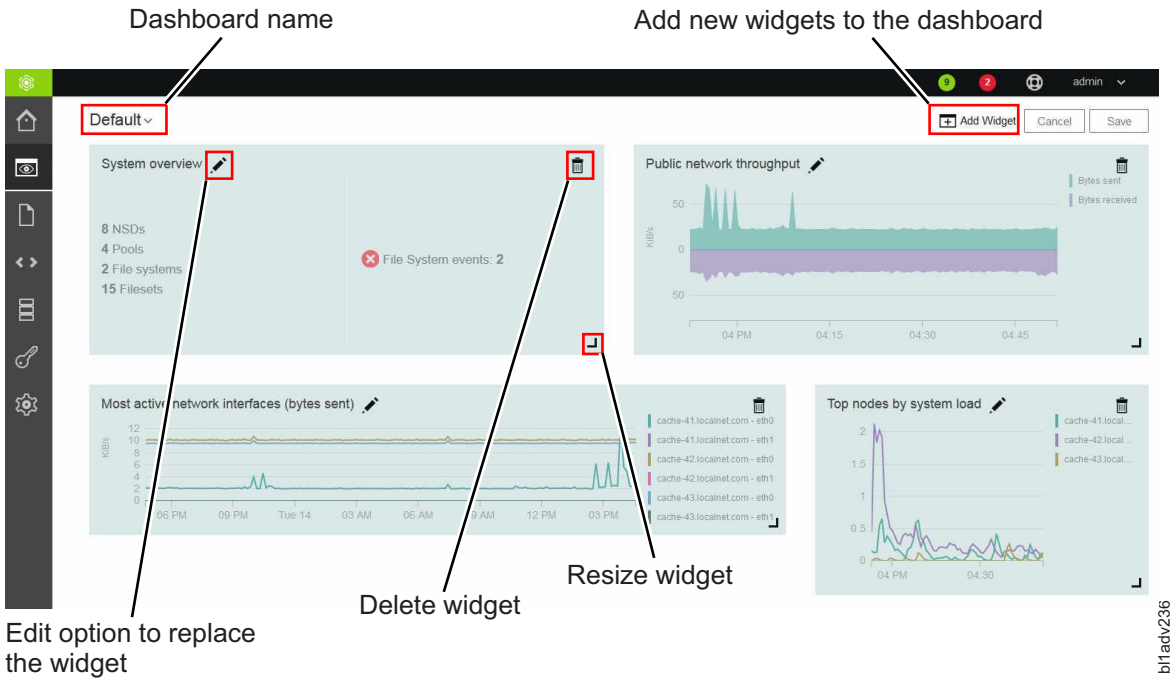

Figure 6. Dashboard page in the edit mode

Layout options

The highly customizable dashboard layout options helps to add or remove widgets and change its display options. Select **Layout Options** option from the menu that is available in the upper right corner of the Dashboard GUI page to change the layout options. While selecting the layout options, you can either select the basic layouts that are available for selection or create a new layout by selecting an empty layout as the starting point.

You can also save the dashboard so that it can be used by other users. Select **Create Dashboard** and **Delete Dashboard** options from the menu that is available in the upper right corner of the Dashboard page to create and delete dashboards respectively. If several GUIs are running by using CCR, saved dashboards are available on all nodes.

When you open the IBM Spectrum Scale GUI after the installation or upgrade, you can see the default dashboards that are shipped with the product. You can further modify or delete the default dashboards to suit your requirements.

Widget options

Several dashboard widgets can be added in the selected dashboard layout. Select **Edit Widgets** option from the menu that is available in the upper right corner of the Dashboard GUI page to edit or remove widgets in the dashboard. You can also modify the size of the widget in the edit mode. Use the **Add Widget** option that is available in the edit mode to add widgets in the dashboard.

The widgets with type *Performance* lists the charts that are marked as favorite charts in the Statistics page of the GUI. Favorite charts along with the predefined charts are available for selection when you add widgets in the dashboard.

To create favorite charts, click the 'star' symbol that is placed next to the chart title in the **Monitoring** > **Statistics** page.

Querying performance data shown in the GUI through CLI

You can query the performance data that is displayed in the GUI through the CLI. This is usually used for external system integration or to troubleshoot any issues with the performance data displayed in the GUI.

The following example shows how to query the performance data through CLI:

```
# mmperfmon query "sum(netdev_bytes_r)"
```

This query displays the following output:

```
Legend:
 1:     mr-31.localnet.com|Network|eth0|netdev_bytes_r
 2:     mr-31.localnet.com|Network|eth1|netdev_bytes_r
 3:     mr-31.localnet.com|Network|lo|netdev_bytes_r

Row           Timestamp netdev_bytes_r netdev_bytes_r netdev_bytes_r
  1 2016-03-15-14:52:09          10024
  2 2016-03-15-14:52:10           9456
  3 2016-03-15-14:52:11           9456
  4 2016-03-15-14:52:12           9456
  5 2016-03-15-14:52:13           9456
  6 2016-03-15-14:52:14           9456
  7 2016-03-15-14:52:15          27320
  8 2016-03-15-14:52:16           9456
  9 2016-03-15-14:52:17           9456
 10 2016-03-15-14:52:18          11387
```

The sensor gets the performance data for the collector and the collector passes it to the performance monitoring tool to display it in the CLI and GUI. If sensors and collectors are not enabled in the system, the system does not display the performance data and when you try to query data from a system resource, it returns an error message. For example, if performance monitoring tools are not configured properly for the resource type *Transparent Cloud Tiering*, the system displays the following output while querying the performance data:

```
mmperfmon query "sum(mcs_total_requests)" number_buckets 1
Error: No data available for query: 3169

mmperfmon: Command failed. Examine previous error messages to determine cause.
```

For more information on how to troubleshoot the performance data issues, see Chapter 23, “Performance issues,” on page 407.

Monitoring performance of nodes

The **Monitoring** > **Nodes** page provides an easy way to monitor the performance, health status, and configuration aspects of all available nodes in the IBM Spectrum Scale cluster.

The Nodes page provides the following options to analyze performance of nodes:

1. A quick view that gives the number of nodes in the system, and the overall performance of nodes based on CPU and memory usages.

   You can access this view by selecting the expand button that is placed next to the title of the page. You can close this view if not required.

   The graphs in the overview show the nodes that have the highest average performance metric over a past period. These graphs are refreshed regularly. The refresh intervals of the top three entities are depended on the displayed time frame as shown below:

   - Every minute for the 5 minutes time frame
   - Every 15 minutes for the 1 hour time frame
   - Every six hours for the 24 hours time frame
   - Every two days for the 7 days time frame

- Every seven days for the 30 days time frame
- Every four months for the 365 days time frame

2. A nodes table that displays many different performance metrics.

   To find nodes with extreme values, you can sort the values displayed in the nodes table by different performance metrics. Click the performance metric in the table header to sort the data based on that metric.

   You can select the time range that determines the averaging of the values that are displayed in the table and the time range of the charts in the overview from the time range selector, which is placed in the upper right corner. The metrics in the table do not update automatically. The refresh button above the table allows to refresh the table content with more recent data.

   You can group the nodes to be monitored based on the following criteria:
   - All nodes
   - NSD server nodes
   - Protocol nodes

3. A detailed view of the performance and health aspects of individual nodes that are listed in the Nodes page.

   Select the node for which you need to view the performance details and select **View Details**. The system displays various performance charts on the right pane.

   The detailed performance view helps to drill-down to various performance aspects. The following list provides the performance details that can be obtained from each tab of the performance view:
   - **Overview** tab provides performance chart for the following:
     - Client IOPS
     - Client data rate
     - Server data rate
     - Server IOPS
     - Network
     - CPU
     - Load
     - Memory
   - **Events** tab helps to monitor the events that are reported in the node. Three filter options are available to filter the events by their status; such as **Current Issues**, **Unread Messages**, and **All Events** displays every event, no matter if it is fixed or marked as read. Similar to the Events page, you can also perform the operations like marking events as read and running fix procedure from this events view.
   - **File Systems** tab provides performance details of the file systems mounted on the node. You can view the file system read or write throughput, average read or write transactions size, and file system read or write latency.
   - **NSDs** tab gives status of the disks that are attached to the node. The NSD tab appears only if the node is configured as an NSD server.
   - **SMB** and **NFS** tabs provide the performance details of the SMB and NFS services hosted on the node. These tabs appear in the chart only if the node is configured as a protocol node.
   - **Network** tab displays the network performance details.

Monitoring performance of file systems

The File Systems page provides an easy way to monitor the performance, health status, and configuration aspects of the all available file systems in the IBM Spectrum Scale cluster.

The following options are available to analyze the file system performance:

1. A quick view that gives the number of protocol nodes, NSD servers, and NSDs that are part of the available file systems that are mounted on the GUI server. It also provides overall capacity and total throughput details of these file systems. You can access this view by selecting the expand button that is placed next to the title of the page. You can close this view if not required.

   The graphs displayed in the quick view are refreshed regularly. The refresh intervals are depended on the displayed time frame as shown below:

   - Every minute for the 5 minutes time frame
   - Every 15 minutes for the 1 hour time frame
   - Every six hours for the 24 hours time frame
   - Every two days for the 7 days time frame
   - Every seven days for the 30 days time frame
   - Every four months for the 365 days time frame
2. A file systems table that displays many different performance metrics. To find file systems with extreme values, you can sort the values displayed in the file systems table by different performance metrics. Click the performance metric in the table header to sort the data based on that metric. You can select the time range that determines the averaging of the values that are displayed in the table and the time range of the charts in the overview from the time range selector, which is placed in the upper right corner. The metrics in the table do not update automatically. The refresh button above the table allows to refresh the table with more recent data.
3. A detailed view of the performance and health aspects of individual file systems. To see the detailed view, you can either double-click on the file system for which you need to view the details or select the file system and click **View Details**.

   The detailed performance view helps to drill-down to various performance aspects. The following list provides the performance details that can be obtained from each tab of the performance view:

   - **Overview**: Provides an overview of the file system, performance, and properties.
   - **Events**: System health events reported for the file system.
   - **NSDs**: Details of the NSDs that are part of the file system.
   - **Pools**: Details of the pools that are part of the file system.
   - **Nodes**: Details of the nodes on which the file system is mounted.
   - **Filesets**: Details of the filesets that are part of the file system.
   - **NFS**: Details of the NFS exports created in the file system.
   - **SMB**: Details of the SMB shares created in the file system.
   - **Object**: Details of the IBM Spectrum Scale object storage on the file system.

Monitoring performance of NSDs

The NSDs page provides an easy way to monitor the performance, health status, and configuration aspects of the all network shared disks (NSD) that are available in the IBM Spectrum Scale cluster.

The following options are available in the NSDs page to analyze the NSD performance:

1. An NSD table that displays the available NSDs and many different performance metrics. To find NSDs with extreme values, you can sort the values that are displayed in the table by different performance metrics. Click the performance metric in the table header to sort the data based on that metric. You can select the time range that determines the averaging of the values that are displayed in the table from the time range selector, which is placed in the upper right corner. The metrics in the table are refreshed based on the selected time frame. You can refresh it manually to see the latest data.
2. A detailed view of the performance and health aspects of individual NSDs are also available in the NSDs page. Select the NSD for which you need to view the performance details and select **View Details.** The system displays various performance charts on the right pane.

   The detailed performance view helps to drill-down to various performance aspects. The following list provides the performance details that can be obtained from each tab of the performance view:

- **Overview:** Provides an overview of the NSD performance details and related attributes.
- **Events:** System health events reported for the NSD.
- **Nodes:** Details of the nodes that serve the NSD.

Performance monitoring limitations

The following section lists the limitations of the performance monitoring tool:

- Performance monitoring is not supported by the following operating systems:
  - x86_64/sles11
  - ppc64le/Ubuntu14.04
  - ppc64/aix
- Performance monitoring tool interface for NFS does not work on SLES 11 nodes.

Chapter 2. Monitoring system health using IBM Spectrum Scale GUI

The following table lists the system health monitoring options that are available in the IBM Spectrum Scale GUI.

Table 28. System health monitoring options available in IBM Spectrum Scale GUI

Option	Function
Monitoring > **Events**	Lists the events that are reported in the system. You can monitor and troubleshoot errors on your system from the Events page.
Monitoring > **Tips**	Lists the tips reported in the system and allows to hide or show tips. The tip events give recommendations to the user to avoid certain issues that might occur in the future.
Home	Provides overall system health of the IBM Spectrum Scale system. This page is displayed in the GUI only if the minimum release level of IBM Spectrum Scale is 4.2.2 or later.
Monitoring > **Nodes**	Lists the events reported at the node level.
Files > **File Systems**	Lists the events reported at the file system level.
Files > **Transparent Cloud Tiering**	Lists the events reported for the Transparent Cloud Tiering service. The GUI displays this page only if the transparent cloud tiering feature is enabled in the system.
Files > **Filesets**	Lists events reported for filesets.
Files > **Active File Management**	Displays health status and lists events reported for AFM cache relationship, AFM disaster recovery (AFMDR) relationship, and gateway nodes.
Storage > **Pools**	Displays health status and lists events reported for storage pools.
Storage > **NSDs**	Lists the events reported at the NSD level.
Health indicator that is available in the upper right corner of the GUI.	Displays the number of events with warning and error status.
System overview widget in the **Monitoring** > **Dashboard** page.	Displays the number of events reported against each component.
System health events widget in the **Monitoring** > **Dashboard** page.	Provides an overview of the events reported in the system.
Timeline widget in the **Monitoring** > **Dashboard** page.	Displays the events that are reported in a particular time frame on the selected performance chart.

Monitoring events using GUI

You can primarily use the **Monitoring** > **Events** page to review the entire set of events that are reported in the IBM Spectrum Scale system.

The events are raised against the respective component, for example, GPFS, NFS, SMB, and so on. Same events might occur multiple times in the system. Such events are grouped together under the **Event**

Groups tab and the number of occurrences of the events are indicated in the **Occurrences** column. The **Individual Events** tab lists all the events irrespective of the multiple occurrences.

You can further filter the events listed in the Events page with the help of the following filter options:

- **Current Issues** displays all unfixed errors and warnings.
- **Unread Messages** displays all unfixed errors and warnings and information messages that are not marked as read.
- **All Events** displays every event, no matter if it is fixed or marked as read.

The status icons help to quickly determine whether the event is informational, a warning, or an error. Click an event and select **Properties** from the **Action** menu to see the detailed information of that event. The event table displays the most recent events first.

Marking events as read

You can mark certain events as read to change the status of the event in the events view. The status icons become gray in case an error or warning is fixed or if it is marked as read.

Running fix procedure

Some issues can be resolved by running a fix procedure. Use action **Run Fix Procedure** to do so. The Events page provides a recommendation for which fix procedure to run next.

For more information on how to set up event notifications, see "Set up event notifications"

Tips events

You can monitor events of type "Tips" from the **Monitoring** > **Tips** page of the GUI. The tip events give recommendations to the user to avoid certain issues that might occur in the future. The system detects the entities with tip event as healthy. A tip disappears from the GUI when the problem behind the tip event is resolved.

Select **Properties** from the **Actions** menu to view the details of the tip. After you review the tip, decide whether it requires attention or can be ignored. Select **Hide** from the **Actions** menu to ignore the events that are not important and select **Show** to mark the tips that require attention.

Set up event notifications

The system can use Simple Network Management Protocol (SNMP) traps and emails to notify you when significant events are detected. Any combination of these notification methods can be used simultaneously. Use **Settings** > **Event Notifications** page in the GUI to configure event notifications.

Notifications are normally sent immediately after an event is raised.

In email notification method, you can also define whether a recipient needs to get a report of events that are reported in the system. These reports are sent only once in a day. Based on the seriousness of the issue, each event that is reported in the system gets a severity level associated with it.

The following table describes the severity levels of event notifications.

Table 29. Notification levels

Notification level	Description
Error	Error notification is sent to indicate a problem that must be corrected as soon as possible. This notification indicates a serious problem with the system. For example, the event that is being reported might indicate a loss of redundancy in the system, and it is possible that another failure might result in loss of access to data. The most typical reason that this type of notification is because of a hardware failure, but some configuration errors or fabric errors also are included in this notification level.
Warning	A warning notification is sent to indicate a problem or unexpected condition with the system. Always immediately investigate this type of notification to determine the effect that it might have on your operation, and make any necessary corrections. Therefore, a warning notification does not require any replacement parts and it does not require IBM Support Center involvement.
Information	An informational notification is sent to indicate that an expected event is occurred. For example, a NAS service is started. No remedial action is required when these notifications are sent.

Configuring email notifications

The email feature transmits operational and error-related data in the form of an event notification email.

To configure an email server, from the Event Notifications page, select **Email Server**. Select **Edit** and then click **Enable email notifications**. Enter required details and when you are ready, click **OK**.

Email notifications can be customized by setting a custom header and footer for the emails and customizing the subject by selecting and combining from the following variables: *&message*, *&messageId*, *&severity*, *&dateAndTime*, *&cluster* and *&component*.

Emails containing the quota reports and other events reported in the following functional areas are sent to the recipients:

- AFM and AFM DR
- Authentication
- CES network
- Transparent Cloud Tiering
- NSD
- File system
- GPFS
- GUI
- Hadoop connector
- iSCSI
- Keystone
- Network
- NFS
- Object
- Performance monitoring
- SMB
- Object authentication
- Node

- CES

You can specify the severity level of events and whether to send a report that contains a summary of the events received.

To create email recipients, select **Email Recipients** from the **Event Notifications** page, and then click **Create Recipient**.

Note: You can change the email notification configuration or disable the email service at any time.

Configuring SNMP manager

Simple Network Management Protocol (SNMP) is a standard protocol for managing networks and exchanging messages. The system can send SNMP messages that notify personnel about an event. You can use an SNMP manager to view the SNMP messages that the system sends.

With an SNMP manager, such as IBM Systems Director, you can view, and act on the messages that the SNMP agent sends. The SNMP manager can send SNMP notifications, which are also known as traps, when an event occurs in the system. Select **Settings** > **Event Notifications** > **SNMP Manager** to configure SNMP managers for event notifications. You can specify up to a maximum of six SNMP managers.

In the SNMP mode of event notification, one SNMP notification (trap) with object identifiers (OID) .1.3.6.1.4.1.2.6.212.10.0.1 is sent by the GUI for each event. The following table provides the SNMP objects included in the event notifications.

Table 30. SNMP objects included in event notifications

OID	Description	Examples
.1.3.6.1.4.1.2.6.212.10.1.1	Cluster ID	317908494245422510
.1.3.6.1.4.1.2.6.212.10.1.2	Entity type	SERVER, FILESYSTEM
.1.3.6.1.4.1.2.6.212.10.1.3	Entity name	gss-11, fs01
.1.3.6.1.4.1.2.6.212.10.1.4	Component	SMB, AUTHENTICATION
.1.3.6.1.4.1.2.6.212.10.1.5	Severity	SEVERE, WARN, INFO
.1.3.6.1.4.1.2.6.212.10.1.6	Date and time	17.02.2016 13:27:42.516
.1.3.6.1.4.1.2.6.212.10.1.7	Event name	MS1014
.1.3.6.1.4.1.2.6.212.10.1.8	Message	At least one CPU of "gss-11" is failed.
1.3.6.1.4.1.2.6.212.10.1.9	Reporting node	The node where the problem is reported.

Understanding the SNMP OID ranges

The following table gives the description of the SNMP OID ranges.

Table 31. SNMP OID ranges

OID range	Description
.1.3.6.1.4.1.2.6.212	IBM Spectrum Scale
.1.3.6.1.4.1.2.6.212.10	IBM Spectrum Scale GUI
.1.3.6.1.4.1.2.6.212.10.0.1	IBM Spectrum Scale GUI event notification (trap)
.1.3.6.1.4.1.2.6.212.10.1.x	IBM Spectrum Scale GUI event notification parameters (objects)

The traps for the core IBM Spectrum Scale and those trap objects are not included in the SNMP notifications that are configured through the IBM Spectrum Scale management GUI. For more information on SNMP traps from the core IBM Spectrum Scale, see Chapter 7, “GPFS SNMP support,” on page 151

Example for SNMP traps

The following example shows the SNMP event notification that is sent when performance monitoring sensor is shut down on a node:

```
SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::enterprises.2.6.212.10.0.1
SNMPv2-SMI::enterprises.2.6.212.10.1.1 = STRING: "317908494245422510"
SNMPv2-SMI::enterprises.2.6.212.10.1.2 = STRING: "NODE"
SNMPv2-SMI::enterprises.2.6.212.10.1.3 = STRING: "gss-11"
SNMPv2-SMI::enterprises.2.6.212.10.1.4 = STRING: "PERFMON"
SNMPv2-SMI::enterprises.2.6.212.10.1.5 = STRING: "ERROR"
SNMPv2-SMI::enterprises.2.6.212.10.1.6 = STRING: "18.02.2016 12:46:44.839"
SNMPv2-SMI::enterprises.2.6.212.10.1.7 = STRING: "pmsensors_down"
SNMPv2-SMI::enterprises.2.6.212.10.1.8 = STRING: "pmsensors service should be started and is stopped"
SNMPv2-SMI::enterprises.2.6.212.10.1.9 = STRING: "gss-11"
```

The following example shows the SNMP event notification that is sent for an SNMP test message:

```
SNMPv2-MIB::snmpTrapOID.0 = OID: SNMPv2-SMI::enterprises.2.6.212.10.0.1
SNMPv2-SMI::enterprises.2.6.212.10.1.1 = STRING: "317908494245422510"
SNMPv2-SMI::enterprises.2.6.212.10.1.2 = STRING: "CLUSTER"
SNMPv2-SMI::enterprises.2.6.212.10.1.3 = STRING: "UNKNOWN"
SNMPv2-SMI::enterprises.2.6.212.10.1.4 = STRING: "GUI"
SNMPv2-SMI::enterprises.2.6.212.10.1.5 = STRING: "INFO"
SNMPv2-SMI::enterprises.2.6.212.10.1.6 = STRING: "18.02.2016 12:47:10.851"
SNMPv2-SMI::enterprises.2.6.212.10.1.7 = STRING: "snmp_test"
SNMPv2-SMI::enterprises.2.6.212.10.1.8 = STRING: "This is a SNMP test message."
SNMPv2-SMI::enterprises.2.6.212.10.1.9 = STRING: "gss-11"
```

SNMP MIBs

The SNMP Management Information Base (MIB) is a collection of definitions that define the properties of the managed objects.

The IBM Spectrum Scale GUI MIB OID range starts with 1.3.6.1.4.1.2.6.212.10. The OID range 1.3.6.1.4.1.2.6.212.10.0.1 denotes IBM Spectrum Scale GUI event notification (trap) and .1.3.6.1.4.1.2.6.212.10.1.x denotes IBM Spectrum Scale GUI event notification parameters (objects). While configuring SNMP, use the MIB file that is available at the following location of each GUI node: `/usr/lpp/mmfs/gui/IBM-SPECTRUM-SCALE-GUI-MIB.txt`.

Related concepts:

Chapter 7, “GPFS SNMP support,” on page 151
GPFS supports the use of the SNMP protocol for monitoring the status and configuration of the GPFS cluster. Using an SNMP application, the system administrator can get a detailed view of the system and be instantly notified of important events, such as a node or disk failure.

Monitoring tip events

You can monitor events of type “Tips” from the **Monitoring** > **Tips** page of the GUI. The tip events give recommendations to the user to avoid certain issues that might occur in the future. Some tips are targeted for optimization, for example, tuning the system for better performance, and certain tips help the users to fully use capabilities. For example, tips on how to enable sensors. A tip disappears from the GUI when the problem behind the tip event is resolved.

Select **Properties** from the **Actions** menu to view the details of the tip. After you review the tip, decide whether it requires attention or can be ignored. Select **Hide** from the **Actions** menu to ignore the events that are not important and select **Show** to mark the tips that require attention.

You can filter the tip events with the help of the following set of filters:

- Active Tips
- Hidden Tips
- All Tips
- Actions > Filter By Date
- Actions > Shows entries within minutes, hours, or days

Use the **Reset Date** filter option to remove the date filter.

Fix procedures are available for certain tip events. To fix such tips, click **Run Fix Procedure** option that is available under the **Action** column and follow the instructions.

Monitoring thresholds

You can configure the IBM Spectrum Scale to raise events when certain thresholds are reached. Use the **Monitoring** > **Thresholds** page to define or modify thresholds for the data that is collected through the performance monitoring sensors.

You can set the following two types of threshold levels for data collected through performance monitoring sensors:

Warning level
: When the data that is being monitored reaches the warning level, the system raises an event with severity "Warning". When the observed value exceeds the current threshold level, the system removes the warning.

Error level
: When the data that is being monitored reaches the error level, the system raises an event with severity "Error". When the observed value exceeds the current threshold level, the system removes the error state.

Certain types of thresholds are predefined in the system. The following predefined thresholds are available:

- Inode utilization at the fileset level
- Data pool capacity utilization
- Metadata pool capacity utilization
- Free memory utilization

Apart from the predefined thresholds, you can also create user-defined thresholds for the data collected through the performance monitoring sensors.

You can use the **Monitoring** > **Thresholds** page in the GUI and the **mmhealth** command to manage both predefined and user-defined thresholds.

Defining thresholds

Use the **Create Thresholds** option to define user-defined thresholds or to modify the predefined thresholds. You can use the **Use as Template** option that is available in the **Actions** menu to use an already defined threshold as the template to create a threshold. You can specify the following details in a threshold rule:

- **Metric category:** Lists all performance monitoring sensors that are enabled in the system and thresholds that are derived by performing certain calculations on certain performance metrics. These derived thresholds are referred as *measurements*. The *measurements* category provides the flexibility to edit certain predefined threshold rules. The following measurements are available for selection:

  Fileset_inode
  Inode capacity utilization at the fileset level. This is calculated as:

  ```
  (sum(gpfs_fset_allocInodes)-sum(gpfs_fset_freeInodes))/sum(gpfs_fset_maxInodes)
  ```

  DataPool_capUtil
  Data pool capacity utilization. This is calculated as:

  ```
  (sum(gpfs_pool_total_dataKB)-sum(gpfs_pool_free_dataKB))/sum(gpfs_pool_total_dataKB)
  ```

  MetaDataPool_capUtil
  Metadata pool capacity utilization. This is calculated as:

  ```
  (sum(gpfs_pool_total_metaKB)-sum(gpfs_pool_free_metaKB))/sum(gpfs_pool_total_metaKB)
  ```

  FsLatency_diskWaitRd
  File system latency for the read operations. Average disk wait time per read operation on the IBM Spectrum Scale client.

  ```
  sum(gpfs_fs_tot_disk_wait_rd)/sum(gpfs_fs_read_ops)
  ```

  FsLatency_diskWaitWr
  File system latency for the write operations. Average disk wait time per write operation on the IBM Spectrum Scale client.

  ```
  sum(gpfs_fs_tot_disk_wait_wr)/sum(gpfs_fs_write_ops)
  ```

  SMBNodeLatency_read
  SMB read latency at the node level.

  ```
  avg(op_time)/avg(op_count)
  ```

  SMBNodeLatency_write
  SMB write latency at the node level.

  ```
  avg(op_time)/avg(op_count)
  ```

  NFSNodeLatency_read
  NFS read latency at the node level.

  ```
  sum(nfs_read_lat)/sum(nfs_read_ops)
  ```

  NFSNodeLatency_write
  NFS write latency at the node level.

  ```
  sum(nfs_write_lat)/sum(nfs_write_ops)
  ```

- **Metric name:** The list of performance metrics that are available under the selected performance monitoring sensor or the measurement.
- **Name:** User-defined name of the threshold rule.
- **Filter by:** Defines the filter criteria for the threshold rule.
- **Group by:** Groups the threshold values by the selected grouping criteria. If you select a value in this field, you must select an aggregator criteria in the **Aggregator** field. By default, there is no grouping, which means that the thresholds are evaluated based on the finest available key.
- **Warning level:** Defines the threshold level for warning events to be raised for the selected metric. When the warning level is reached, the system raises an event with severity "Warning". You can customize the warning message to specify the user action that is required to fix the issue.

- **Error level:** Defines the threshold level for error events to be raised for the selected metric. When the error level is reached, the system raises an event with severity “Error”. You can customize the error message to specify the user action that is required to fix the issue.
- **Aggregator:** When grouping is selected in the **Group by** field, an aggregator must be chosen to define the aggregation function. When the **Rate** aggregator is set, the grouping is automatically set to the finest available grouping.
- **Sensitivity:** Defines the sample interval value. If a sensor is configured with interval period greater than 5 minutes, then the sensitivity is set to the same value as sensors period. The minimum value allowed is 120 seconds. If a sensor is configured with interval period less than 120 seconds, the --sensitivity is set to 120 seconds.
- **Hysteresis:** Defines the percentage of the observed value that must be under or over the current threshold level to switch back to the previous state. The default value is 0%. Hysteresis is used to avoid frequent state changes when the values are close to the threshold. The level needs to be set according to the volatility of the metric.
- **Direction:** Defines whether the events and messages are sent when the value that is being monitored exceeds or goes below the threshold level.

You can also edit and delete a threshold rule.

Threshold configuration - A scenario

The user wants to configure a threshold rule to monitor the maximum disk capacity usage. The following table shows the values against each field of the Create Threshold dialog and their respective functionality.

Table 32. Threshold rule configuration - A sample scenario

GUI fields	Value and Function
Metric Category	GPFSDiskCap Specifies that the threshold rule is going to be defined for the metrics that belong to the GPFSDiskCap sensor.
Metric name	Total Capacity The threshold rule is going to be defined to monitor the threshold levels of total capacity usage.
Name	Total capacity threshold By default, the performance monitoring metric name is used as the threshold rule name. Here, the default value is overwritten with "Total capacity threshold".
Filter by	Cluster The values are filtered at the cluster level.
Group by:	File System Groups the selected metric by file system.

Table 32. Threshold rule configuration - A sample scenario (continued)

GUI fields	Value and Function
Aggregator	Maximum When maximum capacity exceeds the threshold level, the system raises the event. If the following values are selected, the nature of the threshold rule change accordingly: • **Sum:** When the sum of the metric values exceeds the threshold levels, the system raises the events. • **Average:** When the average value exceeds the average, the system raises the events. • **Maximum:** When the maximum value exceeds maximum level, the system raises the events. • **Minimum:** When the minimum value exceeds the sum of or goes below the threshold levels, the system raises the events. • **Rate:** When the rate exceeds the threshold value, the system raises the events. Rate is only added for the "finest" group by clause. If we wanted to get a rate for a "partial key", this is not supported. That is, when **Rate** is selected, the system automatically selects the best possible values in the grouping field.
Warning level	9 GiB The system raises an event with severity Warning when the total capacity usage exceeds 9 GiB.
Error level	10 GiB The system raises an event with severity level Error when the total capacity usage exceeds 10 GiB.
Sensitivity	24 hours The threshold value is being monitored once in a day.
Hysteresis	50% If the value is reduced below 4.5 GiB, the warning state is removed. Similarly, if the value is reduced below 5 GiB, the error state is removed.
Direction	High When the value that is being monitored exceeds the threshold limit, the system raises an event.

Chapter 3. Monitoring system health by using the mmhealth command

The **mmhealth** command displays the results of the background monitoring for the health of a nodelollol and services that are hosted on the node. You can use the **mmhealth** command to view the health status of a whole cluster in a single view.

Every service hosted on an IBM Spectrum Scale node has its own health monitoring service. All the sub-components like the filesystem or network interfaces are monitored through the monitoring service of their main component. Only the sub-components of CES service such as NFS, SMB, Object, and authentication have their own health monitors. The **mmhealth** command gets the health details from these monitoring services. The role of a node in monitoring determines the components that need to be monitored. This is an internal node role and a node can have more than one role. For example, a CES node can also be a node with file systems and performance monitoring. The role of the node also determines the monitoring service that is required on a specific node. For example, you do not need a CES monitoring on a non-CES node. The monitoring services are only started if a specific node role is assigned to the node. Every monitoring service includes at least one monitor.

The following criteria must be met to use the health monitoring functionality on your GPFS cluster:

- Only Linux and AIX nodes are supported.
- Only GPFS monitoring is supported on AIX.
- The AIX nodes must have the Python 2.7.5 installed.The same is required for all operating systems, if IBM Spectrum Scale version 5.0 or later is used.
- CCR must be enabled.
- The cluster must have the minimum release level as 4.2.2.0 or higher to use **mmhealth cluster show** command.

Related concepts:

Chapter 2, “Monitoring system health using IBM Spectrum Scale GUI,” on page 99

Monitoring the health of a node

The following list provides the details of the monitoring services available in the IBM Spectrum Scale system:

1. GPFS
   - Node role: This node role is always active on all IBM Spectrum Scale nodes.
   - Tasks: Monitors all GPFS daemon-related functionalities. For example, `mmfsd` process and `gpfs port accessibility`.
2. NETWORK
   - Node role: This node role is active on every IBM Spectrum Scale node.
   - Tasks: Monitors all IBM Spectrum Scale relevant IP-based (Ethernet + IPoIB) and IB RDMA networks.
3. CES
   - Node role: This node role is active on the *CES* nodes that are listed by **mmlscluster --ces**. Once a node obtains this role, all corresponding CES sub-services are activated on that node. The CES service does not have its own monitoring service or events. The status of the CES is an aggregation of the status of its sub-services. The following sub-services are monitored::
     a. AUTH

- Tasks: Monitors LDAP, AD and or NIS-based authentication services.

b. AUTH_OBJ
   - Tasks: Monitoring the *OpenStack* identity service functionalities.

c. BLOCK
   - Tasks: Checks whether the iSCSI daemon is functioning properly.

d. CESNETWORK
   - Tasks: Monitoring CES network-related adapters and IP addresses.

e. NFS
   - Tasks: Monitoring NFS-related functionalities.

f. OBJECT
   - Tasks: Monitors the IBM Spectrum Scale for object functionality. Especially, the status of relevant system services and accessibility to ports are checked.

g. SMB
   - Tasks: Monitoring SMB-related functionality like the smbd process, the ports and ctdb processes.

4. AFM
   - Node Role: The AFM monitoring service will be active if the node is a gateway node.

     Note: To know if the node is a gateway node, run the **mmlscluster** command.
   - Tasks: Monitors the cache states and different user exit events for all the AFM fileset.
5. CLOUDGATEWAY
   - Node role: Yis identified as a Transparent cloud tiering node. All nodes listed in *mmcloudgateway* node list will get this node role.
   - Tasks: Check if the cloud gateway service functions as expected.
6. DISK
   - Node role: Nodes with node class *nsdNodes* will monitor the DISK service. IBM Spectrum Scale nodes.
   - Tasks: Checking, if IBM Spectrum Scale disks are available and running.
7. FILESYSTEM
   - Node role: This node role is active on all IBM Spectrum Scale nodes.
   - Tasks: Monitors different aspects of IBM Spectrum Scale file systems.
8. GUI
   - Node role: Nodes with node class *GUI_MGMT_SERVERS* will monitor the GUI service.
   - Tasks: Verifies whether the GUI services are functioning properly.
9. HADOOPCONNECTOR
   - Node role: Nodes where the Hadoop service is configured get the Hadoop connector node role.
   - Tasks: Monitors the Hadoop data node and name node services.
10. PERFMON
    - Node role: Nodes where *PerfmonSensors* or *PerfmonCollector* services are running get the *PERFMON* node role. *PerfmonSensors* are determined through the perfmon designation in **mmlscluster**. *PerfmonCollector* are determined through the *colCandidates* line in the configuration file.
    - Tasks: Monitors whether *PerfmonSensors* and *PerfmonCollector* are running as expected.
11. THRESHOLD
    - Node role: Nodes where the performance data collection is configured and enabled. If a node role is not configured to *PERFMON*, it cannot have a *THRESHOLD* role either.
    - Tasks: Monitors whether the node-related thresholds rules evaluation is running as expected, and if the health status has changed as a result of the threshold limits being crossed.

Note: The THRESHOLD service is available only when the cluster belongs to IBM Spectrum Scale version 4.2.3 or later. In a mixed environment with a cluster containing some nodes belonging to IBM Spectrum Scale version 4.2.2 and some nodes belonging to IBM Spectrum Scale version 4.2.3, the overall cluster version is 4.2.2. The threshold service is unavailable in such an environment.

12. MSGQUEUE
    - Node role: A node gets the MSGQUEUE node role if it monitors nodes included in the `kafkaBrokerServers` node class or the `kafkaZookeeperServers` node class.
    - Tasks: Monitors the Zookeeper and Kafka service on the Kafka broker servers and the Kafka zookeeper servers.
13. FILEAUDITLOG
    - Node role: A node gets the FILEAUDITLOG node role if the node is part of the `kafkaBrokerServers` node class.
    - Tasks: Monitors the File Audit Log consumer process of each filesystem that has file audit login enabled, and detects consumer related errors .
14. CESIP
    - Node role: A cluster manager node, which can be detected using the `mmlsmgr -c` command. This node runs a special code module of the monitor, which checks the cluster-wide CES IP distribution.
    - Tasks: Compares the effectively hosted CES IPs with the list of declared CES IPs in the address pool, and report the result. There are three cases:
      – All declared CES IPs are hosted. In this case, the state of the IPs is `HEALTHY`.
      – None of the declared CES IPs are hosted. In this case, the state of the IPs is `FAILED`.
      – Only a subset of the declared CES IPs are hosted . In this case, the state of the IPs is `DEGRADED`.

    Note: A `FAILED` state does not trigger any failover.

For more details on different events, their causes and possible user actions to resolves them, see "Events" on page 473.

Running a user-defined script when an event is raised

The root user has the option to create a callback script that is triggered when an event is raised.

This callback script will be called every time for a new event, irrespective of whether it is an information event or a state-changing event. For state-changing events the script will be triggered when the event becomes active. For more information about the different events, see "Event type and monitoring status for system health" on page 112.

When the script is created, ensure that:

- The script must to be saved in the `/var/mmfs/etc/` location, and must be named `eventsCallback`.
- The script is created by a root user. The script will only be called if it is created by the root user. The uid of the root user is 0.

The script will be called with following space-separated arguments list:
version date time timezone event_name component_name identifier severity_abbreviation state_abbreviation message event_arguments where:

version
The version argument displays the version number of the script. The version is set to1 by default. This value is incremented by 1 for every change in the callout's format or functionality. In the sample output, the version is 1.

date The date is in the yyyy-mm-dd format. In the sample output, the date is `2018-02-23`.

time The time is in the hour:minutes:seconds.milliseconds format. In the sample output, the time is `00:01:07.499834`.

timezone
The timezone is based on the server settings. In the sample output, the timezone is `EST`.

event_name
The event_name argument displays the name of the event. In the sample output, the event_name is `pmsensors_up`.

component_name
The component_name argument displays the name of the reporting component. In the sample output, the component_name is `perfmon`.

identifier
The identifier argument displays the entity's name if the `require_unique` value for the event is set to `True`. If there is no identifier, '' will be returned. In the sample output, the identifier is `''`.

severity_abbreviation
The severity_abbreviation argument usually displays the first letter of Info, TIP, Warning or Error. In the sample output, the severity_abbreviation is `I`.

message
The message is framed with a single quotation mark (' '). Single ticks in the message are encoded to \\'. The `event_arguments` are already included in the message, so parsing from the customer is not needed. In the sample output, the message is `pmsensors service as expected, state is started`.

event_arguments
The event_arguments is framed with single quotation mark (' '). Single ticks in the arguments are encoded to \\'. Arguments are already included in the message. The arguments are comma separated, and without a space. Arguments are %-encoded, so the following characters will be encoded into their hexadecimal unicode representative starting with a %-character:

\t\n !\"#$%&'()*+,-./:;<=>?@[\\]^_`{|}~

For example, % is encoded to %25, and # is encoded to %23.

A sample output of the script is displayed:

```
/var/mmfs/etc/eventsCallback 1 2018-02-23 00:01:07.499834 EST pmsensors_up perfmon
 '' I H 'pmsensors service as expected, state is started' 'started'
```

Important: The script is started synchronously within the monitoring cycle, therefore it must be lightweight and return a value quickly. The recommended runtime is less than 1second. Long running scripts will be detected, logged and killed. The script has a hard timeout of 60 seconds.

Event type and monitoring status for system health

An event might trigger a change in the state of a system.

The following three types of events are reported in the system:

- State-changing events: These events change the state of a component or entity from good to bad or from bad to good depending on the corresponding state of the event.

  Note: An event is raised when the health status of the component goes from good to bad. For example, an event is raised that changes the status of a component from `HEALTHY` to `DEGRADED`. However, if the state was already `DEGRADED` based on another active event, there will be no change in the status of the component. Also if the state of the entity was `FAILED`, a `DEGRADED` event wouldn't change the component's state, because a `FAILED` status is more dominant than the `DEGRADED` status.

- Tip: These are similar to state-changing events, but can be hidden by the user. Like state-changing events, a tip is removed automatically if the problem is resolved. A tip event always changes the state to of a component from HEALTHY to TIPS if the event is not hidden.

  Note: If the state of a component changes to TIPS, it can be hidden. However, you can still view the active hidden events using the **mmhealth show ComponentName --verbose** command, if the cause for the event still exists.
- Information events: These are short notification events that will only be shown in the event log, but do not change the state of the components.

The monitoring interval is between 15 and 30 seconds, depending on the component. However, there are services that are monitored less often (e.g. once per 30 minutes) to save system resources. You can find more information about the events from the **Monitoring** > **Events** page in the IBM Spectrum Scale GUI or by issuing the **mmhealth event show** command.

The following are the possible status of nodes and services:

- UNKNOWN - Status of the node or the service hosted on the node is not known.
- HEALTHY - The node or the service hosted on the node is working as expected. There are no active error events.
- CHECKING - The monitoring of a service or a component hosted on the node is starting at the moment. This state is a transient state and is updated when the startup is completed.
- TIPS - There might be an issue with the configuration and tuning of the components. This status is only assigned to a tip event
- DEGRADED - The node or the service hosted on the node is not working as expected. That is, a problem occurred with the component but it did not result in a complete failure.
- FAILED - The node or the service hosted on the node failed due to errors or cannot be reached anymore.
- DEPEND - The node or the services hosted on the node have failed due to the failure of some components. For example, an NFS or SMB service shows this status if authentication has failed.

The status are graded as follows: HEALTHY < TIPS < DEGRADED < FAILED. For example, the status of the service hosted on a node becomes FAILED if there is at least one active event in the FAILED status for that corresponding service. The FAILED status gets more priority than the DEGRADED which is followed by TIPS and then HEALTHY, while setting the status of the service. That is, if a service has an active event with a HEALTHY status and another active event with a FAILED status, then the system sets the status of the service as FAILED.

Some directed maintenance procedures or DMPs are available to solve issues caused by tip events. For information on DMPs, see “Directed maintenance procedures for tip events” on page 451.

Threshold monitoring for system health

Threshold monitoring pre-requisites

If you did not use the IBM Spectrum Scale installation toolkit or disabled the performance monitoring installation during your system setup (**./spectrumscale config perfmon -r off**), please make sure your system meets the following configuration requirements:

- IBM Spectrum Scale version 4.2.2 or later(on all nodes).
- PMSensors and PMCollectors must be on version 4.2.2 or later.
- CCR must be enabled on the cluster.
- GPFSPool and GPFSFileset sensors are enabled automatically, when all above requirements are met.

The available filesystem available capacity depends on the fullness of its fileset-inode spaces, capacity usage, and memory utilization in each data or metadata pool. Therefore, the predefined capacity threshold limit for a filesystem is broken down to the thresholds rules of:

- Fileset-inode spaces
- Data pool capacity
- Metadata pool capacity
- Memory free utilization

The violation of any rule results in the parent filesystem receiving a capacity issue notification. The pmsensors such as `GPFSPool` and `GPFSFileset` are activated automatically and bound to the first collector node, and tracks the inode and pool space usage of the filesystem. For more information on pmsensors, see "Configuring the performance monitoring tool" on page 46. For a new filesystem, the process can be slow and can be improved by restarting sensors on the first collector node.

For capacity utilization rules, the warn level is set to 80%, and the error level to 90%. For memory utilization rule, the warn level is set to 100 MB, and the error level to 50 MB. The metrics value are frequently compared with rules boundaries by internal monitor process. As soon as one of the metric values exceeds their threshold limit, the system health daemon receives an event notification from monitoring process and generates log event and updates the health status of the filesystem having capacity problems.

Thresholds monitoring known limitations

The filesystem health status change may not get updated in the following situations:

1. The pool or fileset capacity utilization returned from error range to warn range.
2. If pools or inode spaces (independent filesets) have been removed (workaround: The status will be automatically updated with the next restart of the monitoring component on the collector node).

New features for threshold monitoring

Starting with version 4.2.3, the predefined thresholds rules are extended with a new threshold rule monitoring "memory free" utilization on cluster nodes. IBM Spectrum Scale user can also delete or add any or all of the existing thresholds rules.

Starting from IBM Spectrum Scale version 5.0, if multiple thresholds rules have overlapping entities for the same metrics, only one of the concurrent rules is made actively eligible. All rules for the same metric will get a priority rank number, where one is the highest number. This rank is based on a given metric's maximum number of filtering levels and the filter granularity specified in the rule. As a result, a rule that monitors a specific entity or a set of entities becomes high priority, and performs entity thresholds evaluation and status update for this particular entity or a set of entities. This implies that a less specific rule, like the one which is valid for all entities, is disabled for this particular entity or set of entities. For example, a threshold rule which applies to a single file system will take precedence over a rule which applies to several or all the file systems. For more information, see "Use case 3: Creating threshold rules for specific filesets." on page 125

Related concepts:

Chapter 2, "Monitoring system health using IBM Spectrum Scale GUI," on page 99

System health monitoring use cases

The following sections describe the use case for the **mmhealth** command

Use case 1: Checking the health status of the nodes and their corresponding services by using the following commands:

1. To show the health status of the current node:

   ```
   mmhealth node show
   ```

   The system displays output similar to this:

   ```
   Node name:      test_node
   Node status:    HEALTHY
   Status Change:  39 min. ago

   Component          Status        Status Change    Reasons
   ------------------------------------------------------------------
   GPFS               HEALTHY       39 min. ago       -
   NETWORK            HEALTHY       40 min. ago       -
   FILESYSTEM         HEALTHY       39 min. ago       -
   DISK               HEALTHY       39 min. ago       -
   CES                HEALTHY       39 min. ago       -
   PERFMON            HEALTHY       40 min. ago       -
   ```

2. To view the health status of a specific node, issue this command:

   ```
   mmhealth node show -N test_node2
   ```

   The system displays output similar to this:

   ```
   Node name:      test_node2
   Node status:    CHECKING
   Status Change:  Now

   Component       Status        Status Change    Reasons
   ------------------------------------------------------------------
   GPFS            CHECKING      Now              -
   NETWORK         HEALTHY       Now              -
   FILESYSTEM      CHECKING      Now              -
   DISK            CHECKING      Now              -
   CES             CHECKING      Now              -
   PERFMON         HEALTHY       Now              -
   ```

3. To view the health status of all the nodes, issue this command:

   ```
   mmhealth node show -N all
   ```

   The system displays output similar to this:

   ```
   Node name:    test_node
   Node status:  DEGRADED

   Component           Status        Status Change     Reasons
   -------------------------------------------------------------
   GPFS                HEALTHY          Now             -
   CES                 FAILED           Now             smbd_down
   FileSystem          HEALTHY          Now             -

   Node name:            test_node2
   Node status:          HEALTHY

   Component           Status        Status Change    Reasons
   ------------------------------------------------------------
   GPFS                HEALTHY       Now              -
   CES                 HEALTHY       Now              -
   FileSystem          HEALTHY       Now              -
   ```

4. To view the detailed health status of the component and its sub-component, issue this command:

   ```
   mmhealth node show ces
   ```

   The system displays output similar to this:

```
Node name:      test_node

Component       Status        Status Change    Reasons
-----------------------------------------------------------------
CES             HEALTHY       2 min. ago       -
  AUTH          DISABLED      2 min. ago       -
  AUTH_OBJ      DISABLED      2 min. ago       -
  BLOCK         DISABLED      2 min. ago       -
  CESNETWORK    HEALTHY       2 min. ago       -
  NFS           HEALTHY       2 min. ago       -
  OBJECT        DISABLED      2 min. ago       -
  SMB           HEALTHY       2 min. ago       -
```

5. To view the health status of only unhealthy components, issue this command:
   ```
   mmhealth node show --unhealthy
   ```

   The system displays output similar to this:
   ```
   Node name:        test_node
   Node status:       FAILED
   Status Change:  1 min. ago

   Component       Status        Status Change    Reasons
   ---------------------------------------------------------------
   GPFS            FAILED        1 min. ago       gpfs_down, quorum_down
   FILESYSTEM      DEPEND        1 min. ago       unmounted_fs_check
   CES             DEPEND        1 min. ago       ces_network_ips_down, nfs_in_grace
   ```
6. To view the health status of sub-components of a node's component, issue this command:
   ```
   mmhealth node show --verbose
   ```

   The system displays output similar to this:
   ```
   Node name:      gssio1-hs.gpfs.net
   Node status:    HEALTHY

   Component                                 Status             Reasons
   ------------------------------------------------------------------
   GPFS                                      DEGRADED           -
   NETWORK                                   HEALTHY            -
     bond0                                   HEALTHY            -
     ib0                                     HEALTHY            -
     ib1                                     HEALTHY            -
   FILESYSTEM                                DEGRADED           stale_mount, stale_mount, stale_mount
     Basic1                                  FAILED             stale_mount
     Basic2                                  FAILED             stale_mount
     Custom1                                 HEALTHY            -
     gpfs0                                   FAILED             stale_mount
     gpfs1                                   FAILED             stale_mount
   DISK                                      DEGRADED           disk_down
     rg_gssio1_hs_Basic1_data_0              HEALTHY            -
     rg_gssio1_hs_Basic1_system_0            HEALTHY            -
     rg_gssio1_hs_Basic2_data_0              HEALTHY            -
     rg_gssio1_hs_Basic2_system_0            HEALTHY            -
     rg_gssio1_hs_Custom1_data1_0            HEALTHY            -
     rg_gssio1_hs_Custom1_system_0           DEGRADED           disk_down
     rg_gssio1_hs_Data_8M_2p_1_gpfs0         HEALTHY            -
     rg_gssio1_hs_Data_8M_3p_1_gpfs1         HEALTHY            -
     rg_gssio1_hs_MetaData_1M_3W_1_gpfs0     HEALTHY            -
     rg_gssio1_hs_MetaData_1M_4W_1_gpfs1     HEALTHY            -
     rg_gssio2_hs_Basic1_data_0              HEALTHY            -
     rg_gssio2_hs_Basic1_system_0            HEALTHY            -
     rg_gssio2_hs_Basic2_data_0              HEALTHY            -
     rg_gssio2_hs_Basic2_system_0            HEALTHY            -
     rg_gssio2_hs_Custom1_data1_0            HEALTHY            -
     rg_gssio2_hs_Custom1_system_0           HEALTHY            -
     rg_gssio2_hs_Data_8M_2p_1_gpfs0         HEALTHY            -
     rg_gssio2_hs_Data_8M_3p_1_gpfs1         HEALTHY            -
     rg_gssio2_hs_MetaData_1M_3W_1_gpfs0     HEALTHY            -
     rg_gssio2_hs_MetaData_1M_4W_1_gpfs1     HEALTHY            -
   NATIVE_RAID                               DEGRADED           gnr_pdisk_replaceable, gnr_rg_failed, enclosure_needsservice
     ARRAY                                   DEGRADED           -
       rg_gssio2-hs/DA1                      HEALTHY            -
       rg_gssio2-hs/DA2                      HEALTHY            -
       rg_gssio2-hs/NVR                      HEALTHY            -
       rg_gssio2-hs/SSD                      HEALTHY            -
     ENCLOSURE                               DEGRADED           enclosure_needsservice
       SV52122944                            DEGRADED           enclosure_needsservice
   ```

```
    SV53058375                          HEALTHY           -
  PHYSICALDISK                          DEGRADED          gnr_pdisk_replaceable
    rg_gssio2-hs/e1d1s01                FAILED            gnr_pdisk_replaceable
    rg_gssio2-hs/e1d1s07                HEALTHY           -
    rg_gssio2-hs/e1d1s08                HEALTHY           -
    rg_gssio2-hs/e1d1s09                HEALTHY           -
    rg_gssio2-hs/e1d1s10                HEALTHY           -
    rg_gssio2-hs/e1d1s11                HEALTHY           -
    rg_gssio2-hs/e1d1s12                HEALTHY           -
    rg_gssio2-hs/e1d2s07                HEALTHY           -
    rg_gssio2-hs/e1d2s08                HEALTHY           -
    rg_gssio2-hs/e1d2s09                HEALTHY           -
    rg_gssio2-hs/e1d2s10                HEALTHY           -
    rg_gssio2-hs/e1d2s11                HEALTHY           -
    rg_gssio2-hs/e1d2s12                HEALTHY           -
    rg_gssio2-hs/e1d3s07                HEALTHY           -
    rg_gssio2-hs/e1d3s08                HEALTHY           -
    rg_gssio2-hs/e1d3s09                HEALTHY           -
    rg_gssio2-hs/e1d3s10                HEALTHY           -
    rg_gssio2-hs/e1d3s11                HEALTHY           -
    rg_gssio2-hs/e1d3s12                HEALTHY           -
    rg_gssio2-hs/e1d4s07                HEALTHY           -
    rg_gssio2-hs/e1d4s08                HEALTHY           -
    rg_gssio2-hs/e1d4s09                HEALTHY           -
    rg_gssio2-hs/e1d4s10                HEALTHY           -
    rg_gssio2-hs/e1d4s11                HEALTHY           -
    rg_gssio2-hs/e1d4s12                HEALTHY           -
    rg_gssio2-hs/e1d5s07                HEALTHY           -
    rg_gssio2-hs/e1d5s08                HEALTHY           -
    rg_gssio2-hs/e1d5s09                HEALTHY           -
    rg_gssio2-hs/e1d5s10                HEALTHY           -
    rg_gssio2-hs/e1d5s11                HEALTHY           -
    rg_gssio2-hs/e2d1s07                HEALTHY           -
    rg_gssio2-hs/e2d1s08                HEALTHY           -
    rg_gssio2-hs/e2d1s09                HEALTHY           -
    rg_gssio2-hs/e2d1s10                HEALTHY           -
    rg_gssio2-hs/e2d1s11                HEALTHY           -
    rg_gssio2-hs/e2d1s12                HEALTHY           -
    rg_gssio2-hs/e2d2s07                HEALTHY           -
    rg_gssio2-hs/e2d2s08                HEALTHY           -
    rg_gssio2-hs/e2d2s09                HEALTHY           -
    rg_gssio2-hs/e2d2s10                HEALTHY           -
    rg_gssio2-hs/e2d2s11                HEALTHY           -
    rg_gssio2-hs/e2d2s12                HEALTHY           -
    rg_gssio2-hs/e2d3s07                HEALTHY           -
    rg_gssio2-hs/e2d3s08                HEALTHY           -
    rg_gssio2-hs/e2d3s09                HEALTHY           -
    rg_gssio2-hs/e2d3s10                HEALTHY           -
    rg_gssio2-hs/e2d3s11                HEALTHY           -
    rg_gssio2-hs/e2d3s12                HEALTHY           -
    rg_gssio2-hs/e2d4s07                HEALTHY           -
    rg_gssio2-hs/e2d4s08                HEALTHY           -
    rg_gssio2-hs/e2d4s09                HEALTHY           -
    rg_gssio2-hs/e2d4s10                HEALTHY           -
    rg_gssio2-hs/e2d4s11                HEALTHY           -
    rg_gssio2-hs/e2d4s12                HEALTHY           -
    rg_gssio2-hs/e2d5s07                HEALTHY           -
    rg_gssio2-hs/e2d5s08                HEALTHY           -
    rg_gssio2-hs/e2d5s09                HEALTHY           -
    rg_gssio2-hs/e2d5s10                HEALTHY           -
    rg_gssio2-hs/e2d5s11                HEALTHY           -
    rg_gssio2-hs/e2d5s12ssd             HEALTHY           -
    rg_gssio2-hs/n1s02                  HEALTHY           -
    rg_gssio2-hs/n2s02                  HEALTHY           -
  RECOVERYGROUP                         DEGRADED          gnr_rg_failed
    rg_gssio1-hs                        FAILED            gnr_rg_failed
    rg_gssio2-hs                        HEALTHY           -
  VIRTUALDISK                           DEGRADED          -
    rg_gssio2_hs_Basic1_data_0          HEALTHY           -
    rg_gssio2_hs_Basic1_system_0        HEALTHY           -
    rg_gssio2_hs_Basic2_data_0          HEALTHY           -
    rg_gssio2_hs_Basic2_system_0        HEALTHY           -
    rg_gssio2_hs_Custom1_data1_0        HEALTHY           -
    rg_gssio2_hs_Custom1_system_0       HEALTHY           -
    rg_gssio2_hs_Data_8M_2p_1_gpfs0     HEALTHY           -
    rg_gssio2_hs_Data_8M_3p_1_gpfs1     HEALTHY           -
    rg_gssio2_hs_MetaData_1M_3W_1_gpfs0 HEALTHY           -
    rg_gssio2_hs_MetaData_1M_4W_1_gpfs1 HEALTHY           -
    rg_gssio2_hs_loghome                HEALTHY           -
    rg_gssio2_hs_logtip                 HEALTHY           -
    rg_gssio2_hs_logtipbackup           HEALTHY           -
  PERFMON                               HEALTHY           -
```

7. To view the eventlog history of the node for the last hour, issue this command:

```
mmhealth node eventlog --hour
```

The system displays output similar to this:

```
Node name:      test-21.localnet.com
Timestamp                                 Event Name               Severity   Details
2016-10-28 06:59:34.045980 CEST       monitor_started          INFO       The IBM Spectrum Scale monitoring service has been started
2016-10-28 07:01:21.919943 CEST       fs_remount_mount         INFO       The filesystem objfs was mounted internal
2016-10-28 07:01:32.434703 CEST       disk_found               INFO       The disk disk1 was found
2016-10-28 07:01:32.669125 CEST       disk_found               INFO       The disk disk8 was found
2016-10-28 07:01:36.975902 CEST       filesystem_found         INFO       Filesystem objfs was found
2016-10-28 07:01:37.226157 CEST       unmounted_fs_check       WARNING    The filesystem objfs is probably needed, but not mounted
2016-10-28 07:01:52.113691 CEST       mounted_fs_check         INFO       The filesystem objfs is mounted
2016-10-28 07:01:52.283545 CEST       fs_remount_mount         INFO       The filesystem objfs was mounted normal
2016-10-28 07:02:07.026093 CEST       mounted_fs_check         INFO       The filesystem objfs is mounted
2016-10-28 07:14:58.498854 CEST       ces_network_ips_down     WARNING    No CES relevant NICs detected
2016-10-28 07:15:07.702351 CEST       nodestatechange_info     INFO       A CES node state change: Node 1 add startup flag
2016-10-28 07:15:37.322997 CEST       nodestatechange_info     INFO       A CES node state change: Node 1 remove startup flag
2016-10-28 07:15:43.741149 CEST       ces_network_ips_up       INFO       CES-relevant IPs are served by found NICs
2016-10-28 07:15:44.028031 CEST       ces_network_vanished     INFO       CES NIC eth0 has vanished
```

8. To view the eventlog history of the node for the last hour, issue this command:

   ```
   mmhealth node eventlog --hour --verbose
   ```

   The system displays output similar to this:

   ```
   Node name:      test-21.localnet.com
   Timestamp                       Component    Event Name           Event ID Severity Details
   2016-10-28 06:59:34.045980 CEST   gpfs         monitor_started       999726   INFO    The IBM Spectrum Scale monitoring service has been started
   2016-10-28 07:01:21.919943 CEST   filesystem   fs_remount_mount      999306   INFO    The filesystem objfs was mounted internal
   2016-10-28 07:01:32.434703 CEST   disk         disk_found            999424   INFO    The disk disk1 was found
   2016-10-28 07:01:32.669125 CEST   disk         disk_found            999424   INFO    The disk disk8 was found
   2016-10-28 07:01:36.975902 CEST   filesystem   filesystem_found      999299   INFO    Filesystem objfs was found
   2016-10-28 07:01:37.226157 CEST   filesystem   unmounted_fs_check    999298   WARNING The filesystem objfs is probably needed, but not mounted
   2016-10-28 07:01:52.113691 CEST   filesystem   mounted_fs_check      999301   INFO    The filesystem objfs is mounted
   2016-10-28 07:01:52.283545 CEST   filesystem   fs_remount_mount      999306   INFO    The filesystem objfs was mounted normal
   2016-10-28 07:02:07.026093 CEST   filesystem   mounted_fs_check      999301   INFO    The filesystem objfs is mounted
   2016-10-28 07:14:58.498854 CEST   cesnetwork   ces_network_ips_down  999426   WARNING No CES relevant NICs detected
   2016-10-28 07:15:07.702351 CEST   gpfs         nodestatechange_info  999220   INFO    A CES node state change: Node 1 add startup flag
   2016-10-28 07:15:37.322997 CEST   gpfs         nodestatechange_info  999220   INFO    A CES node state change: Node 1 remove startup flag
   2016-10-28 07:15:43.741149 CEST   cesnetwork   ces_network_ips_up    999427   INFO    CES-relevant IPs are served by found NICs
   2016-10-28 07:15:44.028031 CEST   cesnetwork   ces_network_vanished  999434   INFO    CES NIC eth0 has vanished
   ```

9. To view the detailed description of an event, issue **mmhealth event show** command. This is an example for *quorum_down* event:

   ```
   mmhealth event show quorum_down
   ```

   The system displays output similar to this:

   ```
   Event Name:         quorum_down
   Event ID:           999289
   Description:        Reasons could be network or hardware issues, or a shutdown of the cluster service.
                       The event does not necessarily indicate an issue with the cluster quorum state.
   Cause:              The local node does not have quorum. The cluster service might not be running.
   User Action:        Check if the cluster quorum nodes are running and can be reached over the network. Check local firewall settings
   Severity:           ERROR
   State:              DEGRADED
   8:08:54 PM
   2016-09-27 11:31:52.520002 CEST    move_cesip_from    INFO   Address 192.168.3.27 was moved from this node to node 3
   2016-09-27 11:32:40.576867 CEST    nfs_dbus_ok        INFO   NFS check via DBus successful
   2016-09-27 11:33:36.483188 CEST    pmsensors_down     ERROR  pmsensors service should be started and is stopped
   2016-09-27 11:34:06.188747 CEST    pmsensors_up       INFO   pmsensors service as expected, state is started


   2016-09-27 11:31:52.520002 CEST   cesnetwork    move_cesip_from    999244   INFO   Address 192.168.3.27 was moved from this node to node 3
   2016-09-27 11:32:40.576867 CEST   nfs           nfs_dbus_ok        999239   INFO   NFS check via DBus successful
   2016-09-27 11:33:36.483188 CEST   perfmon       pmsensors_down     999342   ERROR  pmsensors service should be started and is stopped
   2016-09-27 11:34:06.188747 CEST   perfmon       pmsensors_up       999341   INFO   pmsensors service as expected, state is started
   ```

10. To view the detailed description of the cluster, issue **mmhealth cluster show** command:

    ```
    mmhealth cluster show
    ```

    The system displays output similar to this:

    ```
    Component     Total   Failed   Degraded   Healthy   Other
    -------------------------------------------------------------------
    NODE             50        1          1        48       -
    GPFS             50        1          -        49       -
    NETWORK          50        -          -        50       -
    FILESYSTEM        3        -          -         3       -
    DISK             50        -          -        50       -
    CES               5        -          5         -       -
    CLOUDGATEWAY      2        -          -         2       -
    PERFMON          48        -          5        43       -
    ```

Note: The cluster must have the minimum release level as 4.2.2.0 or higher to use **mmhealth cluster show** command. Also, this command does not support Windows operating system.

11. To view more information of the cluster health status, issue this command:

```
mmhealth cluster show --verbose
```

The system displays output similar to this:

```
Component      Total   Failed   Degraded   Healthy   Other
-------------------------------------------------------------------
NODE              50        1          1        48       -
GPFS              50        1          -        49       -
NETWORK           50        -          -        50       -
FILESYSTEM
  FS1             15        -          -        15       -
  FS2              5        -          -         5       -
  FS3             20        -          -        20       -
DISK              50        -          -        50       -
CES                5        -          5         -       -
  AUTH             5        -          -         -       5
  AUTH_OBJ         5        5          -         -       -
  BLOCK            5        -          -         -       5
  CESNETWORK       5        -          -         5       -
  NFS              5        -          -         5       -
  OBJECT           5        -          -         5       -
  SMB              5        -          -         5       -
CLOUDGATEWAY       2        -          -         2       -
PERFMON           48        -          5        43       -
```

Threshold monitoring use cases

The following sections describe the use threshold use cases for the **mmhealth** command

Use case 1: Creating a threshold rule and using mmhealth commands for observing the HEALTH status changes.

1. To Monitor the **memory_free** utilization on each node create a new thresholds rule with the following settings:

```
# mmhealth thresholds add mem_memfree --errorlevel 1000000 --warnlevel 1500000
--name myTest_memfree --groupby node
```

The system displays output similar to this:

```
New rule 'myTest_memfree' is created. The monitor process is activated
```

2. To view the list of all threshold rules defined for the system, issue this command:

```
mmhealth thresholds list
```

The system displays output similar to this:

```
### Threshold Rules ###
rule_name            metric               error    warn    direction filterBy  groupBy                        sensitivity
-----------------------------------------------------------------------------------------------------------------------------
myTest_memfree       mem_memfree          1000000  1500000 None              node                           300
InodeCapUtil_Rule    Fileset_inode        90.0     80.0    high              gpfs_cluster_name,
                                                                             gpfs_fs_name,gpfs_fset_name    300
DataCapUtil_Rule     DataPool_capUtil     90.0     80.0    high              gpfs_cluster_name,
                                                                             gpfs_fs_name,gpfs_diskpool_name 300
MemFree_Rule         mem_memfree          50000    100000  low               node                           300
MetaDataCapUtil_Rule MetaDataPool_capUtil 90.0     80.0    high              gpfs_cluster_name,
                                                                             gpfs_fs_name,gpfs_diskpool_name 300
```

3. To show the **THRESHOLD** status of the current node:

```
# mmhealth node show THRESHOLD
```

The system displays output similar to this:

```
Component           Status        Status Change     Reasons
------------------------------------------------------------
THRESHOLD           HEALTHY       13 hours ago      -
  MemFree_Rule      HEALTHY       13 hours ago      -
  myTest_memfree    HEALTHY       10 min   ago      -
```

4. To view the event log history of the node issue the following command on each node:

```
# mmhealth node eventlog
2017-03-17 11:52:33.063550 CET       thresholds_error         ERROR     The value of mem_memfree for the component(s)
                                                                        myTest_memfree/gpfsgui-14.novalocal exceeded
                                                                        threshold error level 1000000 defined in myTest_memfree.

# mmhealth node eventlog
2017-03-17 11:52:32.594932 CET       thresholds_warn          WARNING   The value of mem_memfree for the component(s)
                                                                        myTest_memfree/gpfsgui-11.novalocal exceeded
                                                                        threshold warning level 1500000 defined in myTest_memfree.
2017-03-17 12:00:31.653163 CET       thresholds_normal        INFO      The value of mem_memfree defined in myTest_memfree
                                                                        for component myTest_memfree/gpfsgui-11.novalocal
                                                                        reached a normal level.

# mmhealth node eventlog
2017-03-17 11:52:35.389392 CET       thresholds_error         ERROR     The value of mem_memfree for the component(s)
                                                                        myTest_memfree/gpfsgui-13.novalocal exceeded
                                                                        threshold error level 1000000 defined in myTest_memfree.
```

5. You can view the actual metric values and compare with the rule boundaries by issuing the metric query against pmcollector node. The following example shows the **mem_memfree metric query** command and metric values for each node in the output:

```
# date; echo "get metrics mem_memfree -x -r last 10 bucket_size 300 " |
  /opt/IBM/zimon/zc gpfsgui-11
```

The system displays output similar to this:

```
Fri Mar 17 12:09:00 CET 2017
1:      gpfsgui-11.novalocal|Memory|mem_memfree
2:      gpfsgui-12.novalocal|Memory|mem_memfree
3:      gpfsgui-13.novalocal|Memory|mem_memfree
4:      gpfsgui-14.novalocal|Memory|mem_memfree
Row     Timestamp               mem_memfree     mem_memfree     mem_memfree     mem_memfree
1       2017-03-17 11:20:00     1558888 1598442 717029  768610
2       2017-03-17 11:25:00     1555256 1598596 717328  768207
3       2017-03-17 11:30:00     1554707 1597399 715988  767737
4       2017-03-17 11:35:00     1554945 1598114 715664  768056
5       2017-03-17 11:40:00     1553744 1597234 715559  766245
6       2017-03-17 11:45:00     1552876 1596891 715369  767282
7       2017-03-17 11:50:00     1450204 1596364 714640  766594
8       2017-03-17 11:55:00     1389649 1595493 714228  764839
9       2017-03-17 12:00:00     1549598 1594154 713059  765411
10      2017-03-17 12:05:00     1547029 1590308 706375  766655
...
```

6. To view the **THRESHOLD** status of all the nodes, issue this command::

```
# mmhealth cluster show THRESHOLD
```

The system displays output similar to this:

```
Component               Node                    Status          Reasons
-------------------------------------------------------------------------------------
THRESHOLD               gpfsgui-11.novalocal    HEALTHY         -
THRESHOLD               gpfsgui-13.novalocal    FAILED          thresholds_error
THRESHOLD               gpfsgui-12.novalocal    HEALTHY         -
THRESHOLD               gpfsgui-14.novalocal    FAILED          thresholds_error
```

7. To view the details of the raised event, issue this command:

```
# mmhealth event show thresholds_error
```

The system displays output similar to this:

```
Event Name:             thresholds_error
Event ID:               999892
Description:            The thresholds value reached an error level.
Cause:                  The thresholds value reached an error level.
User Action:            N/A
Severity:               ERROR
State:                  FAILED
```

8. To get an overview about the node reporting unhealthy status you can check the event log for this node, by issuing the following command:

```
# mmhealth node eventlog
```

The system displays output similar to this:

```
...
2017-03-17 11:50:23.252419 CET        move_cesip_from          INFO       Address 192.168.0.158 was moved from this node to node 0
2017-03-17 11:50:23.400872 CET        thresholds_warn          WARNING    The value of mem_memfree for the component(s)
                                                                          myTest_memfree/gpfsgui-13.novalocal exceeded
                                                                          threshold warning level 1500000 defined in myTest_memfree.
2017-03-17 11:50:26.090570 CET        mounted_fs_check         INFO       The filesystem fs2 is mounted
2017-03-17 11:50:26.304381 CET        mounted_fs_check         INFO       The filesystem gpfs0 is mounted
2017-03-17 11:50:26.428079 CET        fs_remount_mount         INFO       The filesystem gpfs0 was mounted normal
2017-03-17 11:50:27.449704 CET        quorum_up                INFO       Quorum achieved
2017-03-17 11:50:28.283431 CET        mounted_fs_check         INFO       The filesystem gpfs0 is mounted
2017-03-17 11:52:32.591514 CET        mounted_fs_check         INFO       The filesystem objfs is mounted
2017-03-17 11:52:32.685953 CET        fs_remount_mount         INFO       The filesystem objfs was mounted normal
2017-03-17 11:52:32.870778 CET        fs_remount_mount         INFO       The filesystem fs1 was mounted normal
2017-03-17 11:52:35.752707 CET        mounted_fs_check         INFO       The filesystem fs1 is mounted
2017-03-17 11:52:35.931688 CET        mounted_fs_check         INFO       The filesystem objfs is mounted
2017-03-17 12:00:36.390594 CET        service_disabled         INFO       The service auth is disabled
2017-03-17 12:00:36.673544 CET        service_disabled         INFO       The service block is disabled
2017-03-17 12:00:39.636839 CET        postgresql_failed        ERROR      postgresql-obj process should be started but is stopped

2017-03-16 12:01:21.389392 CET        thresholds_error         ERROR      The value of mem_memfree for the component(s)
                                                                          myTest_memfree/gpfsgui-13.novalocal exceeded
                                                                          threshold error level 1000000 defined in myTest_memfree.
```

9. To check the last **THRESHOLD** event update for this node, issue the following command:

```
# mmhealth node show THRESHOLD
```

The system displays output similar to this:

```
Node name:      gpfsgui-13.novalocal

Component          Status        Status Change     Reasons
-----------------------------------------------------------------------------------------------
THRESHOLD          FAILED         15 minutes ago      thresholds_error(myTest_memfree/gpfsgui-13.novalocal)
  myTest_memfree   FAILED         15 minutes ago      thresholds_error


Event                Parameter          Severity    Active Since      Event Message
-------------------------------------------------------------------------------------------------------------------------------------
thresholds_error     myTest_memfree     ERROR        15 minutes ago      The value of mem_memfree for the component(s)
                                                                         myTest_memfree/gpfsgui-13.novalocal exceeded
                                                                         threshold error level 1000000 defined in myTest_memfree.
```

10. To review the status of all services for this node, issue the following command:

```
# mmhealth node show
```

The system displays output similar to this:

```
Node name:      gpfsgui-13.novalocal
Node status:    TIPS
Status Change:  15 hours ago

Component      Status        Status Change     Reasons
-----------------------------------------------------------------------------------------------------------------
GPFS           TIPS          15 hours ago      gpfs_maxfilestocache_small, gpfs_maxstatcache_high, gpfs_pagepool_small
NETWORK        HEALTHY       15 hours ago      -
FILESYSTEM     HEALTHY       15 hours ago      -
DISK           HEALTHY       15 hours ago      -
CES            TIPS          15 hours ago      nfs_sensors_inactive
PERFMON        HEALTHY       15 hours ago      -
THRESHOLD      FAILED        15 minutes ago      thresholds_error(myTest_memfree/gpfsgui-13.novalocal)
[root@gpfsgui-13 ~]#
```

Use case 2: Creating multiple threshold rules for the same metric and using `mmhealth` commands for observing the HEALTH status changes for a particular component based on the rules specified in the filter attributes.

1. Empty the thresholds rules list for a better overview of the component status change, dependent on the count and granularity of specified thresholds rules for the same metric, using the following command:

```
# mmhealth thresholds delete all
The rule(s) was(were) deleted successfully
```

2. Create the new rule checking the **mem_memfree utilization** on each node, using the following command:

```
# mmhealth thresholds add mem_memfree --errorlevel 10000000
--warnlevel 15000000 --name all_memfree
New rule 'all_memfree' is created. The monitor process is activated
```

3. Review the new rule priority using the following command:

```
# mmhealth thresholds list -v
###   all_memfree details  ###
attribute          value
---------------------------------
rule_name          all_memfree
frequency          300
tags               thresholds
user_action_warn   None
user_action_error  None
priority           2
type               metric
metric             mem_memfree
metricOp           noOperation
sensitivity        300
computation        None
duration           None
filterBy
groupBy            None
error              10000000
warn               15000000
direction          None
hysteresis         0.0
```

4. Verify the actual metric values for the rule metric using the following query:

```
# date; echo "get metrics mem_memfree last 5 bucket_size 300 "
 | /opt/IBM/zimon/zc gpfsgui-11
Sat May 27 22:42:15 CEST 2017
1:      gpfsgui-11.novalocal|Memory|mem_memfree
2:      gpfsgui-12.novalocal|Memory|mem_memfree
3:      gpfsgui-13.novalocal|Memory|mem_memfree
4:      gpfsgui-14.novalocal|Memory|mem_memfree
5:      gpfsgui-15.novalocal|Memory|mem_memfree
Row     Timestamp              mem_memfree  mem_memfree mem_memfree  mem_memfree  mem_memfree
1       2017-05-27 22:20:00     1222358     1449223     551504       629996       780831
2       2017-05-27 22:25:00     1221110     1448821     551754       631163       781082
3       2017-05-27 22:30:00     1206205     1442715     544871       625573       774282
4       2017-05-27 22:35:00     1191082     1446694     534915       624676       777026
5       2017-05-27 22:40:00     1189409     1434247     520912       624064       775971
```

Note: In this case, the current value is lower than the thresholds error limit. The rule might raise an error.

5. Verify the status of the THRESHOLD services using the following command:

```
# mmhealth node show THRESHOLD

Node name:      gpfsgui-11.novalocal

Component       Status        Status Change     Reasons
------------------------------------------------------------------------------------------
THRESHOLD       FAILED        8 min. ago        thresholds_error(all_memfree/gpfsgui-11.novalocal)
  all_memfree   FAILED        8 min. ago        thresholds_error


Event               Parameter       Severity    Active Since      Event Message
------------------------------------------------------------------------------------------
thresholds_error    all_memfree     ERROR       8 min. ago        The value of mem_memfree for
                                                                  the component(s)
                                                                  all_memfree/gpfsgui-11.novalocal
                                                                  exceeded threshold
                                                                  error level 10000000
                                                                  defined in all_memfree.
```

Note: Status for the local node has changed to **FAILED**.

6. Create another rule to check the **mem_memfree** only for the node gpfsgui-12.

```
# mmhealth thresholds add mem_memfree --filterby node=gpfsgui-12.novalocal
--errorlevel 10600000 --warnlevel 15500000 --name gpfsgui12_memfree
New rule 'gpfsgui12_memfree' is created. The monitor process is activated
```

7. Check the priority for the new rule using the following command:

```
# mmhealth thresholds list -v
###   gpfsgui12_memfree details  ###
attribute          value
-----------------------------------------------
rule_name          gpfsgui12_memfree
frequency          300
tags               thresholds
user_action_warn   None
user_action_error  None
priority           1
type               metric
metric             mem_memfree
metricOp           noOperation
sensitivity        300
computation        None
duration           None
filterBy           node=gpfsgui-12.novalocal
groupBy            None
error              10600000
warn               15500000
direction          None
hysteresis         0.0
```

Note: The priority of the rule **gpfsgui12_memfree** is higher than the priority of **all_memfree**. Threfore, once the rule is active, only the **gpfsgui12_memfree** rule is eligible to evaluate the thresholds limits for the node gpfsgui-12 and update its status.

8. Verify the actual **mem_memfree** values small enough to cause the error event by **gpfsgui12_memfree** rule, using the following command:

```
# date; echo "get metrics mem_memfree last 5 bucket_size 300 "
 | /opt/IBM/zimon/zc gpfsgui-11
Sat May 27 22:47:39 CEST 2017
1:      gpfsgui-11.novalocal|Memory|mem_memfree
2:      gpfsgui-12.novalocal|Memory|mem_memfree
3:      gpfsgui-13.novalocal|Memory|mem_memfree
4:      gpfsgui-14.novalocal|Memory|mem_memfree
5:      gpfsgui-15.novalocal|Memory|mem_memfree
Row     Timestamp               mem_memfree  mem_memfree  mem_memfree  mem_memfree  mem_memfree
1       2017-05-27 22:25:00     1221110      1448821      551754       631163       781082
2       2017-05-27 22:30:00     1206205      1442715      544871       625573       774282
3       2017-05-27 22:35:00     1191082      1446694      534915       624676       777026
4       2017-05-27 22:40:00     1192882      1434523      525189       624599       776626
5       2017-05-27 22:45:00     1200269      1433669      535813       624870       773570
```

9. Verify the status of the THRESHOLD services using the following command:

```
# mmhealth node show THRESHOLD

Node name:      gpfsgui-11.novalocal

Component             Status        Status Change     Reasons
-----------------------------------------------------------------------------------------------------
THRESHOLD             DEGRADED      5 min. ago        thresholds_error(all_memfree/gpfsgui-11.novalocal)
  all_memfree         FAILED        19 min. ago       thresholds_error
  gpfsgui12_memfree   HEALTHY       5 min. ago        -


Event               Parameter       Severity    Active Since   Event Message
-----------------------------------------------------------------------------------------------------------
thresholds_error    all_memfree     ERROR       19 min. ago    The value of mem_memfree for the component(s)
                                                               all_memfree/gpfsgui-11.novalocal exceeded threshold
                                                               error level 10000000 defined in all_memfree.
```

Note: There is new rule **gpfsgui12_memfree** listed, but it shows a **HEALTHY** status. That is correct, because at this point the second rule has not evaluated the status of the node gpfsgui-11, to which the system is connected locally

a. Verify the status of the THRESHOLD services on the node gpfsgui-12 using the following command:

```
# mmhealth node show THRESHOLD -N gpfsgui-12

Node name:      gpfsgui-12.novalocal

Component             Status        Status Change     Reasons
------------------------------------------------------------------------------------------
THRESHOLD             DEGRADED      4 min. ago        thresholds_error
                                                      (gpfsgui12_memfree/gpfsgui-12.novalocal)
  all_memfree         DISABLED      4 min. ago        -
  gpfsgui12_memfree   FAILED        5 min. ago        thresholds_error


Event               Parameter           Severity    Active Since     Event Message
------------------------------------------------------------------------------------------------
thresholds_error    gpfsgui12_memfree    ERROR       5 min. ago       The value of mem_memfree
                                                                      for the component(s)
                                                                      gpfsgui12_memfree/gpfsgui-12.novalocal
                                                                      exceeded threshold
                                                                      error level 10600000
                                                                      defined in gpfsgui12_memfree.
```

Note: There is an event raised by the **gpfsgui12_memfree** rule, and the status of the whole service is DEGRADED.

10. Create a third rule that checks the **mem_memfree** rule for the node gpfsgui-15, using the following command:

```
# mmhealth thresholds add mem_memfree --filterby node=gpfsgui-15.novalocal
--errorlevel 10600000 --warnlevel 15500000 --name gpfsgui15_memfree
New rule 'gpfsgui15_memfree' is created. The monitor process is activated
```

11. Verify the list of active rules sing the following command:

```
# mmhealth thresholds list
### Threshold Rules ###
rule_name          metric       error     warn      direction  filterBy                   groupBy  sensitivity
-----------------------------------------------------------------------------------------------------------------
gpfsgui12_memfree  mem_memfree  10600000  15500000  None       node=gpfsgui-12.novalocal  None     300
all_memfree        mem_memfree  10000000  15000000  None                                  None     300
gpfsgui15_memfree  mem_memfree  10600000  15500000  None       node=gpfsgui-15.novalocal  None     300
```

12. Review the status of the THRESHOLD service on each particular node using the following command:

```
# mmhealth node show THRESHOLD

Node name:      gpfsgui-11.novalocal

Component             Status        Status Change     Reasons
-------------------------------------------------------------------------------------------
THRESHOLD             DEGRADED      28 min. ago      thresholds_error(all_memfree/gpfsgui-11.novalocal)
  all_memfree         FAILED        34 min. ago      thresholds_error
  gpfsgui12_memfree   HEALTHY       18 min. ago      -
  gpfsgui15_memfree   HEALTHY       7 min. ago      -


Event               Parameter      Severity   Active Since     Event Message
-----------------------------------------------------------------------------------------------------------
thresholds_error    all_memfree    ERROR      16 hours ago     The value of mem_memfree for the component(s)
                                                               all_memfree/gpfsgui-11.novalocal exceeded threshold
                                                               error level 10000000 defined in all_memfree.

# mmhealth node show THRESHOLD -N gpfsgui-12

Node name:      gpfsgui-12.novalocal

Component             Status        Status Change     Reasons
-------------------------------------------------------------------------------------------
```

```
THRESHOLD             DEGRADED      28 min. ago      thresholds_error(gpfsgui12_memfree/gpfsgui-12.novalocal)
  all_memfree         DISABLED      18 min. ago      -
  gpfsgui12_memfree   FAILED        28 min. ago      thresholds_error
  gpfsgui15_memfree   HEALTHY       7 min. ago      -


Event              Parameter           Severity   Active Since     Event Message
-------------------------------------------------------------------------------------------------------------------
thresholds_error   gpfsgui12_memfree   ERROR      16 hours ago     The value of mem_memfree for the component(s)
                                                                   gpfsgui12_memfree/gpfsgui-12.novalocal exceeded
                                                                   threshold error level 10600000 defined in
                                                                   gpfsgui12_memfree.


# mmhealth node show THRESHOLD -N gpfsgui-15

Node name:      gpfsgui-15.novalocal

Component             Status        Status Change     Reasons
-------------------------------------------------------------------------------------------------------------
THRESHOLD             DEGRADED      28 min. ago      thresholds_error(gpfsgui15_memfree/gpfsgui-15.novalocal)
  all_memfree         DISABLED      1 min. ago      -
  gpfsgui12_memfree   HEALTHY       28 min. ago      -
  gpfsgui15_memfree   FAILED        7 min. ago     thresholds_error


Event           Parameter         Severity    Active Since      Event Message
-------------------------------------------------------------------------------------------------------------------
                                                                gpfsgui15_memfree/gpfsgui-15.novalocal exceeded
                                                                threshold error level 10600000 defined in
                                                                gpfsgui15_memfree.
```

13. Review the node eventlog of the node `gpfsgui-15` to see the full event history, using the following command:

```
# mmhealth node eventlog -N gpfsgui-15
...
2017-05-27 22:33:56.877481 CEST    thresholds_error       ERROR     The value of mem_memfree for the component(s)
                                                                    all_memfree/gpfsgui-15.novalocal exceeded
                                                                    threshold error level 10000000 defined in
                                                                    all_memfree.
2017-05-27 23:08:26.358688 CEST    thresholds_error       ERROR     The value of mem_memfree for the component(s)
                                                                    gpfsgui15_memfree/gpfsgui-15.novalocal exceeded
                                                                    threshold error level 10600000 defined in
                                                                    gpfsgui15_memfree.
2017-05-27 23:13:56.392194 CEST    thresholds_removed     INFO      The value of mem_memfree for the component(s)
                                                                    all_memfree/gpfsgui-15.novalocal defined in
                                                                    all_memfree was removed.
```

14. Verify that a second rule managing exactly the same metric and component entity is not allowed, using the following command:

```
# mmhealth thresholds add mem_memfree --filterby node=gpfsgui-15.novalocal
 --errorlevel 10600000 --warnlevel 15500000 --name second_gpfsgui15_memfree
The rule 'gpfsgui15_memfree' is already active for the specified filterBy entr(y)ies
```

Use case 3: Creating threshold rules for specific filesets.

The initial cluster file system and fileset might have the following setting:

```
[root@gpfsgui-11 ~]# mmlsfileset nfs_shareFS
Filesets in file system 'nfs_shareFS':
Name                    Status    Path
root                    Linked    /gpfs/nfs_shareFS
nfs_shareFILESET        Linked    /gpfs/nfs_shareFS/nfs_shareFILESET
test_share              Linked    /gpfs/nfs_shareFS/test_share

Per default, the following rules are defined and enabled in the cluster:
[root@gpfsgui-11 ~]# mmhealth thresholds list
### Threshold Rules ###
rule_name            metric               error  warn    direction  filterBy  groupBy                                          sensitivity
-----------------------------------------------------------------------------------------------------------------------------------------------
InodeCapUtil_Rule    Fileset_inode        90     80      None               gpfs_cluster_name,gpfs_fs_name,gpfs_fset_name    300
DataCapUtil_Rule     DataPool_capUtil     90.0   80.0    high               gpfs_cluster_name,gpfs_fs_name,gpfs_diskpool_name 300
MemFree_Rule         mem_memfree          50000  100000  low                node                                             300
MetaDataCapUtil_Rule MetaDataPool_capUtil 90.0   80.0    high               gpfs_cluster_name,gpfs_fs_name,gpfs_diskpool_name 300
```

1. Create new inode capacity utilization rules for the specific filesets.
   a. To create a threshold rule for all filesets in an individual file system use the following command. For example, for the `nfs_shareFS` file system , you need to set the rule filter to the following file system:

```
[root@gpfsgui-11 ~]# mmhealth thresholds add Fileset_inode --errorlevel 85.0
--warnlevel 75.0 --direction high --filterby 'gpfs_fs_name=nfs_shareFS'
--groupby gpfs_cluster_name,gpfs_fs_name,gpfs_diskpool_name
--sensitivity 300 --hysteresis 5.0 --name rule_ForAllFsets_inFS

New rule 'rule_ForAllFsets_inFS' is created. The monitor process is activated
```

   b. To create a threshold rule for individual filesets. For example, for the `nfs_shareFILESET` fileset , you need to specify both, the file system name and the fileset name in the filter as shown:

```
[root@gpfsgui-11 ~]# mmhealth thresholds add Fileset_inode --errorlevel 70.0
--warnlevel 60.0 --direction high --filterby 'gpfs_fs_name=nfs_shareFS,gpfs_fset_name=nfs_shareFILESET'
--groupby gpfs_cluster_name,gpfs_fs_name,gpfs_fset_name --sensitivity 300
--hysteresis 5.0 --name rule_SingleFset_inFS

New rule 'rule_SingleFset_inFS' is created. The monitor process is activated
```

   Once the **mmhealth thresholds add** commands are run, the list will look as follows:

```
# mmhealth thresholds list
### Threshold Rules ###
rule_name              metric               error  warn    direction  filterBy                          groupBy                            sensitivity
------------------------------------------------------------------------------------------------------------------------------------------------------------
InodeCapUtil_Rule      Fileset_inode        90     80      None                                         gpfs_cluster_name,gpfs_fs_name,     300
                                                                                                        gpfs_fset_name
DataCapUtil_Rule       DataPool_capUtil     90.0   80.0    high                                         gpfs_cluster_name,                  300
                                                                                                        gpfs_fs_name,
                                                                                                        gpfs_diskpool_name
MemFree_Rule           mem_memfree          50000  100000  low                                          node                                300
rule_SingleFset_inFS   Fileset_inode        70.0   60.0    high       gpfs_fs_name=nfs_shareFS,         gpfs_cluster_name,                  300
                                                                      gpfs_fset_name=nfs_shareFILESET   gpfs_fs_name,gpfs_fset_name
rule_ForAllFsets_inFS  Fileset_inode        85.0   75.0    high       gpfs_fs_name=nfs_shareFS          gpfs_cluster_name,                  300
                                                                                                        gpfs_fs_name,gpfs_fset_name
MetaDataCapUtil_Rule   MetaDataPool_capUtil 90.0   80.0    high                                         gpfs_cluster_name,
                                                                                                        gpfs_fs_name,gpfs_diskpool_name     300
```

2. Run the **mmhealth thresholds list** command to list the individual rules' priorities. In this example, the `rule_SingleFset_inFS` rule has the highest priority for the `nfs_shareFILESET` fileset. The `rule_ForAllFsets_inFS` rule has the highest priority for the other filesets belonging to the `nfs_shareFS` file system, and the `InodeCapUtil_Rule` rule is valid for all remaining filesets.

```
[root@gpfsgui-11 ~]# mmhealth thresholds list -v -Y|grep priority
mmsysmonc_thresholdslist::0:1::InodeCapUtil_Rule:priority:4:
mmsysmonc_thresholdslist::0:1::DataCapUtil_Rule:priority:4:
mmsysmonc_thresholdslist::0:1::MemFree_Rule:priority:2:
mmsysmonc_thresholdslist::0:1::rule_SingleFset_inFS:priority:1:
mmsysmonc_thresholdslist::0:1::rule_ForAllFsets_inFS:priority:2:
mmsysmonc_thresholdslist::0:1::MetaDataCapUtil_Rule:priority:4:
[root@gpfsgui-11 ~]#
```

Chapter 4. Monitoring events through callbacks

You can configure the callback feature to provide notifications when node and cluster events occur. Starting complex or long-running commands, or commands that involve GPFS files, might cause unexpected and undesired results, including loss of file system availability. Use the **mmaddcallback** command to configure the callback feature.

For more information on how to configure and manage callbacks, see the man page of the following commands in *IBM Spectrum Scale: Command and Programming Reference*:

- **mmaddcallback**
- **mmdelcallback**
- **mmlscallback**

Chapter 5. Monitoring capacity through GUI

You can monitor the capacity of the file system, pools, filesets, NSDs, users, and user groups in the IBM Spectrum Scale system.

The capacity details displayed in the GUI are obtained from the following sources:

- GPFS quota database. The system collects the quota details and stores it in the postgres database.
- Performance monitoring tool. The performance monitoring tool collects the capacity data and displays it in the various pages in the GUI.

Based on the source of the capacity information, different procedures need to be performed to enable capacity and quota data collection.

For both GPFS quota database and performance monitoring tool-based capacity and quota collection, you need to use the **Files** > **Quotas** page to enable quota data collection per file system and enforce quota limit checking. If quota is not enabled for a file system:

- No capacity and inode data is collected for users, groups, and filesets.
- Quota limits for users, groups, and filesets cannot be defined.
- No alerts are sent and the data writes are not restricted.

To enable capacity data collection from the performance monitoring tool, the **GPFSFilesetQuota** sensor must be enabled. For more information on how to enable the performance monitoring sensor for capacity data collection, see *Manual installation of IBM Spectrum Scale GUI* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Capacity data obtained from the GPFS quota database

The capacity and quota information collected from the GPFS quota database is displayed on the Files > Quotas and Files > User Capacity pages in the management GUI.

1. **Files** > **Quotas page**

Use quotas to control the allocation of files and data blocks in a file system. You can create default, user, group, and fileset quotas through the Quotas page.

A quota is the amount of disk space and the amount of metadata that is assigned as upper limits for a specified user, group of users, or fileset. Use the **Actions** menu to create or modify quotas. The management GUI allows you to only manage capacity-related quota. The inode-related quota management is only possible in the command-line interface.

You can specify a soft limit, a hard limit, or both. When you set a soft limit quota, a warning is sent to the administrator when the file system is close to reaching its storage limit. A grace period starts when the soft quota limit is reached. Data is written until the grace period expires, or until the hard quota limit is reached. Grace time resets when used capacity goes below the soft limit. If you set a hard limit quota, you cannot save data after the quota is reached. If the quota is exceeded, you must delete files or raise the quota limit to store more data.

Note:

- User or user group quotas for filesets are only supported if the *Per Fileset* option is enabled at the file system level. Use the command-line interface to set the option. See the manpages of **mmcrfs** and **mmchfs** commands for more detail.

- You need to unmount a file system to change the quota enablement method from per file system to per fileset or vice versa.

You can set default user quotas at the file system level rather than defining user quotas explicitly for each user. Default quota limits can be set for users. You can specify the general quota collection scope such as per file system or per fileset to define whether the default quota needs to be defined at file system level or fileset level and set the default user quota. After this value is set, all child objects that are created under the file system or fileset will be configured with the default soft and hard limits. You can assign a custom quota limit to individual child objects, but the default limits remain the same unless changed at the file system or fileset level.

After reconfiguring quota settings, it is recommended to run the **mmcheckquota** command for the affected file system to verify the changes.

For more information on how to manage quotas, see *Managing GPFS quotas* section in the *IBM Spectrum Scale: Administration Guide*.

Capacity data from users, groups, and filesets with no quota limit set are not listed in the Quotas page. Use the **Files** > **User Capacity** page to see capacity information of such users and groups. Use the **Files** > **Filesets** page to view current and historic capacity information of filesets.

2. **Files** > **User Capacity page**

The **Files** > **User Capacity** page provides predefined capacity reports for users and groups. While capacity information of file systems, pools, and filesets is available in the respective areas of the GUI, the **Files** > **User Capacity** page is the only place where information on used capacity per user or group is available.

The User Capacity page depends on the quota accounting method of the file system. You need to enable quota for a file system to display the user capacity data. If quota is not enabled, you can follow the fix procedure in the **Files** > **Quotas** page or use the **mmchfs <Device> -Q yes** CLI command to enable quota. Even if the capacity limits are not set, the User Capacity page shows data as soon as the quota accounting is enabled and users write data. This is different in the Quotas page, where only users and groups with quota limits defined are listed. The user and group capacity quota information is automatically collected once a day by the GUI.

For users and user groups, you can see the total capacity and whether quotas are set for these objects. you can also see the percentage of soft limit and hard limit usage. When the hard limit is exceeded, no more files belong to the respective user, user group, or fileset can be written. However, exceeding the hard limit allows a certain grace period before disallowing more file writes. Soft and hard limits for disk capacity are measured in units of kilobytes (KiB), megabytes (MiB), or gigabytes (GiB). Use the **Files** > **Quotas** page to change the quota limits.

Capacity data collected through the performance monitoring tool

The historical capacity data collection for file systems, pools, and filesets depend on the correctly configured data collection sensors for fileset quota and disk capacity. When the IBM Spectrum Scale system is installed through the installation toolkit, the capacity data collection is configured by default. In other cases, you need to enable capacity sensors manually.

If the capacity data collection is not configured correctly you can use **mmperfmon** CLI command or the **Services** > **Performance Monitoring** > **Sensors** page.

The **Services** > **Performance Monitoring** > **Sensors** page allows to view and edit the sensor settings. By default, the collection periods of capacity collection sensors are set to collect data with a period of up to one day. Therefore, it might take a while until the data is refreshed in the GUI.

The following sensors are collecting capacity related information and are used by the GUI.

GPFSDiskCap
: NSD, Pool and File system level capacity. Uses the **mmdf** command in the background and typically runs once per day as it is resource intensive. Should be restricted to run on a single node only.

GPFSPool
: Pool and file system level capacity. Requires a mounted file system and typically runs every 5 minutes. Should be restricted to run on a single node only.

GPFSFilesetQuota
: Fileset capacity based on the quota collection mechanism. Typically, runs every hour. Should be restricted to run only on a single node.

GPFSFileset
: Inode space (independent filest) capacity and limits. Typically runs every 5 minutes. Should be restricted to run only on a single node.

DiskFree
: Overall capacity and local node capacity. Can run on every node.

The **Monitoring** > **Statistics** page allows to create customized capacity reports for file systems, pools and filesets. You can store these reports as favorites and add them to the dashboard as well.

The dedicated GUI pages combine information about configuration, health, performance, and capacity in one place. The following GUI pages provide the corresponding capacity views:

- **Files** > **File Systems**
- **Files** > **Filesets**
- **Storage** > **Pools**
- **Storage** > **NSDs**

The Filesets grid and details depend on quota that is obtained from the GPFS quota database and the performance monitoring sensor *GPFSFilesetQuota*. If quota is disabled, the system displays a warning dialog in the Filesets page.

Troubleshooting issues with capacity data displayed in the GUI

Due to the impact that capacity data collection can have on the system, different capacity values are collected on a different schedule and are provided by different system components. The following list provides insight on the issues that can arise from the multitude of schedules and subsystems that provide capacity data:

Capacity in the file system view and the total amount of the capacity for pools and volumes view do not match.
: The capacity data in the file system view is collected every 10 minutes by performance monitoring collector, but the capacity data for pools and Network Shared Disks (NSD) are not updated. By default, NSD data is only collected once per day by performance monitoring collector and it is cached. Clicking the refresh icon gathers the last two records from performance monitoring tool and it displays the last record values if they are not null. If the last record has null values, the system displays the previous one. If the values of both records are null, the system displays N/A and the check box for displaying a time chart is disabled. The last update date is the record date that is fetched from performance monitoring tool if the values are not null.

Capacity in the file system view and the total amount of used capacity for all filesets in that file system do not match.
: There are differences both in the collection schedule as well as in the collection mechanism that contributes to the fact that the fileset capacities do not add up to the file system used capacity.

Scheduling differences:

Capacity information that is shown for filesets in the GUI is collected once per hour by performance monitoring collector and displayed on Filesets page. When you click the refresh icon you get the information of the last record from performance monitoring. If the last two records have null values, you get a 'Not collected' warning for used capacity. The file system capacity information on the file systems view is collected every 10 minutes by performance monitoring collector and when you click the refresh icon you get the information of the last record from performance monitoring.

Data collection differences:

Quota values show the sum of the size of all files and are reported asynchronously. The quota reporting does not consider metadata, snapshots, or capacity that cannot be allocated within a subblock. Therefore, the sum of the fileset quota values can be lower than the data shown in the file system view. You can use the CLI command **mmlsfileset** with the **-d** and **-i** options to view capacity information. The GUI does not provide a means to display this values because of the performance impact due to data collection.

The sum of all fileset inode values on the view quota window does not match the number of inodes that are displayed on the file system properties window.

The quota value only accounts for user-created inodes while the properties for the file system also display inodes that are used internally. Refresh the quota data to update these values.

No capacity data shown on a new system or for a newly created file system

Capacity data may show up with a delay of up to 1 day. The capacity data for file systems, NSDs, and pools is collected once a day as this is a resource intensive operation. Line charts do not show a line if only a single data point exists. You can use the hover function in order to see the first data point in the chart.

The management GUI displays negative fileset capacity or an extremely high used capacity like millions of Petabytes or 4000000000 used inodes.

This problem can be seen in the quota and filesets views. This problem is caused when the quota accounting is out of sync. To fix this error, issue the **mmcheckquota** command. This command recounts inode and capacity usage in a file system by user, user group, and fileset, and writes the collected data into the database. It also checks quota limits for users, user groups, and filesets in a file system. Running this command can impact performance of I/O operations.

No capacity data is displayed on the performance monitoring charts

Verify whether the GPFSFilesetQuota sensor is enabled. You can check the sensor status form the Services > Performance Monitoring page in the GUI. For more information on how to enable the performance monitoring sensor for capacity data collection, see *Manual installation of IBM Spectrum Scale GUI* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Chapter 6. Monitoring AFM and AFM DR

The following sections inform you how to monitor and troubleshoot AFM and AFM DR filesets.

Monitoring fileset states for AFM

AFM fileset can have different states depending on the mode and queue states.

To view the current cache state, run the
`mmafmctl` *`filesystem`* `getstate`

command, or the
`mmafmctl` *`filesystem`* `getstate -j` *`cache_fileset`*

command.

See the following table for the explanation of the cache state:

Table 33. AFM states and their description

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Inactive	The AFM cache is created	Operations were not initiated on the cache cluster after the last daemon restart.	Healthy	None
FlushOnly	Operations are queued	Operations have not started to flush.	Healthy	This is a temporary state and should move to Active when a write is initiated.
Active	The AFM cache is active	The cache cluster is ready for an operation.	Healthy	None
Dirty	The AFM is active	The pending changes in the cache cluster are not played at the home cluster. This state does not hamper the normal activity.	Healthy	None
Recovery	The cache is accessed after primary gateway failure	A new gateway is taking over a fileset as primary gateway after the old primary gateway failed.	Healthy	None
QueueOnly	The cache is running some operation.	Operations such as recovery, resync, failover are being executed, and operations are being queued and not flushed.	Healthy	This is a temporary state.

Table 33. AFM states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Disconnected	Primary gateway cannot connect to the NFS server at the home cluster.	Occurs only in a cache cluster that is created over an NFS export. When parallel data transfer is configured, this state shows the connectivity between the primary gateway and the mapped home server, irrespective of other gateway nodes.	Unhealthy	Correct the errant NFS servers on the home cluster.
Unmounted	The cache that is using NFS has detected a change in the home cluster - sometimes during creation or in the middle of an operation if home exports are meddled with.	• The home NFS is not accessible • The home exports are not exported properly • The home export does not exist	Unhealthy	1. Fix the NFS export issue in the Home setup section and retry for access. 2. Relink the cache cluster if the cache cluster does not recover. After **mountRetryInterval** of the primary gateway, the cache cluster retries connecting with home.
Unmounted	The cache that is using the GPFS protocol detects a change in the home cluster, sometimes during creation or in the middle of an operation.	There are problems accessing the local mount of the remote file system.	Unhealthy	Check remote filesystem mount on the cache cluster and remount if necessary.
Dropped	Recovery failed.	The local file system is full, space is not available on the cache or the primary cluster, or case of a policy failure during recovery.	Unhealthy	Fix the issue and access the fileset to retry recovery.
Dropped	IW Failback failed.	The local file system is full, space is not available on the cache or the primary cluster, or there is a policy failure during recovery.	Unhealthy	Fix the issue and access the fileset to retry failback.

Table 33. AFM states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Dropped	A cache with active queue operations is forcibly unlinked.	All queued operations are being de-queued, and the fileset remains in the Dropped state and moves to the Inactive state when the unlinking is complete.	Healthy	This is a temporary state.
Dropped	The old GW node starts functioning properly after a failure	AFM internally performs queue transfers from one gateway to another to handle gateway node failures.	Healthy	The system resolves this state on the next access.
Dropped	Cache creation or in the middle of an operation if the home exports changed.	Export problems at home such as following: • The home path is not exported on all NFS server nodes that are interacting with the cache clusters. • The home cluster is exported after the operations have started on the fileset. Changing `fsid` on the home cluster after the fileset operations have begun. • All home NFS servers do not have the same `fsid` for the same export path.	Unhealthy	1. Fix the NFS export issue in the Home setup section and retry for access. 2. Relink the cache cluster if the cache cluster does not recover. After **`mountRetryInterval`** the primary gateway retries connecting with home cluster.
Dropped	During recovery or normal operation	If gateway queue memory is exceeded, the queue can get dropped. The memory has to be increased to accommodate all requests and bring the queue back to the Active state.	Unhealthy	Increase **`afmHardMemThreshold`**.
Expired	The RO cache that is configured to expire	An event that occurs automatically after prolonged disconnection when the cached contents are not accessible.	Unhealthy	Fix the errant NFS servers on the home cluster

Table 33. AFM states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
NeedsFailback	The IW cache that needs to complete failback	A failback initiated on an IW cache cluster is interrupted and is incomplete.	Unhealthy	Failback is automatically triggered on the fileset, or the administrator can run failback again.
FailbackInProgress	Failback initiated on IW cache	Failback is in progress and automatically moves to `failbackCompleted`	Healthy	None
FailbackCompleted	The IW cache after failback	Failback successfully completes on the IW cache cluster.	Healthy	Run **mmafmctl failback --stop** on the cache cluster.
NeedsResync	The SW cache cluster during home corruption	Occurs when the home cluster is accidentally corrupted	Unhealthy	Run **mmafmctl resync** on the cache.
NeedsResync	Recovery on the SW cache	A rare state possible only under error conditions during recovery	Unhealthy	No administrator action required. The system would fix this in the subsequent recovery.
Stopped	Replication stopped on fileset.	Fileset stops sending changes to the gateway node. Mainly used during planned downtime.	Unhealthy	After planned downtime, run **mmafmctl <fs> start -j <fileset>** to start sending changes/modification to the gateway node and continue replication.

Monitoring fileset states for AFM DR

AFM DR fileset can have different states depending on the mode and queue states.

Run the **mmafmctl getstate** command to view the current cache state.

See the following table:

Table 34. AFM DR states and their description

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Inactive	AFM primary is created	Operations have not been initiated on the primary after last daemon restart.	Healthy	None

Table 34. AFM DR states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
FlushOnly	Operations are queued	Operations have not started to flush. This is a temporary state and moves to Active when a write is initiated.	Healthy	
Active	AFM primary is active	Primary is ready for operation	Healthy	None
Dirty	AFM primary is active	Indicates there are pending changes in primary not yet played at secondary. Does not hamper normal activity.	Healthy	None
Recovery	The primary is accessed after MDS failure	Can occur when a new gateway is taking over a fileset as MDS after the old MDS failed.	Healthy	None
QueueOnly	The primary is running some operation	Can occur when operations such as recovery are being executed and operations are being queued and are not yet flushed.	Healthy	This is a temporary state.
Disconnected	It occurs when the MDS cannot connect to the NFS server at secondary	Occurs only in a cache cluster that is created over NFS export. When parallel I/O is configured, this state shows the connectivity between the MDS and the mapped home server, irrespective of other gateway nodes.	Unhealthy	Correct the errant NFS servers on the secondary cluster.
Unmounted	Primary using NFS detects a change in secondary - sometimes during creation or in the middle of operation if secondary exports are interfered	This can occur if: • Secondary NFS is not accessible • Secondary exports are not exported properly • Secondary export does not exist	Unhealthy	1. Rectify the NFS export issue as in secondary setup section and retry access 2. Relink primary if it does not recover. After **mountRetryInterval** of the MDS, the primary retries connecting with secondary

Table 34. AFM DR states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Unmounted	The primary that is using the GPFS protocol detects a change in the secondary cluster, sometimes during creation or in the middle of an operation	Occurs when there are problems accessing the local mount of the remote file system.	Unhealthy	Check remote filesystem mount on the primary cluster and remount if necessary.
Dropped	Recovery failed.	Occurs when the local file system is full, space is not available on the primary, or a policy failure during recovery.	Unhealthy	Fix the issue and access the fileset to retry recovery.
Dropped	A primary with active queue operations is forcibly unlinked	All queued operations are being de-queued, and the fileset remains in the Dropped state and moves to the Inactive state when the unlinking is complete.	Healthy	This is a temporary state.
Dropped	Old GW node starts functioning properly after a failure	AFM internally performs queue transfers from one gateway to another to handle gateway node failures.	Healthy	The system resolves this state on the next access.
Dropped	Primary creation or in the middle of an operation if the home exports changed.	Export problems at secondary such as: • The home path is not exported on all NFS server nodes that are interacting with the cache clusters. Even if the home cluster is exported after the operations have started on the fileset, problems might persist. • Changing `fsid` on the home cluster after the fileset operations have begun. • All home nfs servers do not have the same `fsid` for the same export path.	Unhealthy	1. Fix the NFS export issue in the secondary setup section and retry for access. 2. Relink the primary if the cache cluster does not recover. After **`mountRetryInterval`** the MDS retries connecting with the secondary.

Table 34. AFM DR states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
Dropped	During recovery or normal operation	If gateway queue memory is exceeded, the queue can get dropped. The memory has to be increased to accommodate all requests and bring the queue back to the Active state.	Unhealthy	Increase **afmHardMemThreshold**.
NeedsResync	Recovery on primary	This is a rare state and is possible only under error conditions during recovery.	Unhealthy	The problem gets fixed automatically in the subsequent recovery.
NeedsResync	Failback on primary or conversion from GPFS/SW to primary	This is a rare state and is possible only under error conditions during failback or conversion.	Unhealthy	Rerun failback or conversion.
PrimInitProg	Setting up primary and secondary relationship during - • creation of a primary fileset. • conversion of gpfs, sw, or iw fileset to primary fileset. • change secondary of a primary fileset.	This state is used while primary and secondary are in the process of establishing a relationship while the psnap0 is in progress. All operations are disallowed till psnap0 is taken locally. This should move to active when psnap0 is queued and played on the secondary side.	Healthy	Review errors on **psnap0** failure if fileset state is not active.
PrimInitFail	Failed to set up primary and secondary relationship during - • creation of a primary fileset. • conversion of gpfs, sw, or iw fileset to primary fileset. • change secondary of a primary fileset.	This is a rare failure state when the **psnap0** has not been created at the primary. In this state no data is moved from the primary to the secondary. The administrator should check that the gateway nodes are up and file system is mounted on them on the primary. The secondary fileset should also be setup correctly and available for use.	Unhealthy	• Review errors after **psnap0** failure. • Re-running the **mmafmctl convertToPrimary** command without any parameters ends this state.

Table 34. AFM DR states and their description (continued)

AFM fileset state	Condition	Description	Healthy or Unhealthy	Administrator's action
FailbackInProgress	Primary failback started	This is the state when failback is initiated on the primary.	Healthy	None
Stopped	Replication stopped on fileset.	Fileset stops sending changes to the gateway node. Mainly used during planned downtime.	Unhealthy	After planned downtime, run **mmafmctl <fs> start -j <fileset>** to start sending changes/modification to the gateway node and continue replication.

Monitoring health and events

You can use **mmhealth** to monitor health.

To monitor callback events, you can use **mmaddcallback** and **mmdelcallback**.

Monitoring with mmhealth

You can use **mmhealth** to monitor AFM and AFM DR.

Use the following **mmhealth** command to display the health status of the gateway node:

```
# mmhealth node show AFM
Node name:      p7fbn10.gpfs.net

Component            Status        Status Change     Reasons
-------------------------------------------------------------
AFM                  HEALTHY       3 days ago        -
  fs1/p7fbn10ADR-4   HEALTHY       3 days ago        -
  fs1/p7fbn10ADR-5   HEALTHY       3 days ago        -

There are no active error events for the component AFM on this node (p7fbn10.gpfs.net).
p7fbn10 Wed Mar 15 04:34:41 1]~# mmhealth node show AFM  -Y
mmhealth:State:HEADER:version:reserved:reserved:node:component:entityname:entitytype:status:laststatuschange:
mmhealth:Event:HEADER:version:reserved:reserved:node:component:entityname:entitytype:event:arguments:
activesince:identifier:ishidden:
mmhealth:State:0:1:::p7fbn10.gpfs.net:NODE:p7fbn10.gpfs.net:NODE:DEGRADED:2017-03-11 18%3A48%3A20.600167 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:p7fbn10.gpfs.net:NODE:HEALTHY:2017-03-11 19%3A56%3A48.834633 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:fs1/p7fbn10ADR-5:FILESET:HEALTHY:2017-03-11 19%3A56%3A48.834753 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:fs1/p7fbn10ADR-4:FILESET:HEALTHY:2017-03-11 19%3A56%3A19.086918 EDT:
```

Use the following **mmhealth** command to display the health status of all the monitored AFM components in the cluster:

```
# mmhealth cluster show AFM
Node name:      p7fbn10.gpfs.net

Component            Status        Status Change     Reasons
-------------------------------------------------------------
AFM                  HEALTHY       3 days ago        -
  fs1/p7fbn10ADR-4   HEALTHY       3 days ago        -
  fs1/p7fbn10ADR-5   HEALTHY       3 days ago        -

There are no active error events for the component AFM on this node (p7fbn10.gpfs.net).
p7fbn10 Wed Mar 15 04:34:41 1]~# mmhealth node show AFM  -Y
mmhealth:State:HEADER:version:reserved:reserved:node:component:entityname:entitytype:status:laststatuschange:
mmhealth:Event:HEADER:version:reserved:reserved:node:component:entityname:entitytype:event:arguments:
activesince:identifier:ishidden:
mmhealth:State:0:1:::p7fbn10.gpfs.net:NODE:p7fbn10.gpfs.net:NODE:DEGRADED:2017-03-11 18%3A48%3A20.600167 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:p7fbn10.gpfs.net:NODE:HEALTHY:2017-03-11 19%3A56%3A48.834633 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:fs1/p7fbn10ADR-5:FILESET:HEALTHY:2017-03-11 19%3A56%3A48.834753 EDT:
mmhealth:State:0:1:::p7fbn10.gpfs.net:AFM:fs1/p7fbn10ADR-4:FILESET:HEALTHY:2017-03-11 19%3A56%3A19.086918 EDT:
```

Monitoring callback events for AFM and AFM DR

You can use events to monitor AFM and AFM DR fileset.

All events are at the fileset level. To add the events, run the **mmaddcallback** command.

An example of the command is

```
#mmdelcallback callback3
mmaddcallback callback3 --command /tmp/recovery_events.sh --event
afmRecoveryStart --parms "%eventName %homeServer %fsName %filesetName
    %reason"
```

Table 35. List of events that can be added using **mmaddcallback**

Events	Applicable to..	Description
afmprepopend	All AFM filesets	Completion of the prefetch task.
afmRecoveryStart	SW, IW, DR filesets	Beginning of the recovery process.
afmRecoveryEnd	SW, IW, DR filesets	End of the recovery process.
afmRPOMiss	primary	Indicates that RPO is missed due to a network delay or a failure to create snapshot on secondary side. Failed RPOs are queued and tried again on the secondary.
afmHomeDisconnected	All AFM filesets, DR filesets	For NFS target: The AFM home/DR secondary is not reachable.
afmHomeConnected	All AFM filesets, DR filesets	For NFS target: The AFM home/DR secondary is reachable.
afmFilesetExpired	RO fileset	For RO fileset: Fileset has expired
afmFilesetUnexpired	RO fileset	For RO fileset: Fileset is back to Active after expiration.
afmManualResyncComplete	SW, IW, DR filesets	The SW resync or failover process is complete after - • conversion of gpfs, sw, or iw fileset to primary fileset. • change secondary of a primary fileset.
afmQueueDropped	All AFM filesets, DR filesets	The queue is dropped.
afmfilesetunmounted	All AFM filesets, DR filesets	The fileset is in the Unmounted state.
afmFilesetCreate	All AFM filesets	The fileset is created successfully.
afmFilesetLink	All AFM filesets	The fileset is linked successfully.
afmFilesetChange	All AFM filesets	The fileset is changed successfully. If the fileset was renamed, then the new name is mentioned in %reason.
afmFilesetUnlink	All AFM filesets	The fileset is unlinked successfully.
afmFilesetDelete	All AFM filesets	The fileset is deleted successfully.

Monitoring performance

You can use **mmperfmon** and **mmpmon** commands to monitor AFM and AFM DR.

Monitoring using mmpmon

You can use **mmpmon** to monitor AFM and AFM DR.

1. To reset some statistics on a gateway node, run the following commands:

   ```
   echo "afm_s reset" | mmpmon
   echo "afm_s fset all" | mmpmon
   ```

2. To reset all statistics, run the following command:

   ```
   mmfsadm afm resetall
   ```

3. To view the statistics, run the following command:

   ```
   echo afm_s | mmpmon -s -r 0 -d 2000
   ```

   This command shows statistics from the time the Gateway is functioning. Every gateway recycle resets the statistics.

   The following example is from an AFM Gateway node. The example shows how many operations of each type were executed on the gateway node.

   ```
   c2m3n10 Tue May 10 09:55:59 0]~# echo afm_s | mmpmon

   mmpmon> mmpmon node 192.168.2.20 name c2m3n10 afm_s s OK
   Name                Queued     Inflight   Completed  Errors     Filtered   ENOENT
   lookup              0          0          1          0          0          0
   create              0          0          20         0          10         0
   remove              0          0          0          0          10         0
   open                0          0          2          0          0          0
   read                0          0          0          0          1          0
   write               0          0          20         0          650        0
   BytesWritten = 53320860 (50.85 MB) (26035.58 KB/s) BytesToWrite = 0 (0.00 KB)
   Queue Delay (s) (min:0 max:19 avg:18)
   Async Msgs (expire:50 force:0 sync:4 revoke:0)
   NumMsgExecuted = 715
   NumHomeconn    = 292
   NumHomedisc    = 292
   NumRPOMisses   = 1
   ```

The fields are described in the following table.

Table 36. Field description of the example

Field name	Description
BytesWritten	The amount of data synchronized to home.
BytesToWrite	The amount of data in queue.
QueueDelay	The maximum delay experienced by operations before sync to home.
NumMsgExecuted	The number of operations executed at home.
NumHomeconn	The number of times home reconnected after disconnection.
NumHomedisc	The number of times home disconnected.
NumRPOMisses	Related to RPOs for AFM primary fileset.

Monitoring using mmperfmon

You can use **mmperfmon** to monitor AFM and AFM DR.

Complete the following steps to enable Performance Monitoring tool and query data.

Note: Ensure that monitoring is initialized, performance monitoring is enabled, and other sensors are collecting data.

1. Run the following command to configure the gateway nodes as performance monitoring nodes: **mmcrnodeclass afmGateways -N gw1,gw2**.
2. Set perfmon designation for the gateway nodes: **mmchnode –perfmon -N afmGateways**.
3. Enable the monitoring tool on the gateway nodes to set the collection periods to 10 or higher: **mmperfmon config update GPFSAFM.period=10 GPFSAFMFS.period=10 GPFSAFMFSET.period=10**
4. Restrict the gateway nodes to collect AFM data: **mmperfmon config update GPFSAFM.restrict=afmGateways GPFSAFMFS.restrict=afmGateways GPFSAFMFSET.restrict=afmGateways**
5. Run the query to display time series data: **mmperfmon query gpfs_afm_fset_bytes_written --bucket-size 60 --number-buckets 1 -N gw1**

   The system displays output similar to - `Legend: 1: gw1|GPFSAFMFSET|gpfs0|independentwriter|gpfs_afm_fset_bytes_written Row Timestamp gpfs_afm_fset_bytes_written 1 2017-03-10-13:28:00 133546`

   Note: You can use the GUI or the Grafana bridge to query collected data.

Monitoring prefetch

You can display the status of an AFM prefetch request by running the **mmafmctl prefetch** command without the `list-file` option.

For example, for file system gpfs1 and fileset iw_1, run the following command:

```
#  mmafmctl gpfs1  prefetch -j iw_1

Fileset Name    Async Read (Pending) Async Read (Failed) Async Read (Already Cached) Async Read(Total)
Async Read (Data in Bytes)
------------    ------------------------------------   ------------------ ------------------------
  iw_1           11             0             0                   11                0
```

This output displays that there are 11 inodes that must be prefetched Async Read (Pending). When the job has completed, the status command displays:

```
#  mmafmctl gpfs1  prefetch -j iw_1
 Fileset Name    Async Read (Pending) Async Read (Failed) Async Read (Already Cached) Async Read(Total)
 Async Read (Data in Bytes)
 ------------    ------------------------------------  ------------------ ------------------------
  iw_1           0                  0                   10                          11
```

Monitoring status using `mmdiag`

You can use the mmdiag command to monitor AFM and AFM DR in the following ways:

- Use the following **mmdiag --afm** command to display all active AFM-relationships on a gateway node:
  `# mmdiag --afm`

  The system displays output similar to -

```
=== mmdiag: afm ===
 AFM Gateway: fin23p Active

 AFM-Cache:  fileset_2 (/cache_fs0/fs2) in Device: cache_fs0
    Mode: independent-writer
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs0)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 603
 AFM-Cache:  fileset1 (/cache_fs0/fs1) in Device: cache_fs0
    Mode: single-writer
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs1)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 2
 AFM-Cache:  fileset1 (/test_cache/fs1) in Device: test_cache
```

```
    Mode: read-only
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs2)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 3
[root@fin23p ~]# mmdiag --afm -Y
mmdiag:afm_fset:HEADER:version:reserved:reserved:cacheName:cachePath:deviceName
:cacheMode:HomeNode:HomePath:filesetStatus:handlerState:cacheState:qState:qLen:qNumExec
mmdiag:afm_gw:HEADER:version:reserved:reserved:gwNode:gwActive:gwDisconn
:Recov:Resync:NodeChg:QLen:QMem:softQMem:hardQMem:pingState
mmdiag:afm_gw:0:1:::fin23p:Active:::::::
mmdiag:afm_fset:0:1:::fileset_2:/cache_fs0/fs2:cache_fs0:independent-writer
:fin21p:nfs%3A//fin21p/test_fs0/cache_fs0:Linked:Mounted:Active:Normal:0:603:
mmdiag:afm_fset:0:1:::fileset1:/cache_fs0/fs1:cache_fs0:single-writer
:fin21p:nfs%3A//fin21p/test_fs0/cache_fs1:Linked:Mounted:Active:Normal:0:2:
mmdiag:afm_fset:0:1:::fileset1:/test_cache/fs1:test_cache:read-only
:fin21p:nfs%3A//fin21p/test_fs0/cache_fs2:Linked:Mounted:Active:Normal:0:3:
```

- Use the following **mmdiag --afm** command to display only the specified fileset's relationship:
  `# mmdiag --afm fileset=cache_fs0:fileset_2`

  The system displays output similar to -

```
 === mmdiag: afm ===
 AFM-Cache:  fileset_2 (/cache_fs0/fs2) in Device: cache_fs0
    Mode: independent-writer
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs0)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 603
[root@fin23p ~]# mmdiag --afm fset=cache_fs0:fileset_2 -Y
mmdiag:afm_fset:HEADER:version:reserved:reserved:cacheName:cachePath:deviceName
:cacheMode:HomeNode:HomePath:filesetStatus:handlerState:cacheState:qState:qLen:qNumExec
mmdiag:afm_fset:0:1:::fileset_2:/cache_fs0/fs2:cache_fs0:
independent-writer:fin21p:nfs%3A//fin21p/test_fs0/cache_fs0
:Linked:Mounted:Active:Normal:0:603:
```

- Use the following **mmdiag --afm** command to display detailed gateway-specific attributes:
  `# mmdiag --afm gw`

  The system displays output similar to -

```
  === mmdiag: afm ===
 AFM Gateway: fin23p Active

 QLen: 0 QMem: 0 SoftQMem: 2147483648 HardQMem 5368709120
 Ping thread: Started
[root@fin23p ~]# mmdiag --afm gw -Y
mmdiag:afm_gw:HEADER:version:reserved:reserved:gwNode:gwActive:gwDisconn
:Recov:Resync:NodeChg:QLen:QMem:softQMem:hardQMem:pingState
mmdiag:afm_gw:0:1:::fin23p:Active:::::0:0:2147483648:5368709120:Started
[root@fin23p ~]#
```

- Use the **mmdiag --afm** command to display all active filesets known to the gateway node:
  `# mmdiag --afm fileset=all`

  The system displays output similar to -

```
=== mmdiag: afm ===
 AFM-Cache:  fileset1 (/test_cache/fs1) in Device: test_cache
    Mode: read-only
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs2)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 3
 AFM-Cache:  fileset1 (/cache_fs0/fs1) in Device: cache_fs0
    Mode: single-writer
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs1)
```

```
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 2
 AFM-Cache:  fileset_2 (/cache_fs0/fs2) in Device: cache_fs0
    Mode: independent-writer
    Home: fin21p (nfs://fin21p/test_fs0/cache_fs0)
    Fileset Status: Linked
  Handler-state: Mounted
  Cache-state: Active
  Q-state: Normal Q-length: 0 Q-executed: 603
[root@fin23p ~]# mmdiag --afm fileset=all -Y
mmdiag:afm_fset:HEADER:version:reserved:reserved:cacheName:cachePath:deviceName
:cacheMode:HomeNode:HomePath:filesetStatus:handlerState:cacheState:qState:qLen:qNumExec
mmdiag:afm_fset:0:1:::fileset1:/test_cache/fs1:test_cache
:read-only:fin21p:nfs%3A//fin21p/test_fs0/cache_fs2
:Linked:Mounted:Active:Normal:0:3:
mmdiag:afm_fset:0:1:::fileset1:/cache_fs0/fs1:cache_fs0
:single-writer:fin21p:nfs%3A//fin21p/test_fs0/cache_fs1
:Linked:Mounted:Active:Normal:0:2:
mmdiag:afm_fset:0:1:::fileset_2:/cache_fs0/fs2:cache_fs0
:independent-writer:fin21p:nfs%3A//fin21p/test_fs0/cache_fs0
:Linked:Mounted:Active:Normal:0:603:
```

Policies used for monitoring AFM and AFM DR

You can monitor AFM and AFM DR using some policies and commands.

Following are the policies used for monitoring:

1. The following file attributes are available through the policy engine:

Table 37. Attributes with their description

Attribute	Description
P	The file is managed by AFM and AFM DR.
u	The file is managed by AFM and AFM DR, and the file is fully cached. When a file originates at the home, it indicates that the entire file is copied from the home cluster.
v	A file or a soft link is newly created, but not copied to the home cluster.
w	The file has outstanding data updates.
x	A hard link is newly created, but not copied to the home cluster.
y	A file metadata was changed and the change not copied to the home cluster.
z	A file is local to the cache and is not queued at the home cluster.
j	A file is appended, but not copied to the home cluster. This attribute also indicates complete directories.
k	All files and directories that are not orphan and are repaired.

2. A list of dirty files in the cache cluster:

   This is an example of a LIST policy that generates a list of files in the cache with pending changes that have not been copied to the home cluster.

```
RULE 'listall' list 'all-files' SHOW( varchar(kb_allocated) || ' ' || varchar(file_size) || ' ' ||
          varchar(misc_attributes) || ' ' || fileset_name) WHERE REGEX(misc_attributes,'[P]') AND
          REGEX(misc_attributes,'[w|v|x|y|j]')
```

If there are no outstanding updates, an output file is not created.

3. A list of partially cached files:

   The following example is that of a LIST policy that generates a list of partially-cached files. If the file is in progress, partial caching is enabled or the home cluster becomes unavailable before the file is completely copied.

   ```
   RULE 'listall' list 'all-files'
            SHOW(varchar(kb_allocated) || ' ' || varchar(file_size) || ' ' ||
            varchar(misc_attributes) || ' ' || fileset_name )
            WHERE REGEX(misc_attributes,'[P]') AND NOT REGEX(misc_attributes,'[u]') AND kb_allocated > 0
   ```

   This list does not include files that are not cached. If partially-cached files do not exist, an output file is not created

4. The custom eviction policy:

   The steps to use policies for AFM file eviction are - generate a list of files and run the eviction. This policy lists all the files that are managed by AFM are not accessed in the last seven days.

   ```
   RULE 'prefetch-list'
            LIST 'toevict'
            WHERE CURRENT_TIMESTAMP - ACCESS_TIME > INTERVAL '7' DAYS
            AND REGEX(misc_attributes,'[P]') /* only list AFM managed files */
   ```

   To limit the scope of the policy or to use it on different filesets run **mmapplypolicy** by using a directory path instead of a file system name. `/usr/lpp/mmfs/bin/mmapplypolicy $path -f $localworkdir -s $localworkdir -P $sharedworkdir/${policy} -I defer`

   Use **mmafmctl** to evict the files: **mmafmctl datafs evict --list-file $localworkdir/list.evict**

5. A policy of uncached files:

   a. The following example is of a LIST policy that generates a list of uncached files in the cache directory:

      ```
      RULE EXTERNAL LIST 'u_list'
               RULE 'u_Rule' LIST 'u_list' DIRECTORIES_PLUS FOR FILESET ('sw1') WHERE NOT
               REGEX(misc_attributes,'[u]')
      ```

   b. An example of a LIST policy that generates a list of files with size and attributes belonging to the cache fileset is as under - (`cacheFset1` is the name of the cache fileset in the example.)

      ```
      RULE 'all' LIST 'allfiles' FOR FILESET ('cacheFset1') SHOW( '/' || VARCHAR(kb_allocated)
      || '/' || varchar(file_size) || '/' ||
      VARCHAR(BLOCKSIZE) || '/' || VARCHAR(MISC_ATTRIBUTES) )
      ```

Monitoring AFM and AFM DR using GUI

The **Files** > **Active File Management** page in the IBM Spectrum Scale provides an easy way to monitor the performance, health status, and configuration aspects of the AFM and AFM DR relationships in the IBM Spectrum Scale cluster. It also provides details of the gateway nodes that are part of the AFM or AFM DR relationships.

The following options are available to monitor AFM and AFM DR relationships and gateway nodes:

1. A quick view that gives the details of top relationships between cache and home sites in an AFM or AFM DR relationship. It also provides performance of gateway nodes by used memory and number of queued messages. The graphs that are displayed in the quick view are refreshed regularly. The refresh intervals are depended on the selected time frame. The following list shows the refresh intervals corresponding to each time frame:
   - Every minute for the 5 minutes time frame
   - Every 15 minutes for the 1 hour time frame

- Every 6 hours for the 24 hours time frame
- Every two days for the 7 days time frame
- Every seven days for the 30 days time frame
- Every four months for the 365 days time frame

2. Different performance metrics and configuration details in the tabular format. The following tables are available:

   Cache Provides information about configuration, health, and performance of the AFM feature that is configured for data caching and replication.

   Disaster Recovery
   Provides information about configuration, health, and performance of AFM DR configuration in the cluster.

   Gateway Nodes
   Provides details of the nodes that are designated as the gateway node in the AFM or AFM DR configuration.

   To find an AFM or AFM DR relationship or a gateway node with extreme values, you can sort the values that are displayed on the table by different attributes. Click the performance metric in the table header to sort the data based on that metric. You can select the time range that determines the averaging of the values that are displayed in the table and the time range of the charts in the overview from the time range selector, which is placed in the upper right corner. The metrics in the table do not update automatically. The refresh button that is placed above the table allows to refresh the table with more recent data.

3. A detailed view of the performance and health aspects of the individual AFM or AFM DR relationship or gateway node. To see the detailed view, you can either double-click the row that lists the relationship or gateway node of which you need to view the details or select the item from the table and click **View Details**. The following details are available for each item:

   Cache

   - **Overview:** Provides number of available cache inodes and displays charts that show the amount of data that is transferred, data backlog, and memory used for the queue.
   - **Events:** Provides details of the system health events reported for the AFM component.
   - **Snapshots:** Provides details of the snapshots that are available for the AFM fileset.
   - **Gateway Nodes:** Provides details of the nodes that are configured as gateway node in the AFM configuration.

   Disaster Recovery

   - **Overview:** Provides number of available primary inodes and displays charts that show the amount of data that is transferred, data backlog, and memory used for the queue.
   - **Events:** Provides details of the system health events reported for the AFM component.
   - **Snapshots:** Provides details of the snapshots that are available for the AFM fileset.
   - **Gateway Nodes:** Provides details of the nodes that are configured as gateway node in the AFM configuration.

   Gateway Nodes
   The details of gateway nodes are available under the following tabs:

   - **Overview** tab provides performance chart for the following:
     - Client IOPS
     - Client data rate
     - Server data rate
     - Server IOPS
     - Network
     - CPU

- Load
- Memory

- **Events** tab helps to monitor the events that are reported in the node. Similar to the Events page, you can also perform the operations like marking events as read and running fix procedure from this events view. Only current issues are shown in this view. The Monitoring > Events page displays the entire set of events that are reported in the system.
- **File Systems** tab provides performance details of the file systems that are mounted on the node. File system's read or write throughput, average read or write transactions size, and file system read or write latency are also available.

  Use the **Mount File System** or **Unmount File System** options to mount or unmount individual file systems or multiple file systems on the selected node. The nodes on which the file system need to be mounted or unmounted can be selected individually from the list of nodes or based on node classes.
- **NSDs** tab gives status of the disks that are attached to the node. The NSD tab appears only if the node is configured as an NSD server.
- **SMB** and **NFS** tabs provide the performance details of the SMB and NFS services that are hosted on the node. These tabs appear in the chart only if the node is configured as a protocol node.
- The **AFM** tab provides details of the configuration and status of the AFM and AFM DR relationships for which the node is configured as the gateway node.

  It also displays the number of AFM filesets and the corresponding export server maps. Each export map establishes a mapping between the gateway node and the NFS host name to allow parallel data transfers from cache to home. One gateway node can be mapped only to a single NFS server and one NFS server can be mapped to multiple gateway nodes.
- **Network** tab displays the network performance details.
- **Properties** tab displays the basic attributes of the node and you can use the **Prevent file system mounts** option to specify whether you can prevent file systems from mounting on the node.

Monitoring AFM and AFM DR configuration and performance in the remote cluster

The IBM Spectrum Scale GUI can monitor only a single cluster. If you want to monitor the AFM and AFM DR configuration, health, and performance across clusters, the GUI node of the local cluster must establish a connection with the GUI node of the remote cluster. By establishing a connection between GUI nodes, both the clusters can monitor each other. To enable remote monitoring capability among clusters, the GUI nodes that are communicating each other must be in the same software level.

To establish a connection with the remote cluster, perform the following steps:

1. Perform the following steps on the local cluster raise the access request:
   a. Select the **Request Access** option that is available under the **Outgoing Requests** tab to raise the request for access.
   b. In the **Request Remote Cluster Access** dialog, enter an alias for the remote cluster name and specify the GUI nodes to which the local GUI node must establish the connection.
   c. If you know the credentials of the security administrator of the remote cluster, you can also add the user name and password of the remote cluster administrator and skip step 2.
   d. Click **Send** to submit the request.
2. Perform the following steps on the remote cluster to grant access:
   a. When the request for connection is received in, the GUI displays the details of the request in the **Access** > **Remote Connections** > **Incoming Requests** page.
   b. Select **Grant Access** to grant the permission and establish the connection.

Now, the requesting cluster GUI can monitor the remote cluster. To enable both clusters to monitor each other, repeat the procedure with reversed roles through the respective GUIs.

Note: Only the GUI user with *Security Administrator* role can grant access to the remote connection requests.

When the remote cluster monitoring capabilities are enabled, you can view the following remote cluster details in the local AFM GUI:

- On home and secondary, you can see the AFM relationships configuration, health status, and performance values of the Cache and Disaster Recovery grids.
- On the Overview tab of the detailed view, the available home and secondary inodes are available.
- On the Overview tab of the detailed view, the details such as NFS throughput, IOPs, and latency details are available, if the protocol is NFS.

The performance and status information on gateway nodes are not transferred to home.

Creating and deleting peer and RPO snapshots through GUI

When a peer snapshot is taken, it creates a snapshot of the cache fileset and then queues a snapshot creation at the home site. This ensures application consistency at both cache and home sites. The recovery point objective (RPO) snapshot is a type of peer snapshot that is used in the AFM DR setup. It is used to maintain consistency between the primary and secondary sites in an AFM DR configuration.

Use the **Create Peer Snapshot** option in the **Files** > **Snapshots** page to create peer snapshots. You can view and delete these peer snapshots from the Snapshots page and also from the detailed view of the **Files** > **Active File Management** page.

Chapter 7. GPFS SNMP support

GPFS supports the use of the SNMP protocol for monitoring the status and configuration of the GPFS cluster. Using an SNMP application, the system administrator can get a detailed view of the system and be instantly notified of important events, such as a node or disk failure.

The Simple Network Management Protocol (SNMP) is an application-layer protocol that facilitates the exchange of management information between network devices. It is part of the Transmission Control Protocol/Internet Protocol (TCP/IP) protocol suite. SNMP enables network administrators to manage network performance, find and solve network problems, and plan for network growth.

SNMP consists of commands to enumerate, read, and write managed variables that are defined for a particular device. It also has a **trap** command, for communicating events asynchronously.

The variables are organized as instances of objects, known as management information bases (MIBs). MIBs are organized in a hierarchical tree by organization (for example, IBM). A GPFS MIB is defined for monitoring many aspects of GPFS.

An SNMP agent software architecture typically consists of a master agent and a set of subagents, which communicate with the master agent through a specific agent/subagent protocol (the AgentX protocol in this case). Each subagent handles a particular system or type of device. A GPFS SNMP subagent is provided, which maps the SNMP objects and their values.

You can also configure SNMP by using the **Settings** > **Event Notifications** > **SNMP Manager** page of the IBM Spectrum Scale management GUI. For more information on the SNMP configuration options that are available in the GUI, see "Configuring SNMP manager" on page 102.

Installing Net-SNMP

The SNMP subagent runs on the collector node of the GPFS cluster. The collector node is designated by the system administrator.

For more information, see "Collector node administration" on page 153.

The Net-SNMP master agent (also called the SNMP daemon, or **snmpd**) must be installed on the collector node to communicate with the GPFS subagent and with your SNMP management application. Net-SNMP is included in most Linux distributions and should be supported by your Linux vendor. Source and binaries for several platforms are available from the download section of the Net-SNMP website (www.net-snmp.org/download.html).

Note: Currently, the collector node must run on the Linux operating system. For an up-to-date list of supported operating systems, specific distributions, and other dependencies, refer to the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

The GPFS subagent expects to find the following shared object libraries:

```
libnetsnmpagent.so       -- from Net-SNMP
libnetsnmphelpers.so     -- from Net-SNMP
libnetsnmpmibs.so        -- from Net-SNMP
libnetsnmp.so            -- from Net-SNMP
libwrap.so               -- from TCP Wrappers
libcrypto.so             -- from OpenSSL
```

Note: TCP Wrappers and OpenSSL are prerequisites and should have been installed when you installed Net-SNMP.

The installed libraries will be found in **/lib64** or **/usr/lib64** or **/usr/local/lib64**. They may be installed under names like **libnetsnmp.so.5.1.2**. The GPFS subagent expects to find them without the appended version information in the name. Library installation should create these symbolic links for you, so you will rarely need to create them yourself. You can ensure that symbolic links exist to the versioned name from the plain name. For example,

```
# cd /usr/lib64
# ln -s libnetsnmpmibs.so.5.1.2 libnetsnmpmibs.so
```

Repeat this process for all the libraries listed in this topic.

Note: For possible Linux platform and Net-SNMP version compatibility restrictions, see the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

Configuring Net-SNMP

The GPFS subagent process connects to the Net-SNMP master agent, **snmpd**.

The following entries are required in the **snmpd** configuration file on the collector node (usually, **/etc/snmp/snmpd.conf**):

```
master agentx
AgentXSocket tcp:localhost:705
trap2sink managementhost
```

where:

managementhost
: Is the host name or IP address of the host to which you want SNMP traps sent.

If your GPFS cluster has a large number of nodes or a large number of file systems for which information must be collected, you must increase the timeout and retry parameters for communication between the SNMP master agent and the GPFS subagent to allow time for the volume of information to be transmitted. The **snmpd** configuration file entries for this are:

```
agentXTimeout 60
agentXRetries 10
```

where:

agentXTimeout
: Is set to 60 seconds for subagent to master agent communication.

agentXRetries
: Is set to 10 for the number of communication retries.

Note: Other values may be appropriate depending on the number of nodes and file systems in your GPFS cluster.

After modifying the configuration file, restart the SNMP daemon.

Configuring management applications

To configure any SNMP-based management applications you might be using (such as Tivoli® NetView® or Tivoli Netcool®, or others), you must make the GPFS MIB file available on the processor on which the management application runs.

You must also supply the management application with the host name or IP address of the collector node to be able to extract GPFS monitoring information through SNMP. To do this, you must be familiar with your SNMP-based management applications.

For more information about Tivoli NetView or Tivoli Netcool, see IBM Knowledge Center (www.ibm.com/support/knowledgecenter).

Installing MIB files on the collector node and management node

The GPFS management information base (MIB) file is found on the collector node in the **/usr/lpp/mmfs/data** directory with the name **GPFS-MIB.txt**.

To install this file on the collector node, do the following:

1. Copy or link the **/usr/lpp/mmfs/data/GPFS-MIB.txt** MIB file into the **SNMP MIB** directory (usually, **/usr/share/snmp/mibs**).

   Alternatively, you could add the following line to the **snmp.conf** file (usually found in the directory **/etc/snmp**):

   ```
   mibdirs +/usr/lpp/mmfs/data
   ```

2. Add the following entry to the **snmp.conf** file (usually found in the directory **/etc/snmp**):

   ```
   mibs +GPFS-MIB
   ```

3. Restart the SNMP daemon.

Different management applications have different locations and ways for installing and loading a new MIB file. The following steps for installing the GPFS MIB file apply only to Net-SNMP. If you are using other management applications, such as NetView and NetCool, refer to corresponding product manuals (listed in "Configuring management applications" on page 152) for the procedure of MIB file installation and loading.

1. Remotely copy the **/usr/lpp/mmfs/data/GPFS-MIB.txt** MIB file from the collector node into the **SNMP MIB** directory (usually, **/usr/share/snmp/mibs**).
2. Add the following entry to the **snmp.conf** file (usually found in the directory **/etc/snmp**):

   ```
   mibs +GPFS-MIB
   ```

3. You might need to restart the SNMP management application. Other steps might be necessary to make the GPFS MIB available to your management application.

Collector node administration

Collector node administration includes: assigning, unassigning, and changing collector nodes. You can also see if a collector node is defined.

To assign a collector node and start the SNMP agent, enter:

```
mmchnode --snmp-agent -N NodeName
```

To unassign a collector node and stop the SNMP agent, enter:

```
mmchnode --nosnmp-agent -N NodeName
```

To see if there is a GPFS SNMP subagent collector node defined, enter:

```
mmlscluster | grep snmp
```

To change the collector node, issue the following two commands:

```
mmchnode --nosnmp-agent -N OldNodeName
```

```
mmchnode --snmp-agent -N NewNodeName
```

Starting and stopping the SNMP subagent

The SNMP subagent is started and stopped automatically.

The SNMP subagent is started automatically when GPFS is started on the collector node. If GPFS is already running when the collector node is assigned, the **mmchnode** command will automatically start the SNMP subagent.

The SNMP subagent is stopped automatically when GPFS is stopped on the node (**mmshutdown**) or when the SNMP collector node is unassigned (**mmchnode**).

The management and monitoring subagent

The GPFS SNMP management and monitoring subagent runs under an SNMP master agent such as Net-SNMP. It handles a portion of the SNMP OID space.

The management and monitoring subagent connects to the GPFS daemon on the collector node to retrieve updated information about the status of the GPFS cluster.

SNMP data can be retrieved using an SNMP application such as Tivoli NetView. NetView provides a MIB browser for retrieving user-requested data, as well as an event viewer for displaying asynchronous events.

Information that is collected includes status, configuration, and performance data about GPFS clusters, nodes, disks, file systems, storage pools, and asynchronous events. The following is a sample of the data that is collected for each of the following categories:

- Cluster status and configuration (see "Cluster status information" on page 155 and "Cluster configuration information" on page 156)
  - Name
  - Number of nodes
  - Primary and secondary servers
- Node status and configuration (see "Node status information" on page 156 and "Node configuration information" on page 156)
  - Name
  - Current status
  - Type
  - Platform
- File system status and performance (see "File system status information" on page 157 and "File system performance information" on page 158)
  - Name
  - Status
  - Total space
  - Free space
  - Accumulated statistics
- Storage pools (see "Storage pool information" on page 158)
  - Name
  - File system to which the storage pool belongs
  - Total storage pool space
  - Free storage pool space
  - Number of disks in the storage pool
- Disk status, configuration, and performance (see "Disk status information" on page 159, "Disk configuration information" on page 159, and "Disk performance information" on page 160)
  - Name
  - Status
  - Total space

  - Free space
  - Usage (metadata/data)
  - Availability
  - Statistics
- Asynchronous events (traps) (see "Net-SNMP traps" on page 160)
  - File system mounted or unmounted
  - Disks added, deleted, or changed
  - Node failure or recovery
  - File system creation, deletion, or state change
  - Storage pool is full or nearly full

Note: If file systems are not mounted on the collector node at the time that an SNMP request is received, the subagent can still obtain a list of file systems, storage pools, and disks, but some information, such as performance statistics, will be missing.

SNMP object IDs

This topic defines the SNMP object IDs.

The management and monitoring SNMP subagent serves the OID space defined as **ibm.ibmProd.ibmGPFS**, which is the numerical **enterprises.2.6.212** OID space.

Underneath this top-level space are the following:

- **gpfsTraps** at **ibmGPFS.0**
- **gpfsMIBObjects** at **ibmGPFS.1**
- **ibmSpectrumScaleGUI** at **ibmGPFS.10**

You can also configure SNMP by using the **Settings** > **Event Notifications** > **SNMP Manager** page of the IBM Spectrum Scale management GUI. The object identifiers (OID) .1.3.6.1.4.1.2.6.212.10.0.1 is sent by the GUI for each event.

MIB objects

gpfsMIBObjects provides a space of objects that can be retrieved using a MIB browser application. Net-SNMP provides the **snmpget**, **snmpgetnext**, **snmptable**, and **snmpwalk** commands, which can be used to retrieve the contents of these fields.

Cluster status information

The following table lists the values and descriptions for the GPFS cluster:

Table 38. gpfsClusterStatusTable: Cluster status information

Value	Description
gpfsClusterName	The cluster name.
gpfsClusterId	The cluster ID.
gpfsClusterMinReleaseLevel	The currently enabled cluster functionality level.
gpfsClusterNumNodes	The number of nodes that belong to the cluster.
gpfsClusterNumFileSystems	The number of file systems that belong to the cluster.

Cluster configuration information

The following table lists the values and descriptions for the GPFS cluster configuration:

Table 39. gpfsClusterConfigTable: Cluster configuration information

Value	Description
gpfsClusterConfigName	The cluster name.
gpfsClusterUidDomain	The UID domain name for the cluster.
gpfsClusterRemoteShellCommand	The remote shell command being used.
gpfsClusterRemoteFileCopyCommand	The remote file copy command being used.
gpfsClusterPrimaryServer	The primary GPFS cluster configuration server.
gpfsClusterSecondaryServer	The secondary GPFS cluster configuration server.
gpfsClusterMaxBlockSize	The maximum file system block size.
gpfsClusterDistributedTokenServer	Indicates whether the distributed token server is enabled.
gpfsClusterFailureDetectionTime	The desired time for GPFS to react to a node failure.
gpfsClusterTCPPort	The TCP port number.
gpfsClusterMinMissedPingTimeout	The lower bound on a missed **ping** timeout (seconds).
gpfsClusterMaxMissedPingTimeout	The upper bound on missed **ping** timeout (seconds).

Node status information

The following table provides description for each GPFS node:

Table 40. gpfsNodeStatusTable: Node status information

Node	Description
gpfsNodeName	The node name used by the GPFS daemon.
gpfsNodeIp	The node IP address.
gpfsNodePlatform	The operating system being used.
gpfsNodeStatus	The node status (for example, up or down).
gpfsNodeFailureCount	The number of node failures.
gpfsNodeThreadWait	The longest hung thread's wait time (milliseconds).
gpfsNodeHealthy	Indicates whether the node is healthy in terms of hung threads. If there are hung threads, the value is no.
gpfsNodeDiagnosis	Shows the number of hung threads and detail on the longest hung thread.
gpfsNodeVersion	The GPFS product version of the currently running daemon.

Node configuration information

The following table lists the collected configuration data for each GPFS node:

Table 41. gpfsNodeConfigTable: Node configuration information

Node	Description
gpfsNodeConfigName	The node name used by the GPFS daemon.
gpfsNodeType	The node type (for example, manager/client or quorum/nonquorum).

Table 41. gpfsNodeConfigTable: Node configuration information (continued)

Node	Description
gpfsNodeAdmin	Indicates whether the node is one of the preferred admin nodes.
gpfsNodePagePoolL	The size of the cache (low 32 bits).
gpfsNodePagePoolH	The size of the cache (high 32 bits).
gpfsNodePrefetchThreads	The number of prefetch threads.
gpfsNodeMaxMbps	An estimate of how many megabytes of data can be transferred per second.
gpfsNodeMaxFilesToCache	The number of inodes to cache for recently-used files that have been closed.
gpfsNodeMaxStatCache	The number of inodes to keep in the stat cache.
gpfsNodeWorker1Threads	The maximum number of worker threads that can be started.
gpfsNodeDmapiEventTimeout	The maximum time the file operation threads will block while waiting for a DMAPI synchronous event (milliseconds).
gpfsNodeDmapiMountTimeout	The maximum time that the mount operation will wait for a disposition for the mount event to be set (seconds).
gpfsNodeDmapiSessFailureTimeout	The maximum time the file operation threads will wait for the recovery of the failed DMAPI session (seconds).
gpfsNodeNsdServerWaitTimeWindowOnMount	Specifies a window of time during which a mount can wait for NSD servers to come up (seconds).
gpfsNodeNsdServerWaitTimeForMount	The maximum time that the mount operation will wait for NSD servers to come up (seconds).
gpfsNodeUnmountOnDiskFail	Indicates how the GPFS daemon will respond when a disk failure is detected. If it is "true", any disk failure will cause only the local node to forcibly unmount the file system that contains the failed disk.

File system status information

The following table shows the collected status information for each GPFS file system:

Table 42. gpfsFileSystemStatusTable: File system status information

Value	Description
gpfsFileSystemName	The file system name.
gpfsFileSystemStatus	The status of the file system.
gpfsFileSystemXstatus	The executable status of the file system.
gpfsFileSystemTotalSpaceL	The total disk space of the file system in kilobytes (low 32 bits).
gpfsFileSystemTotalSpaceH	The total disk space of the file system in kilobytes (high 32 bits).
gpfsFileSystemNumTotalInodesL	The total number of file system inodes (low 32 bits).
gpfsFileSystemNumTotalInodesH	The total number of file system inodes (high 32 bits).
gpfsFileSystemFreeSpaceL	The free disk space of the file system in kilobytes (low 32 bits).
gpfsFileSystemFreeSpaceH	The free disk space of the file system in kilobytes (high 32 bits).

Table 42. gpfsFileSystemStatusTable: File system status information (continued)

Value	Description
gpfsFileSystemNumFreeInodesL	The number of free file system inodes (low 32 bits).
gpfsFileSystemNumFreeInodesH	The number of free file system inodes (high 32 bits).

File system performance information

The following table shows the GPFS file system performance information:

Table 43. gpfsFileSystemPerfTable: File system performance information

Value	Description
gpfsFileSystemPerfName	The file system name.
gpfsFileSystemBytesReadL	The number of bytes read from disk, not counting those read from cache (low 32 bits).
gpfsFileSystemBytesReadH	The number of bytes read from disk, not counting those read from cache (high 32 bits).
gpfsFileSystemBytesCacheL	The number of bytes read from the cache (low 32 bits).
gpfsFileSystemBytesCacheH	The number of bytes read from the cache (high 32 bits).
gpfsFileSystemBytesWrittenL	The number of bytes written, to both disk and cache (low 32 bits).
gpfsFileSystemBytesWrittenH	The number of bytes written, to both disk and cache (high 32 bits).
gpfsFileSystemReads	The number of read operations supplied from disk.
gpfsFileSystemCaches	The number of read operations supplied from cache.
gpfsFileSystemWrites	The number of write operations to both disk and cache.
gpfsFileSystemOpenCalls	The number of file system open calls.
gpfsFileSystemCloseCalls	The number of file system close calls.
gpfsFileSystemReadCalls	The number of file system read calls.
gpfsFileSystemWriteCalls	The number of file system write calls.
gpfsFileSystemReaddirCalls	The number of file system readdir calls.
gpfsFileSystemInodesWritten	The number of inode updates to disk.
gpfsFileSystemInodesRead	The number of inode reads.
gpfsFileSystemInodesDeleted	The number of inode deletions.
gpfsFileSystemInodesCreated	The number of inode creations.
gpfsFileSystemStatCacheHit	The number of stat cache hits.
gpfsFileSystemStatCacheMiss	The number of stat cache misses.

Storage pool information

The following table lists the collected information for each GPFS storage pool:

Table 44. gpfsStgPoolTable: Storage pool information

Value	Description
gpfsStgPoolName	The name of the storage pool.
gpfsStgPoolFSName	The name of the file system to which the storage pool belongs.

Table 44. gpfsStgPoolTable: Storage pool information (continued)

Value	Description
gpfsStgPoolTotalSpaceL	The total disk space in the storage pool in kilobytes (low 32 bits).
gpfsStgPoolTotalSpaceH	The total disk space in the storage pool in kilobytes (high 32 bits).
gpfsStgPoolFreeSpaceL	The free disk space in the storage pool in kilobytes (low 32 bits).
gpfsStgPoolFreeSpaceH	The free disk space in the storage pool in kilobytes (high 32 bits).
gpfsStgPoolNumDisks	The number of disks in the storage pool.

Disk status information

The following table lists the status information collected for each GPFS disk:

Table 45. gpfsDiskStatusTable: Disk status information

Value	Description
gpfsDiskName	The disk name.
gpfsDiskFSName	The name of the file system to which the disk belongs.
gpfsDiskStgPoolName	The name of the storage pool to which the disk belongs.
gpfsDiskStatus	The status of a disk (values: NotInUse, InUse, Suspended, BeingFormatted, BeingAdded, To Be Emptied, Being Emptied, Emptied, BeingDeleted, BeingDeleted-p, ReferencesBeingRemoved, BeingReplaced or Replacement).
gpfsDiskAvailability	The availability of the disk (Unchanged, OK, Unavailable, Recovering).
gpfsDiskTotalSpaceL	The total disk space in kilobytes (low 32 bits).
gpfsDiskTotalSpaceH	The total disk space in kilobytes (high 32 bits).
gpfsDiskFullBlockFreeSpaceL	The full block (unfragmented) free space in kilobytes (low 32 bits).
gpfsDiskFullBlockFreeSpaceH	The full block (unfragmented) free space in kilobytes (high 32 bits).
gpfsDiskSubBlockFreeSpaceL	The sub-block (fragmented) free space in kilobytes (low 32 bits).
gpfsDiskSubBlockFreeSpaceH	The sub-block (fragmented) free space in kilobytes (high 32 bits).

Disk configuration information

The following table lists the configuration information collected for each GPFS disk:

Table 46. gpfsDiskConfigTable: Disk configuration information

Value	Description
gpfsDiskConfigName	The disk name.
gpfsDiskConfigFSName	The name of the file system to which the disk belongs.
gpfsDiskConfigStgPoolName	The name of the storage pool to which the disk belongs.
gpfsDiskMetadata	Indicates whether the disk holds metadata.

Table 46. gpfsDiskConfigTable: Disk configuration information (continued)

Value	Description
gpfsDiskData	Indicates whether the disk holds data.

Disk performance information

The following table lists the performance information collected for each disk:

Table 47. gpfsDiskPerfTable: Disk performance information

Value	Description
gpfsDiskPerfName	The disk name.
gpfsDiskPerfFSName	The name of the file system to which the disk belongs.
gpfsDiskPerfStgPoolName	The name of the storage pool to which the disk belongs.
gpfsDiskReadTimeL	The total time spent waiting for disk read operations (low 32 bits).
gpfsDiskReadTimeH	The total time spent waiting for disk read operations (high 32 bits).
gpfsDiskWriteTimeL	The total time spent waiting for disk write operations in microseconds (low 32 bits).
gpfsDiskWriteTimeH	The total time spent waiting for disk write operations in microseconds (high 32 bits).
gpfsDiskLongestReadTimeL	The longest disk read time in microseconds (low 32 bits).
gpfsDiskLongestReadTimeH	The longest disk read time in microseconds (high 32 bits).
gpfsDiskLongestWriteTimeL	The longest disk write time in microseconds (low 32 bits).
gpfsDiskLongestWriteTimeH	The longest disk write time in microseconds (high 32 bits).
gpfsDiskShortestReadTimeL	The shortest disk read time in microseconds (low 32 bits).
gpfsDiskShortestReadTimeH	The shortest disk read time in microseconds (high 32 bits).
gpfsDiskShortestWriteTimeL	The shortest disk write time in microseconds (low 32 bits).
gpfsDiskShortestWriteTimeH	The shortest disk write time in microseconds (high 32 bits).
gpfsDiskReadBytesL	The number of bytes read from the disk (low 32 bits).
gpfsDiskReadBytesH	The number of bytes read from the disk (high 32 bits).
gpfsDiskWriteBytesL	The number of bytes written to the disk (low 32 bits).
gpfsDiskWriteBytesH	The number of bytes written to the disk (high 32 bits).
gpfsDiskReadOps	The number of disk read operations.
gpfsDiskWriteOps	The number of disk write operations.

Net-SNMP traps

Traps provide asynchronous notification to the SNMP application when a particular event has been triggered in GPFS. The following table lists the defined trap types:

Table 48. Net-SNMP traps

Net-SNMP trap type	This event is triggered by:
Mount	By the mounting node when the file system is mounted on a node.
Unmount	By the unmounting node when the file system is unmounted on a node.
Add Disk	By the file system manager when a disk is added to a file system on a node.
Delete Disk	By the file system manager when a disk is deleted from a file system.
Change Disk	By the file system manager when the status of a disk or the availability of a disk is changed within the file system.
SGMGR Takeover	By the cluster manager when a file system manager takeover is successfully completed for the file system.
Node Failure	By the cluster manager when a node fails.
Node Recovery	By the cluster manager when a node recovers normally.
File System Creation	By the file system manager when a file system is successfully created.
File System Deletion	By the file system manager when a file system is deleted.
File System State Change	By the file system manager when the state of a file system changes.
New Connection	When a new connection thread is established between the events exporter and the management application.
Event Collection Buffer Overflow	By the collector node when the internal event collection buffer in the GPFS daemon overflows.
Hung Thread	By the affected node when a hung thread is detected. The GPFS Events Exporter Watchdog thread periodically checks for threads that have been waiting for longer than a threshold amount of time.
Storage Pool Utilization	By the file system manager when the utilization of a storage pool becomes full or almost full.

Chapter 8. Monitoring the IBM Spectrum Scale system by using call home

The call home feature collects files, logs, traces, and details of certain system health events from different nodes and services.

Understanding call home

The **mmcallhome** command provides options to configure, enable, run, schedule, and monitor call home related tasks in the IBM Spectrum Scale cluster. Information from each node within a call home group is collected and securely uploaded to the IBM ECuRep server.

Call home groups help to distribute the data-gather and data-upload workload to prevent bottlenecks. You can create groups of any size between one and the number of nodes in you cluster. The larger the group is, the higher is the workload on the callhome node. It is recommended to limit the group size to 32 nodes. Larger groups are also possible but it might result in performance issues.

The following figure depicts the basic call home group structure.

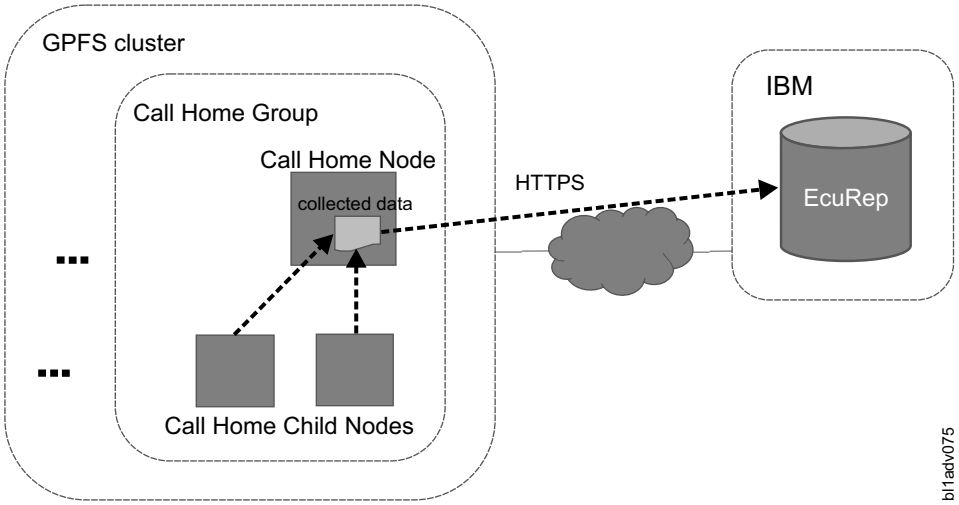

Figure 7. Call home architecture

Call home group
: A group of nodes configured by using the **mmcallhome group** command. A call home group consists of at least one child node, which also acts as its call home node. A call home group can have more than one child node, but has only one call home node. Multiple call home groups can be configured within a IBM Spectrum Scale cluster. You can automate the call home group creation by using the **mmcallhome group auto** command.

Call home node
: This node performs the data upload. If scheduled data gathering is enabled, this node initiates the data collection within its call home group and uploads the data package to IBM support center. A gather-send task process that runs on the call home node collects data from the child nodes and uploads the data to a specific IBM server. This server then sends the data to the IBM backend, ECuRep (Enhanced Customer Data Repository). For more information, see ECuRep. The gather-send configuration file includes information about the data collected from the child nodes.

Note: The call home node is also a child node of the group. If the call home node becomes unavailable, the whole call home group cannot perform any data uploads until the node is online again.

Important: The call home node needs to have access to the external network via port 443. For more information on network-related requirements, see the *Installing call home* section in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Call home child node
A child node is a member of a call home group. The call home node can collect data from all the call home child nodes in a call home group.

Note: If a call home child node, which is not a call home node, becomes unavailable, the rest of its call home group would continue to work properly. Only the details of the unavailable node will be missing in the scheduled uploads.

To configure the call home feature, see "Configuring call home to enable manual and automated data upload."

mmcallhome commands impact

mmcallhome command options react differently when applied to nodes which belong to a call home group and to nodes that do not belong to a call home group:

mmcallhome group, mmcallhome capability, mmcallhome info, and mmcallhome proxy
These commands respond the same when executed on nodes that belong or do not belong to a call home group.

Note: For compatibility reasons, in a mixed cluster configuration the **mmcallhome capability**, **mmcallhome info**, and **mmcallhome proxy** commands only apply to the global settings, if the corresponding nodes are not a part of a call home group. If the call home node of this group also has IBM Spectrum Scale 4.2.3 PTF 6 or older nodes, they also manage a separate settings configuration for their group.

All other mmcallhome command options
All other **mmcallhome** commands can only be run from a node which is a member of a call home group.

For more information on **mmcallhome** command, see the *mmcallhome command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Configuring call home to enable manual and automated data upload

The call home component needs to be configured before it can be used to perform manual and automated data uploads.

The configuration process consists of the following steps:

1. Configure the call home settings.
2. Create call home groups.

After performing these steps, you will be able to use the `mmcallhome run SendFile --file file [--desc DESC | --pmr xxxxx.yyy.zzz]` command to upload a specific file to the ECuRep. Any data collection schedules that have been configured will be regularly executed.

Configuring call home settings

To configure call home settings, perform the following steps:

1. If you are using proxy, configure the proxy settings:
   a. Set the proxy location and authentication settings, using the following command:
      ```
      mmcallhome proxy change --proxy-location ProxyLocation
       –proxy-port ProxyPort [--proxy-username ProxyUsername
       –proxy-password ProxyPassword]
      ```
   b. Enable the proxy using the following command:
      ```
      mmcallhome proxy enable [--with-proxy-auth]
      ```
2. Set up the customer information using the following command:
   ```
   mmcallhome info change --customer-name CustomerName
    --customer-id CustomerId --email Email --country-code CountryCode
   ```
3. Set up the scheduled data collection if needed, using the following commands:
   ```
   mmcallhome schedule add --task DAILY
   mmcallhome schedule add --task WEEKLY
   ```
4. Enable the call home capability using the following command:
   ```
   mmcallhome capability enable
   ```

Creating call home groups

There are two ways to create call home groups: automatically and manually.

Automatic group creation allows users to create homogenous groups, assigning all compatible cluster nodes to one of the groups. Automatic group creation is performed using the **mmcallhome group auto** command. For more information regarding the automatic group creation, see “Configuring the call home groups automatically” on page 166.

Manual group creation allows the users to fine tune the contents of the call home groups. For example, you can make the following changes to the call home groups:

- Assign only a part of the compatible cluster nodes to a call home group.
- Grouping specific nodes within a group.
- Create call home groups with inhomogeneous sizes.
- Change the contents of an existing group without influencing other groups.

Manual group creation is performed using the **mmcallhome group add** command. For more information regarding the automatic group creation, see “Configuring the call home groups manually.”

Configuring the call home groups manually

Manual group creation is meant for specific use cases, where you need to customize the contents for the new groups.

It is performed using the following command:

```
mmcallhome group add GroupName server
 [--node {all | ChildNode[,ChildNode...]}]
```

This assigns the nodes specified via the `--node` option and the node specified as server to the new call home group named `GroupName`. The server node is set to be the call home node of the created group.

All group nodes must have call home packages installed, and the call home node must be able to connect to esupport.ibm.com. If the proxy call home settings are enabled, a proxy is used to test the connectivity. Otherwise a direct connection is attempted.

If the call home node has a release version of 5.0.1.0 or later, the new group is automatically set to track the global call home settings.

Note: If you are using a mixed cluster setup and specify a node with a release version earlier than 5.0.1.0 as the call home node, the created group will have a group-specific configuration with default values, and will have to be manually configured. In such cases, execute the same commands that you have executed to configure global call home settings from one of the nodes of the newly created groups. For more details, see the *Configuring call home settings* section in "Configuring call home to enable manual and automated data upload" on page 164.

Configuring the call home groups automatically

If you want to distribute all compatible cluster nodes into call home groups automatically and create homogenous groups, you must use the **mmcallhome group auto** command after configuring the call home settings.

All compatible nodes that have call home packages and are not yet a part of any call home groups can be redistributed into new call home groups using the **mmcallhome group auto** command. All the newly created call home groups then use the global call home configuration. The following actions are executed in this process:

1. The nodes, that can access esupport.ibm.com and have call home packages installed on them, are detected.

   Note: If a proxy is specified by the **mmcallhome proxy change** command and enabled by the **mmcallhome proxy enable** command, then the specified proxy will be used for detecting the nodes that have access to esupport.ibm.com. If the proxy configuration is disabled, direct connections will be attempted instead.
2. A minimal subset of these nodes is selected, so that all nodes, which are supposed to be distributed into groups, can be distributed into groups with a maximum recommended size of 32 nodes.
3. New groups are created and set to use the global call home settings.

If you want to redistribute all nodes, which are currently assigned to any groups, use the `--force` option as shown. The use of the `--force` option effectively deletes all current groups prior to creating new ones.

```
mmcallhome group auto --force
```

If you want to manually specify the call home nodes to use for the new groups, you can use the `--server` option as shown:

```
mmcallhome group auto [--server {ServerName[,ServerName...]}]
```

In such cases, the following rules apply:

- The number of groups created is the same as the number of the specified call home nodes.
- The access to esupport.ibm.com is not checked for any call home nodes.
- Each group gets one of the specified call home nodes assigned to it.
- All compatible nodes are distributed between these groups

Monitoring, uploading, and sharing collected data with IBM Support

The send file task, that runs on a call home node, uploads files and packages that are collected at this call home node. The call home feature can upload any file to the IBM ECuRep backend.

The call home component uses the directory, specified in the IBM Spectrum Scale settings variable *dataStructureDump*, for saving the temporary data. By default, this directory is `/tmp/mmfs`, but it can be changed by the customer by using the **mmchconfig** command. The current value may be read by executing the following command:

```
mmdiag --config | grep "dataStructureDump"
```

The space requirements differ, depending on the use case. The following two situations are possible:

- For uploading daily and weekly packages and uploading of not PMR-related files the space requirements are three times the size of the file that is to be uploaded. For example, if you need to upload 1 GB of data, then there needs to be at least 3 GB of disk space for the file to be properly uploaded.
- For uploading the PMR-related files introduced in IBM Spectrum Scale version 5.0.0, a dynamical buffered chunking is used. In this case, the minimum space required for an upload is 1/250 of the file size (0.4%), while at least 200 MB is recommended to minimize chunking overhead and maximize the transfer speed.

The call home function has a built-in data collection mechanism that collects pre-defined data on a daily or weekly basis. You can find the definition for this data collection in the `/usr/lpp/mmfs/data/callhome/gather.d` folder in the files `weekly.conf` and `daily.conf`.

The uploaded data is stored for at least two weeks on IBM ECuRep and can be identified using your customer information. If you need to access this data, contact IBM® support. For more information, see ECuRep. The PMR-related data is saved as long as the PMR is still open.

Note: You can also upload data using the following command:

```
mmcallhome run SendFile --file file [--desc DESC | --pmr xxxxx.yyy.zzz]
```

Discuss this procedure with the IBM support before using it.

Important: You can schedule call home uploads for Ubuntu starting with IBM Spectrum Scale version 5.0.0.2. (Previous IBM Spectrum Scale versions do not support scheduled call home uploads for Ubuntu.)

Use the following steps to monitor and analyze the data, and then share it with IBM support:

1. Register the tasks:
   - To register a daily task with cron, issue the **mmcallhome schedule add** command as shown in the following example:

     ```
     mmcallhome schedule add --task daily

     /etc/cron.d/gpfscallhome_GatherSend_daily.conf registered
     41 command entries are defined for this task
     ```

   - To register a weekly task with cron, issue the **mmcallhome schedule add** command as shown in the following example:

     ```
     mmcallhome schedule add --task weekly

     /etc/cron.d/gpfscallhome_GatherSend_weekly.conf registered
     14 command entries are defined for this task
     ```

2. Monitor the tasks.
   - To monitor the call home tasks, issue the **mmcallhome status list** command as shown in the following example:

     ```
     mmcallhome status list

     Task   Start time          Status        Package file name
     daily  20150930132656.582  success       ...aultDaily.g_daily.20150930132656582.cl0.DC
     daily  20150930133134.802  success       ...aultDaily.g_daily.20150930133134802.cl0.DC
     daily  20150930133537.509  success       ...aultDaily.g_daily.20150930133537509.cl0.DC
     daily  20150930133923.063  success       ...aultDaily.g_daily.20150930133923063.cl0.DC
     RunSendFile 20150930133422.843 success
     ...group2.MyTestData.s_file.20150930133422843.cl0.DC
     ```

   - To view the status of the currently running and the already completed call home tasks, issue the **mmcallhome status list** command as shown in the following example:

     ```
     mmcallhome status list --verbose
     ```

```
Task   Start time          Updated time          Status       RC or Step
       Package file name
       [ additional info: value ]
-------------------------------------------------------------------------------
       31790849425327.4_2_1_0.x.abc.autoGroup_1.gat_weekly.g_weekly.20160412160447854.cl0.DC
-------------------------------------------------------------------------------
       31790849425327.4_2_1_0.x.abc.autoGroup_1.gat_weekly.g_weekly.20160412173941161.cl0.DC
-------------------------------------------------------------------------------
weekly  20160412174030.803  20160412174034        failed       RC=6 (lock err) NA
-------------------------------------------------------------------------------
        31790849425327.4_2_1_0.x.abc.autoGroup_1.gat_weekly.g_weekly.20160412175159390.cl0.DC
```

Note: Sometimes, the output of `mmcallhome status list --verbose` displays a single line without detailed information about RC indicating successful completion of call home tasks. The failed status indicates an issue with the call home task and the RC numeral indicates the respective issue. If the value of RC is zero, it indicates that the upload procedure is successful, but some automatically resolvable issue occurred while uploading the data. The value, RC != 0, indicates that the upload procedure is not successful. The detailed information about the upload procedure is available in the logs at `/var/mmfs/callhome/logs/`.

- To list the registered tasks for gather-send, issue the **mmcallhome schedule list** command as shown in the following example:

```
mmcallhome schedule list

Registered Tasks for GatherSend:
ConfFile               CronParameters
daily.conf             3 2 * * *
weekly.conf            54 3 * * sun
```

Note: The CronParameter indicates the date and time settings for the execution of the command. It displays the values for minutes (0-59), hours (0-23), day of month (1-31), month (1-12 or Jan-Dec), and day of week (0-6, where sun=0 or sun-sat). For example, `CronParameter 54 3 * * sun` indicates that the command runs on every Sunday at 3:54 AM. By default, call home schedules **daily** task to be executed at 02:xx AM each day, and the **weekly** task to be executed at 03:yy AM each Sunday, where xx and yy are random numbers from 00 to 59. These values may be changed if necessary by editing the corresponding *.conf files, but it is recommended that you contact the support or development team before making these changes. For more details, see crontab(5) - Linux man page.

3. Upload the collected data. The call home functionality provides the following data upload methods to collect and upload the data:
   a. **File upload:** Any file can be specified for upload.
   b. **Package upload:** Collects predefined data package regularly. The call home feature provides *weekly.conf* schema to collect the package weekly and *daily.conf* schema to collect the package daily. These gather schemas are at: `/usr/lpp/mmfs/data/callhome/gather.d`. After the upload, the data packages are stored in the data package directory for backup.

Attention: This upload is done internally by the call home function based on the type of call home function that is registered in step 1. Every time call home collects data or the call home command is started to upload a specific file, call home first creates a data package file. The data package file is stored in the directory `/tmp/mmfs/callhome` as a tar file. This tar file is deleted once the data is uploaded successfully to ECuRep. In case the upload was not successful, old data (undeleted tar files) from a weekly or daily gather task will be uploaded together with the new data. This data will be available till the upload is successful, or till the data package file is deleted manually.

If the data collection is specified weekly, Cron is started once a week, and data from call home child node is gathered by the call home node as specified in the *weekly.conf* file. When the gather task is finished, the data is uploaded from the call home node to the IBM Support. The following commands are issued internally to generate the data that needs to be shared with IBM Support:

- lscpu
- cat /proc/interrupts
- lsblk

- /usr/lpp/mmfs/bin/mmdiag --config -Y
- /usr/lpp/mmfs/bin/mmdiag --version -Y
- mmdiag --/usr/lpp/mmfs/bin/mmlslicense -L -Y
- /usr/lpp/mmfs/bin/mmlsnodeclass --system -Y
- For each file system: tsstatus && /usr/lpp/mmfs/bin/mmdf <fs> -Y
- rpm - qi <all spectrum scale rpms>
- dpkg-query --list <all spectrum scale rpms>
- /usr/lpp/mmfs/bin/mmlscluster --cnfs
- /usr/lpp/mmfs/bin/mmlspool <fs>
- mmpmon ver,io_s,fs_io_s,vio_s
- numactl --hardware
- numastat -nm
- lspci -vv
- /usr/lpp/mmfs/bin/mmces service list --verbose -Y
- /usr/lpp/mmfs/bin/mmobj s3 list
- /usr/lpp/mmfs/bin/mmobj multiregion list -Y
- /usr/lpp/mmfs/bin/mmobj policy list -Y
- /usr/lpp/mmfs/bin/mmobj config list --ccrfile proxy-server.conf --section pipeline:main
- count of accounts, containers, objects
- object.callhome.py:
- swift configuration files located under /etc/swift
- /usr/lpp/mmfs/bin/mmces address list --attribute object_database_node
- keystone config: http://<object_database_node>:35357/
- gui.callhome.py <lcdir> (GUI activity log)
- network.callhome.py <lcdir>
- ethtool <for any device in /sys/class/net>
- ethtool -g <for any device in /sys/class/net>
- /usr/lpp/mmfs/bin/mmuserauth service list -Y
- cat /etc/os-release

If the data collection is specified daily, Cron is started once every day and data from call home child node is gathered by the call home node as specified in the *daily.conf* file. When the gather task is finished, the data is uploaded from the call home node to the IBM Support. The following commands are issued internally to generate the data that needs to be shared with IBM Support:

- /usr/lpp/mmfs/bin/mmces address list -Y
- /usr/lpp/mmfs/bin/mmlsnode -a
- /usr/lpp/mmfs/bin/mmlsmgr -Y
- /usr/lpp/mmfs/bin/mmlscallback -Y
- /usr/lpp/mmfs/bin/mmremotecluster show all -Y
- /usr/lpp/mmfs/bin/mmremotefs show all -Y
- /usr/lpp/mmfs/bin/mmauth show -Y
- /usr/lpp/mmfs/bin/tsstatus
- /usr/lpp/mmfs/bin/mmsdrquery 40 all
- For each file system: tsstatus && mmlsfs gpfs1 -aABdDEfiIkKLnoPQStTuVwYz --create-time --encryption --fastea --filesetdf --inode-limit --is4KAligned --log-replicas --mount-priority --perfileset --rapid-repair --write-cache-threshold --striped-logs --fileset-count --afm --snc --uid --snapid
- For each file system: tsstatus && /usr/lpp/mmfs/bin/mmlspolicy <fs> -L -Y

- For each file system: tsstatus && /usr/lpp/mmfs/bin/mmlsfileset <fs> -L -Y
- For each file system: tsstatus && /usr/lpp/mmfs/bin/mmlssnapshot <fs> -Y
- Files under /var/mmfs/ccr/committed
- route -n
- /proc/net/def
- ip a
- network.callhome.py <lcdir>
- ethtool <for any device in /sys/class/net>
- ethtool -g <for any device in /sys/class/net>
- gui.callhome.py <lcdir> (GUI activity log)
- sysctl -a
- uptime
- "df -kl" ensure timeout is set
- lsmod
- cat /proc/device-tree/system-id
- ppc64_cpu --smt;ppc64_cpu --cores-present;ppc64_cpu --cores-on
- cat /proc/cpuinfo
- cat /proc/meminfo
- dmidecode
- ofed_info -s
- ibstat
- ibdev2netdev
- uname -a
- ls -1 /sys/class/sas_host | sort -k1.5 -n | while read a ; do echo "# $a"; cat /sys/class/sas_host/${a}/device/scsi_host/${a}/version* ; done
- /usr/lpp/mmfs/bin/mmdiag --memory -Y
- /usr/lpp/mmfs/bin/mmdiag --health -Y/
- usr/lpp/mmfs/bin/mmdiag --lroc -Y
- /usr/lpp/mmfs/bin/mmdiag --commands -Y
- /usr/lpp/mmfs/bin/mmdiag --waiters -Y
- /usr/lpp/mmfs/bin/mmdiag --deadlock -Y
- /usr/lpp/mmfs/bin/mmdiag --network -Y
- /usr/lpp/mmfs/bin/mmdiag --nsd -Y
- /usr/lpp/mmfs/bin/mmdiag --afm -Y
- /usr/lpp/mmfs/bin/mmdiag --stats -Y
- /usr/lpp/mmfs/bin/mmdiag --tokenmgr -Y
- /usr/lpp/mmfs/bin/mmhealth cluster show --verbose -Y
- netstat -s

If a call home group is configured to upload data to IBM support, the various components that are running on a node of this group can upload the files. The **mmhealth** command collects and uploads data using the **mmcallhome** command for the following events:

- nfsd_down
- ctdb_down
- ctdb_state_down
- smbd_down

4. Share the collected information with IBM support.

   The call home feature allows data upload in following two ways:

   a. **Manual upload:** The call home feature provides manual upload option to upload the files or packages manually to the IBM server. To upload any data manually, issue mmcallhome run in one of the following way:

      - To manually initiate the daily data upload:

        ```
        mmcallhome run gather send --task daily
        ```

      - To manually upload a specific file you can use one of the following commands:
        - `mmcallhome run SendFile --file myfile`

          This command will upload the file, and put it into a common directory for all customers, available for support. For this type of upload:

          - Support must be specifically told that there is a file to consider.
          - All support people can read the file.
        - `mmcallhome run SendFile --file myFile --pmr 12345.678.910`

          This command will upload the file to a specific location, linked to the specified PMR. For this type of upload:

          - Support will automatically notice changes for a PMR, if they are working on it.
          - Only the support representatives who are allowed to work on this PMR can read the contents.

   b. **Automatic upload:** Use the **mmcallhome schedule** command to schedule a weekly or daily schema to upload the predefined data. If system health detects a specific event it will collect the data, and upload the data using the following command:

      ```
      mmcallhome run SendFile --file file
      ```

      This is only possible if the node where the system health process is running is a member of an enabled group.

   The manual and automatic upload options can upload the data to the IBM ECuRep. This data is not analyzed automatically, and gets deleted after a specified time period (generally, 2 weeks). Please contact the IBM Support for more information about the usage of the uploaded data.

Configuring call home using GUI

The call home feature provides a communication channel that automatically notifies the IBM service personnel about the issues reported in the system. You can also manually upload diagnostic data files and associate them with a PMR through the GUI.

You can use the Call Home page in the GUI to perform the following tasks:

- Enable call home feature on the cluster.
- Select one or more call home nodes that share the data with the IBM Support.
- Specify the contact information to be used by the IBM Support if there are any issues.
- Specify the proxy information that is needed to create a communication channel between the call home nodes and IBM support.
- Test connection with the IBM server.

Collecting data and sharing it with IBM Support

The call home shares support information and your contact information with IBM on a schedule basis. The IBM Support monitors the details that are shared by the call home and takes necessary action in case of any issues or potential issues. Enabling call home reduces the response time for the IBM Support to address the issues. The call home automatically shares data with the IBM support based on a schedule. The GUI does not support to change the data gathering and sharing schedules.

You can also manually upload the diagnostic data that is collected through the **Settings** > **Diagnostic Data** page in the GUI to share the diagnostic data to resolve a Problem Management Record (PMR). To upload data manually, perform the following steps:

1. Go to **Settings** > **Diagnostic Data**.
2. Collect diagnostic data based on the requirement. You can also use the previously collected data for the upload.
3. Select the relevant data set from the **Previously Collected Diagnostic Data** section and then right-click and select **Upload to PMR**.
4. Select the PMR to which the data must be uploaded and then click **Upload**.

Call home configuration examples

The following section gives some examples of the call home configuration.

Until IBM Spectrum Scale 4.2.3.7, each call home group had its own configuration, which had to be configured and managed separately. All call home nodes with the release version between 4.2.3.7 to 5.0.0.x are set to use global call home settings, after the first change of the corresponding setting from the default value. The change of values happens automatically if the groups are created using the **mmcallhome group auto** command.

For all call home nodes with the release version of 5.0.1.0, the call home nodes are automatically set to use the global call home configuration upon their creation.

For the following use cases we assume the following customer information:

- Customer name: User1
- Customer ID: 123456
- E-mail: customer@ibm.com
- Country-code: JP

Use Case 1: To automatically create call home groups for all Linux nodes in the cluster where call home packages are installed, and enable all call home features.

Note: For this use case, we assume the following:

- Call home has not been configured before.
- Some of the nodes have a direct connectivity to esupport.ibm.com.
- Automatic daily and weekly data collection is to be enabled.

1. Set the customer information:

```
[root@g5001-21 ~]# mmcallhome info change --customer-name
User1 --customer-id 123456 --email customer@ibm.com --country-code JP
Call home country-code has been set to JP
Call home customer-name has been set to User1
Call home customer-id has been set to 123456
Call home email has been set to customer@ibm.com
```

2. Enable the daily and weekly schedule:

```
[root@g5001-21 ~]# mmcallhome schedule add --task DAILY
Call home daily has been set to enabled
[root@g5001-21 ~]# mmcallhome schedule add --task WEEKLY
Call home weekly has been set to enabled
```

3. Enable call home to actually send data:

```
[root@g5001-21 ~]# mmcallhome capability enable
By accepting this request, you agree to allow IBM
and its subsidiaries to store and use your contact information
and your support information anywhere they do business worldwide.
For more information, please refer to the Program license
```

```
agreement and documentation. If you agree, please respond
with "accept" for acceptance, else with "not accepted" to decline.
(accept / not accepted)
accept
Call home enabled has been set to true

Additional messages:
License was accepted. Callhome enabled.
```

4. Create the call home groups automatically:

```
[root@g5001-21 ~]# mmcallhome group auto
Analysing cluster: [I] In progress: Collect group information.
mmcallhautoconfig: [I] In progress: Create 1 new call home groups.
mmcallhautoconfig: [I] In progress: Nodes without call home: 0
See /var/mmfs/tmp/callhome/log/callhomeutils.log for details.
Analysing cluster: [I] In progress: Collect group information.
group: autoGroup_1 successfully added
mmcallhome: [I] deploy configuration.
Success
```

Use Case 2: To distribute all Linux nodes in the cluster where call home packages are installed into two call home groups, and set the nodes g5001-21 and g5001-22 as their call home nodes.

Note: For this use case, we assume the following:

- Call home has not been configured before.
- Both the call home nodes require an authenticated proxy.
- Enable only weekly data collection.

1. Set the customer information:

```
[root@g5001-21 ~]# mmcallhome info change --customer-name
User1 --customer-id 123456 --email customer@ibm.com --country-code JP
Call home country-code has been set to JP
Call home customer-name has been set to User1
Call home customer-id has been set to 123456
Call home email has been set to customer@ibm.com
```

2. Enable the weekly schedule:

```
[root@g5001-21 ~]# mmcallhome schedule add --task WEEKLY
Call home weekly has been set to enabled
```

3. Define the proxy settings and enable proxy:

```
[root@g5001-21 ~]# mmcallhome proxy change
--proxy-location 192.168.0.10 --proxy-port 5085
--proxy-username clusteradmin --proxy-password MyPass
Call home proxy-port has been set to 5085
Call home proxy-username has been set to clusteradmin
Call home proxy-password has been set to MyPass
Call home proxy-location has been set to 192.168.0.10
[root@g5001-21 ~]# mmcallhome proxy enable --with-proxy-auth
Call home proxy-enabled has been set to true
Call home proxy-auth-enabled has been set to true
```

4. Enable call home to send data:

```
[root@g5001-21 ~]# mmcallhome capability enable
By accepting this request, you agree to allow
IBM and its subsidiaries to store and use your
contact information and your support information
anywhere they do business worldwide. For more
information, please refer to the Program license
agreement and documentation. If you agree, please
respond with "accept" for acceptance, else with
"not accepted" to decline.
(accept / not accepted)
accept
```

```
Call home enabled has been set to true

Additional messages:
License was accepted. Callhome enabled.
```

5. Create the call home groups automatically, while specifying the call home nodes:

```
[root@g5001-21 ~]# mmcallhome group auto --server g5001-21,g5001-22
Analysing cluster: [I] In progress: Collect group information.
mmcallhautoconfig: [I] In progress: Create 2 new call home groups.
mmcallhautoconfig: [I] In progress: Nodes without call home: 0
See /var/mmfs/tmp/callhome/log/callhomeutils.log for details.
Analysing cluster: [I] In progress: Collect group information.
group: autoGroup_1 successfully added
Analysing cluster: [I] In progress: Collect group information.
group: autoGroup_2 successfully added
mmcallhome: [I] deploy configuration.
Success
```

Use Case 3: To automatically create call home groups for all Linux nodes in the cluster where call home packages are installed, but disable the scheduled data collection.

Note: For this use case, we assume the following:

- Call home has been configured before, but must be reconfigured. Ensure that the old settings are removed, and the old groups are deleted.
- Both the call home nodes require an authenticated proxy.
- Neither weekly nor daily data collection must be enabled, as data upload is only be done on demand. For example, in case of PMRs.

1. Set the customer information:

```
[root@g5001-21 ~]# mmcallhome info change --customer-name
User1 --customer-id 123456 --email customer@ibm.com --country-code JP
Call home country-code has been set to JP
Call home customer-name has been set to User1
Call home customer-id has been set to 123456
Call home email has been set to customer@ibm.com
```

2. Disable the proxy configuration:

```
[root@g5001-21 ~]# mmcallhome proxy disable
Call home proxy-enabled has been set to false
Call home proxy-auth-enabled has been set to false
```

3. Disable the task schedule:

```
[root@g5001-21 ~]# mmcallhome schedule delete --task DAILY
Call home daily has been set to disabled
[root@g5001-21 ~]# mmcallhome schedule delete --task WEEKLY
Call home weekly has been set to disabled
```

4. Enable call home to send data when needed:

```
[root@g5001-21 ~]# mmcallhome capability enable
By accepting this request, you agree to allow IBM and its
subsidiaries to store and use your contact information
and your support information anywhere they do business worldwide.
For more information, please refer to the Program license agreement
and documentation. If you agree, please respond with "accept" for
acceptance, else with "not accepted" to decline.
(accept / not accepted)
accept
Call home enabled has been set to true

Additional messages:
License was accepted. Callhome enabled.
```

5. Create the call home groups automatically, while removing all previously existing groups:

```
[root@g5001-21 ~]# mmcallhome group auto –force
Analysing cluster: [I] In progress: Collect group information.
mmcallhautoconfig: [I] In progress: Create 1 new call home groups.
mmcallhautoconfig: [I] In progress: Nodes without call home: 0
See /var/mmfs/tmp/callhome/log/callhomeutils.log for details.
mmcallhautoconfig: [I] In progress: Delete existing groups.
Call home group autoGroup_1 has been deleted
Call home group autoGroup_2 has been deleted
Analysing cluster: [I] In progress: Collect group information.
Analysing cluster: [I] In progress: Collect group information.
group: autoGroup_1 successfully added
mmcallhome: [I] deploy configuration.
Success
```

Chapter 9. Monitoring remote cluster through GUI

The IBM Spectrum Scale GUI can monitor only a single cluster. To monitor other IBM Spectrum Scale clusters, which is also referred as remote clusters, the GUI node must establish a connection with the GUI node of the other cluster. By establishing a connection between the GUI nodes, both the clusters can monitor the other cluster. To enable remote monitoring capability among clusters, the GUI nodes that are communicating with each other must be in the same software level. The software level of the participating nodes must be 5.0.0 or later.

To establish a connection with the remote cluster, perform the following steps:

1. Perform the following steps on the local cluster to raise the access request:
   a. Select the **Request Access** option that is available under the **Outgoing Requests** tab to raise the request for access.
   b. In the **Request Remote Cluster Access** dialog, enter an alias for the remote cluster name and specify the GUI nodes to which the local GUI node must establish the connection.
   c. If you know the credentials of the security administrator of the remote cluster, you can also add the user name and password of the remote cluster administrator and skip the step 2.
   d. Click **Send** to submit the request.
2. Perform the following steps on the remote cluster to grant access:
   a. When the request for connection is received in, the GUI displays the details of the request in the **Access** > **Remote Connections** > **Incoming Requests** page.
   b. Select **Grant Access** to grant the permission and establish the connection.

Now, the requesting cluster GUI can monitor the remote cluster. To enable both clusters to monitor each other, repeat the procedure with reversed roles through the respective GUIs.

Note: Only the GUI user with *Security Administrator* role can grant access to the remote connection requests.

Monitoring performance of the remote cluster

You can monitor the performance of the remote cluster with the help of performance monitoring tools that are configured in both the remote and local clusters. The performance details collected in the remote cluster is shared with the local cluster using REST APIs.

After establishing the connection with the remote cluster by using the **Access** > **Remote Connections** page, you can access the performance details of the remote cluster from the following GUI pages:

- **Monitoring** > **Statistics**
- **Monitoring** > **Dashboard**
- **Files** > **File Systems**

To monitor performance details of the remote cluster in the Statistics page, you need to create customized performance charts by performing the following steps:

1. Access the edit mode by clicking the icon that is available on the upper right corner of the performance chart and selecting **Edit.**
2. In the edit mode, select the remote cluster to be monitored from the **Cluster** field. You can either select the local cluster or remote cluster from this field.
3. Select **Resource type**. This is the area from which the data is taken to create the performance analysis.

4. Select **Aggregation level**. The aggregation level determines the level at which the data is aggregated. The aggregation levels that are available for selection varies based on the resource type.
5. Select the entities that need to be graphed. The table lists all entities that are available for the chosen resource type and aggregation level. When a metric is selected, you can also see the selected metrics in the same grid and use methods like sorting, filtering, or adjusting the time frame to select the entities that you want to select.
6. Select **Metrics**. Metrics is the type of data that need to be included in the performance chart. The list of metrics that is available for selection varies based on the resource type and aggregation type.
7. Click **Apply** to create the customized chart.

After creating the customized performance chart, you can mark it as favorite charts to get them displayed on the Dashboard page.

If a file system is mounted on the remote cluster nodes, the performance details of such remote cluster nodes are available in the **Remote Nodes** tab of the detailed view of file systems in the **Files** > **File Systems** page.

Chapter 10. Monitoring file audit logging

The following topics describe various ways to monitor file audit logging in IBM Spectrum Scale.

Monitoring the message queue server and ZooKeeper status

Issue the **mmmsgqueue status** command to determine which nodes are running the processes and to see their current state.

```
# mmmsgqueue status

Node                 Contains  Broker   Contains   ZooKeeper
Name                 Broker    Status   ZooKeeper  Status
c6f2bc3n10.gpfs.net  no                 yes        good
c6f2bc3n2.gpfs.net   no                 yes        good
hs22n55.gpfs.net     yes       good     yes        good
hs22n56.gpfs.net     yes       good     no
hs22n95.gpfs.net     yes       good     yes        good
```

For more information, see the *mmmsgqueue command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Displaying the port that the Kafka broker servers are using

Issue the **mmmsgqueue list --servers** to determine the nodes and ports that the Kafka broker servers are running on.

```
# mmmsgqueue list --servers
hs22n55.gpfs.net:9092,hs22n56.gpfs.net:9092,hs22n95.gpfs.net:9092
```

For more information, see the *mmmsgqueue command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Determining the current topic generation number that is being used in the file system

Issue the **mmaudit replicate list -Y** command.

```
# mmaudit replicate list -Y
mmaudit::HEADER:version:RESERVED:RESERVED:RESERVED:auditDeviceName:
clusterID:auditFilesetDeviceName:auditFilesetName:auditRetention
:topicGenNum:RESERVED:RESERVED:RESERVED:RESERVED
mmaudit:::1:::replicate:6372129557625143312:replicate:regressaudit:365:26::::
# mmmsgqueue list --topics
153_6372129557625143312_26_audit
```

The value reported by the *topicGenNum* represents the generation number. This number is used when creating the topic for the file system. The topic is used as a subdirectory to store the file audit logging records. The topic equals `deviceMinorNumber_clusterId_topicGenNum_audit`. You can find the topic that is being used by running the following command:

```
ps -ef | grep consumer | grep file_system_name
```

The **-T** flag shows the topic for that file system.

For more information, see the *mmaudit command* in the *IBM Spectrum Scale: Command and Programming Reference* and the *mmmsgqueue command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Monitoring the consumer status

Issue the **mmaudit** command to determine the current status of your consumer process. You can also stop and restart the consumers on a per node basis.

To determine the consumer status, issue a command similar to the following example:

```
 # mmaudit all consumerStatus -N hs22n56,c6f2bc3n2
Dev Name  Cluster ID                          Num Nodes
newfs     6372129557625143312                 2
          Node Name                           Is Consumer?  Status
          c6f2bc3n2.gpfs.net                  yes           AUDIT_CONS_OK
          Node Name                           Is Consumer?  Status
          hs22n56.gpfs.net                    yes           AUDIT_CONS_OK
```

To stop the consumers on a set of specific nodes, issue:

```
# mmaudit all consumerStop -N hs22n56,c6f2bc3n2
[I] Node: c6f2bc3n2.gpfs.net is a consumer node, and consumer for device: newfs successfully stopped.
[I] Node: hs22n56.gpfs.net is a consumer node, and consumer for device: newfs successfully stopped.
```

To start the consumers on a set of specific nodes, issue:

```
# mmaudit all consumerStart -N hs22n56,c6f2bc3n2
[I] Node: c6f2bc3n2.gpfs.net is a consumer node, and consumer for device: newfs successfully started.
[I] Node: hs22n56.gpfs.net is a consumer node, and consumer for device: newfs successfully started.
```

For more information, see the *mmaudit command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Monitoring file audit logging states

File audit logging is integrated into the system health infrastructure. Alerts are generated for elements of the message queue and the processes that consume the events and create the audit logs.

The following table details the different **FILEAUDITLOG** states and what action (if any) needs to be taken by the administrator.

Table 49. FILEAUDITLOG states

FILEAUDITLOG state	Description	Condition	Administrator's action
auditc_ok	File audit consumer is running.	**Fileauditlogging** is healthy.	None
auditc_initlockauditfile	Failed to indicate to **systemctl** on the successful startup sequence to enable **fileauditlogging** for the indicated file system.	**Fileauditlogging** is in the failed state.	Disable and re-enable auditing using the **mmaudit** command.
auditc_compress	Could not compress the indicated **auditLogFile** for the given file system.	**Fileauditlogging** is still healthy. Indicates to the customer that the attempt to compress the **auditLogFile<...>** failed.	Administrator can manually invoke compression on the **auditLogFile** using the **mmchattr –compression yes auditLogFile<...>** command.

Table 49. FILEAUDITLOG states (continued)

FILEAUDITLOG state	Description	Condition	Administrator's action
auditc_setimmutablity	Could not set immutablity attribute on the indicated **auditLogFile** for the given file system.	**Fileauditlogging** is still healthy. Indicates to the customer that the attempt to set the immutability attribute on the indicated **auditLogFile<....>** failed.	Administrator can manually set the **auditLogFile<...>** to be immutable using the **mmchattr -i yes auditLogFile<...>** command.
auditc_topicsubscription	Failed to subscribe to the indicated topic in the message queue for the given file system.	**Fileauditlogging** is in the failed state.	Check on the **topicName** and initial configuration details using the **mmmsgqueue list --topics** command and then retry the **mmaudit** command.
auditc_offsetfetch	Failed to fetch offset (from Kafka brokers, which serve as the marker from where events are to be consumed) for the given topic for the indicated file system.	**Fileauditlogging** is in the failed state.	Check on the **topicName** using **mmmsgqueue list --topics** command and then retry the **mmaudit** command.
auditc_auditlogfile	Unable to open or append to the indicated **auditLogFile<...>** files for the given file system.	**Fileauditlogging** is in the failed state.	Check if the file system is mounted on the node and then retry the **mmaudit** command.
auditc_mmauditlog	Unable to open or append to the feature log file (`/var/adm/ras/mmaudit.log`).	**Fileauditlogging** is in the failed state.	Check if the local file system (`/var/adm/ras/`) on this node has enough space to write into and then return the **mmaudit** command.
auditc_loadkafkalib	Unable to enable file auditing for the indicated file system due to failure in loading the `librdkafka` library on this node.	**Fileauditlogging** is in the failed state.	Check the installation of the `librdkafka` libraries and then retry the **mmaudit** command.
auditc_createkafkahandle	Failed to create audit consumer **kafkaHandle** for the given file system.	**Fileauditlogging** is in the failed state.	Check the Kafka configuration and then retry the **mmaudit** command.
auditc_brokerconnect	Unable to connect to the indicated Kafka broker servers for the indicated file system.	**Fileauditlogging** is in the failed state.	Check the node status and the network connectivity to the **brokerServer** nodes as defined in **mmlsnodeclass** command for **kafkaBrokerServers**.
auditc_offsetstore	Failed to store the offset (in Kafka brokers to persist the marker up to which events are consumed) for the indicated file system.	**Fileauditlogging** is in the failed state.	Check on the **topicName** using **mmmsgqueue list --topics** command and then retry the **mmaudit** command.
auditc_flush_auditlogfile	Unable to flush the indicated **auditLogFile<...>** for the given file system.	**Fileauditlogging** is in the failed state.	Check if the file system is mounted on the node and still has space available in it.

Table 49. FILEAUDITLOG states (continued)

FILEAUDITLOG state	Description	Condition	Administrator's action
auditc_flush_errlogfile	Unable to flush the feature log file (/var/adm/ras/ mmaduit.log).	**Fileauditlogging** is in the failed state.	Check if the local file system (/var/adm/ras/) on this node has enough space to write into.
auditc_warn	A warning was encountered in the audit consumer for the indicated file system.	**Fileauditlogging** is in the degraded state.	Check /var/adm/ras/ mmaudit.log for more details.
auditc_err	An error was encountered in the audit consumer for the indicted file system.	**Fileauditlogging** is in the failed state.	Check /var/adm/ras/ mmaudit.log for more details.
auditc_found	An audit consumer listed in the **spectrumscale** configuration was detected.	**Fileauditlogging** is healthy.	None
auditc_vanished	An audit consumer, listed in the **spectrumscale** configuration, has been removed.	**Fileauditlogging** is in the transient state.	This can be a temporary/transient state. Issue command **mmaudit all list** to check the status.

Monitoring file audit logging using mmhealth commands

You can use **mmhealth** commands to monitor the status of file audit logging and the message queue.

The following **mmhealth** commands allow you to view the consumer status on a per node or per cluster basis and file audit logging overall.

Use to view the consumer status on a per node basis:

```
# mmhealth node show FILEAUDITLOG
```

You will see output similar to the following example:

```
Node name:      ibmnode1.ibm.com

Component     Status        Status Change     Reasons
------------------------------------------------------------------------------------
FILEAUDITLOG  HEALTHY       4 days ago        -
device0       HEALTHY       4 days ago        -
device1       HEALTHY       4 days ago        -

There are no active error events for the component FILEAUDITLOG on this node (ibmnode1.ibm.com).
```

Use to view more details about the consumers on a per node basis:

```
# mmhealth node show FILEAUDITLOG -v
```

You will see output similar to the following example:

```
Node name:      ibmnode1.ibm.com

Component       Status        Status Change             Reasons
-------------------------------------------------------------------------------------------
FILEAUDITLOG    HEALTHY       2018-04-09 15:28:27       -
device0         HEALTHY       2018-04-09 15:28:56       -
device1         HEALTHY       2018-04-09 15:31:27       -

Event              Parameter  Severity   Active Since          Event Message
-------------------------------------------------------------------------------------------------------------------------------------
```

```
auditc_ok          device0    INFO       2018-04-09 15:28:27   File Audit consumer for file system device0 is running
auditc_service_ok  device0    INFO       2018-04-09 15:28:27   File Audit consumer service for file system device0 is running
auditc_service_ok  device1    INFO       2018-04-09 15:31:26   File Audit consumer service for file system device1 is running
```

Use to view the status for the entire cluster:

```
# mmhealth cluster show FILEAUDITLOG
```

You will see output similar to the following example:

```
Component        Node                 Status            Reasons
-------------------------------------------------------------------------------
FILEAUDITLOG     ibmnode1.ibm.com     HEALTHY           -
FILEAUDITLOG     ibmnode2.ibm.com     HEALTHY           -
FILEAUDITLOG     ibmnode3.ibm.com     HEALTHY           -
FILEAUDITLOG     ibmnode4.ibm.com     HEALTHY           -
```

Use to view more details about each file system that has file audit logging enabled:

```
# mmhealth cluster show FILEAUDITLOG -v
```

You will see output similar to the following example:

```
Component         Node                 Status            Reasons
-------------------------------------------------------------------------
FILEAUDITLOG      ibmnode1.ibm.com     HEALTHY           -
device0                                HEALTHY           -
device1                                HEALTHY           -
FILEAUDITLOG      ibmnode2.ibm.com     HEALTHY           -
device0                                HEALTHY           -
device1                                HEALTHY           -
FILEAUDITLOG      ibmnode3.ibm.com     HEALTHY           -
device0                                HEALTHY           -
device1                                HEALTHY           -
FILEAUDITLOG      ibmnode4.ibm.com     HEALTHY           -
device0                                HEALTHY           -
device1                                HEALTHY           -
```

The following **mmhealth** commands allow you to view the message queue status on a per node or per cluster basis.

Use to view any active error events for the message queue on a single node:

```
# mmhealth node show MSGQUEUE
```

You will see output similar to the following example:

```
Node name:      ibmnode1.ibm.com

Component Status     Status     Change      Reasons
-------------------------------------------------------------------------------
MSGQUEUE             HEALTHY    3 days ago  -

There are no active error events for the component MSGQUEUE on this node (ibmnode1.ibm.com).
```

Use to view more details about a single node's broker and ZooKeeper status:

```
# mmhealth node show MSGQUEUE -v
```

You will see output similar to the following example:

```
Node name:      ibmnode1.ibm.com

Component      Status         Status Change         Reasons
-----------------------------------------------------------------------------------
MSGQUEUE       HEALTHY        2018-04-09 15:25:26   -

Event          Parameter     Severity    Active Since         Event Message
```

```
--------------------------------------------------------------------------------------------------------------
kafka_ok        MSGQUEUE      INFO       2018-04-09 15:25:26  kafka process as expected, state is started
zookeeper_ok    MSGQUEUE      INFO       2018-04-09 15:25:26  zookeeper process as expected, state is started
```

Use either of the following commands to view the state of the message queue for the entire cluster:

```
# mmhealth cluster show MSGQUEUE
# mmhealth cluster show MSGQUEUE -v
```

You will see output similar to the following example:

```
Component        Node                 Status            Reasons
-------------------------------------------------------------------------------------------------
MSGQUEUE         ibmnode1.ibm.com     HEALTHY           -
MSGQUEUE         ibmnode2.ibm.com     HEALTHY           -
MSGQUEUE         ibmnode3.ibm.com     HEALTHY           -
MSGQUEUE         ibmnode4.ibm.com     HEALTHY           -
MSGQUEUE         ibmnode5.ibm.com     HEALTHY           -
```

Monitoring file audit logging using the GUI

You can use the GUI to monitor file audit logging.

- To monitor the health of the file audit logging nodes, file systems, and see any events that might be present, use the **Services** > **File Auditing** page.
- To monitor the health of the message queue components like the brokers and ZooKeepers or see any events generated by them, use the **Services** > **Message Queue** page.

Chapter 11. Best practices for troubleshooting

Following certain best practices make the troubleshooting process easier.

How to get started with troubleshooting

Troubleshooting the issues reported in the system is easier when you follow the process step-by-step.

When you experience some issues with the system, go through the following steps to get started with the troubleshooting:

1. Check the events that are reported in various nodes of the cluster by using the **mmhealth node eventlog** command.
2. Check the user action corresponding to the active events and take the appropriate action. For more information on the events and corresponding user action, see "Events" on page 473.
3. If you are facing a deadlock issue, see Chapter 14, "Managing deadlocks," on page 271 to know how to resolve the issue.
4. Check for events which happened before the event you are trying to investigate. They might give you an idea about the root cause of problems. For example, if you see an event nfs_in_grace and node_resumed a minute before you get an idea about the root cause why NFS entered the grace period, it means that the node has been resumed after a suspend.
5. Collect the details of the issues through logs, dumps, and traces. You can use various CLI commands and **Settings** > **Diagnostic Data** GUI page to collect the details of the issues reported in the system. For more information, see Chapter 13, "Collecting details of the issues," on page 195.
6. Based on the type of issue, browse through the various topics that are listed in the troubleshooting section and try to resolve the issue.
7. If you cannot resolve the issue by yourself, contact IBM Support. For more information on how to contact IBM Support, see Chapter 31, "Support for troubleshooting," on page 469.

Back up your data

You need to back up data regularly to avoid data loss. It is also recommended to take backups before you start troubleshooting. The IBM Spectrum Scale provides various options to create data backups.

Follow the guidelines in the following sections to avoid any issues while creating backup:

- *GPFS(tm) backup data* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*
- *Backup considerations for using IBM Spectrum Protect* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*
- *Configuration reference for using IBM Spectrum Protect with IBM Spectrum Scale(tm)* in *IBM Spectrum Scale: Administration Guide*
- *Protecting data in a file system using backup* in *IBM Spectrum Scale: Administration Guide*
- *Backup procedure with SOBAR* in *IBM Spectrum Scale: Administration Guide*

The following best practices help you to troubleshoot the issues that might arise in the data backup process:

1. Enable the most useful messages in **mmbackup** command by setting the **MMBACKUP_PROGRESS_CONTENT** and **MMBACKUP_PROGRESS_INTERVAL** environment variables in the command environment prior to issuing the **mmbackup** command. Setting **MMBACKUP_PROGRESS_CONTENT=7** provides the most useful messages. For more information on these variables, see *mmbackup command* in *IBM Spectrum Scale: Command and Programming Reference*.

2. If the mmbackup process is failing regularly, enable debug options in the backup process:

   Use the **DEBUGmmbackup** environment variable or the **-d** option that is available in the **mmbackup** command to enable debugging features. This variable controls what debugging features are enabled. It is interpreted as a bitmask with the following bit meanings:

   0x001 Specifies that basic debug messages are printed to STDOUT. There are multiple components that comprise mmbackup, so the debug message prefixes can vary. Some examples include:

   ```
   mmbackup:mbackup.sh
   DEBUGtsbackup33:
   ```

   0x002 Specifies that temporary files are to be preserved for later analysis.

   0x004 Specifies that all dsmc command output is to be mirrored to STDOUT.

   The **-d** option in the **mmbackup** command line is equivalent to **DEBUGmmbackup = 1** .
3. To troubleshoot problems with backup subtask execution, enable debugging in the tsbuhelper program.

   Use the **DEBUGtsbuhelper** environment variable to enable debugging features in the mmbackup helper program tsbuhelper.

Resolve events in a timely manner

Resolving the issues in a timely manner helps to get attention on the new and most critical events. If there are a number of unfixed alerts, fixing any one event might become more difficult because of the effects of the other events. You can use either CLI or GUI to view the list of issues that are reported in the system.

You can use the **mmhealth node eventlog** to list the events that are reported in the system.

The **Monitoring** > **Events** GUI page lists all events reported in the system. You can also mark certain events as read to change the status of the event in the events view. The status icons become gray in case an error or warning is fixed or if it is marked as read. Some issues can be resolved by running a fix procedure. Use the action **Run Fix Procedure** to do so. The Events page provides a recommendation for which fix procedure to run next.

Keep your software up to date

Check for new code releases and update your code on a regular basis.

This can be done by checking the IBM support website to see if new code releases are available: IBM Spectrum Scale support website . The release notes provide information about new function in a release plus any issues that are resolved with the new release. Update your code regularly if the release notes indicate a potential issue.

Note: If a critical problem is detected on the field, IBM may send a flash, advising the user to contact IBM for an efix. The efix when applied might resolve the issue.

Subscribe to the support notification

Subscribe to support notifications so that you are aware of best practices and issues that might affect your system.

Subscribe to support notifications by visiting the IBM support page on the following IBM website: http://www.ibm.com/support/mynotifications.

By subscribing, you are informed of new and updated support site information, such as publications, hints and tips, technical notes, product flashes (alerts), and downloads.

Know your IBM warranty and maintenance agreement details

If you have a warranty or maintenance agreement with IBM, know the details that must be supplied when you call for support.

For more information on the IBM Warranty and maintenance details, see Warranties, licenses and maintenance.

Know how to report a problem

If you need help, service, technical assistance, or want more information about IBM products, you find a wide variety of sources available from IBM to assist you.

IBM maintains pages on the web where you can get information about IBM products and fee services, product implementation and usage assistance, break and fix service support, and the latest technical information. The following table provides the URLs of the IBM websites where you can find the support information.

Table 50. IBM websites for help, services, and information

Website	Address
IBM home page	http://www.ibm.com
Directory of worldwide contacts	http://www.ibm.com/planetwide
Support for IBM Spectrum Scale	IBM Spectrum Scale support website
Support for IBM System Storage® and IBM Total Storage products	http://www.ibm.com/support/entry/portal/product/system_storage/

Note: Available services, telephone numbers, and web links are subject to change without notice.

Before you call

Make sure that you have taken steps to try to solve the problem yourself before you call. Some suggestions for resolving the problem before calling IBM Support include:

- Check all hardware for issues beforehand.
- Use the troubleshooting information in your system documentation. The troubleshooting section of the IBM Knowledge Center contains procedures to help you diagnose problems.

To check for technical information, hints, tips, and new device drivers or to submit a request for information, go to the IBM Spectrum Scale support website .

Using the documentation

Information about your IBM storage system is available in the documentation that comes with the product. That documentation includes printed documents, online documents, readme files, and help files in addition to the IBM Knowledge Center. See the troubleshooting information for diagnostic instructions. To access this information, go to http://www.ibm.com/support/entry/portal/product/system_storage/storage_software/software_defined_storage/ibm_spectrum_scale and follow the instructions. The entire product documentation is available at: https://www.ibm.com/support/knowledgecenter/STXKQY/ibmspectrumscale_welcome.html?lang=en.

Other problem determination hints and tips

These hints and tips might be helpful when investigating problems related to logical volumes, quorum nodes, or system performance that can be encountered while using GPFS.

See these topics for more information:

- "Which physical disk is associated with a logical volume in AIX systems?"
- "Which nodes in my cluster are quorum nodes?"
- "What is stored in the /tmp/mmfs directory and why does it sometimes disappear?" on page 189
- "Why does my system load increase significantly during the night?" on page 189
- "What do I do if I receive message 6027-648?" on page 189
- "Why can't I see my newly mounted Windows file system?" on page 190
- "Why is the file system mounted on the wrong drive letter?" on page 190
- "Why does the offline mmfsck command fail with "Error creating internal storage"?" on page 190
- "Questions related to active file management" on page 190

Which physical disk is associated with a logical volume in AIX systems?

Earlier releases of GPFS allowed AIX logical volumes to be used in GPFS file systems. Their use is now discouraged because they are limited with regard to their clustering ability and cross platform support.

Existing file systems using AIX logical volumes are, however, still supported. This information might be of use when working with those configurations.

If an error report contains a reference to a logical volume pertaining to GPFS, you can use the **lslv -l** command to list the physical volume name. For example, if you want to find the physical disk associated with logical volume **gpfs7lv**, issue:

```
lslv -l gpfs44lv
```

Output is similar to this, with the physical volume name in column one.

```
  gpfs44lv:N/A
  PV               COPIES        IN BAND       DISTRIBUTION
  hdisk8           537:000:000   100%          108:107:107:107:108
```

Which nodes in my cluster are quorum nodes?

Use the **mmlscluster** command to determine which nodes in your cluster are quorum nodes.

Output is similar to this:

```
GPFS cluster information
========================
GPFS cluster name: cluster2.kgn.ibm.com
GPFS cluster id: 13882489265947478002
GPFS UID domain: cluster2.kgn.ibm.com
Remote shell command: /usr/bin/ssh
Remote file copy command: /usr/bin/scp
Repository type: CCR

Node Daemon node name IP address Admin node name Designation
-----------------------------------------------------------------------------
1 k164n04.kgn.ibm.com 198.117.68.68 k164n04.kgn.ibm.com quorum
2 k164n05.kgn.ibm.com 198.117.68.71 k164n05.kgn.ibm.com quorum
3 k164n06.kgn.ibm.com 198.117.68.70 k164n06.kgn.ibm.com
```

In this example, **k164n04** and **k164n05** are quorum nodes and **k164n06** is a non-quorum node.

To change the quorum status of a node, use the **mmchnode** command. To change one quorum node to nonquorum, GPFS does not have to be stopped. If you are changing more than one node at the same time, GPFS needs to be down on all the affected nodes. GPFS does not have to be stopped when changing nonquorum nodes to quorum nodes, nor does it need to be stopped on nodes that are not affected.

For example, to make **k164n05** a non-quorum node, and **k164n06** a quorum node, issue these commands:

```
mmchnode --nonquorum -N k164n05
mmchnode --quorum -N k164n06
```

To set a node's quorum designation at the time that it is added to the cluster, see **mmcrcluster** or **mmaddnode** *command* in *IBM Spectrum Scale: Command and Programming Reference*.

What is stored in the /tmp/mmfs directory and why does it sometimes disappear?

When GPFS encounters an internal problem, certain state information is saved in the GPFS dump directory for later analysis by the IBM Support Center.

The default dump directory for GPFS is **/tmp/mmfs**. This directory might disappear on Linux if cron is set to run the **/etc/cron.daily/tmpwatch** script. The **tmpwatch** script removes files and directories in **/tmp** that have not been accessed recently. Administrators who want to use a different directory for GPFS dumps can change the directory by issuing this command:

```
mmchconfig dataStructureDump=/name_of_some_other_big_file_system
```

Note: This state information (possibly large amounts of data in the form of GPFS dumps and traces) can be dumped automatically as part the first failure data capture mechanisms of GPFS, and can accumulate in the (default `/tmp/mmfs`) directory that is defined by the **dataStructureDump** configuration parameter. It is recommended that a cron job (such as `/etc/cron.daily/tmpwatch`) be used to remove `dataStructureDump` directory data that is older than two weeks, and that such data is collected (for example, via **gpfs.snap**) within two weeks of encountering any problem that requires investigation.

Why does my system load increase significantly during the night?

On some Linux distributions, cron runs the **/etc/cron.daily/slocate.cron** job every night. This will try to index all the files in GPFS. This will put a very large load on the GPFS token manager.

You can exclude all GPFS file systems by adding **gpfs** to the **excludeFileSytemType** list in this script, or exclude specific GPFS file systems in the **excludeFileSytemType** list.

```
/usr/bin/updatedb -f "excludeFileSystemType" -e "excludeFileSystem"
```

If indexing GPFS file systems is desired, only one node should run the **updatedb** command and build the database in a GPFS file system. If the database is built within a GPFS file system it will be visible on all nodes after one node finishes building it.

What do I do if I receive message 6027-648?

The **mmedquota** or **mmdefedquota** commands can fail with message **6027-648: EDITOR environment variable must be full path name**.

To resolve this error, do the following:

1. Change the value of the EDITOR environment variable to an absolute path name.
2. Check to see if the EDITOR variable is set in the login profiles such as *$HOME/.bashrc* and *$HOME/.kshrc*. If it is set, check to see if it is an absolute path name because the **mmedquota** or **mmdefedquota** command could retrieve the EDITOR environment variable from that file.

Why can't I see my newly mounted Windows file system?

On Windows, a newly mounted file system might not be visible to you if you are currently logged on to a system. This can happen if you have mapped a network share to the same drive letter as GPFS.

Once you start a new session (by logging out and logging back in), the use of the GPFS drive letter will supersede any of your settings for the same drive letter. This is standard behavior for all local file systems on Windows.

Why is the file system mounted on the wrong drive letter?

Before mounting a GPFS file system, you must be certain that the drive letter required for GPFS is freely available and is not being used by a local disk or a network-mounted file system on all nodes where the GPFS file system will be mounted.

Why does the offline mmfsck command fail with "Error creating internal storage"?

Use **mmfsck** command on the file system manager for storing internal data during a file system scan. The command fails if the GPFS fails to provide a temporary file of the required size.

The **mmfsck** command requires some temporary space on the file system manager for storing internal data during a file system scan. The internal data will be placed in the directory specified by the **mmfsck -t** command line parameter (**/tmp** by default). The amount of temporary space that is needed is proportional to the number of inodes (used and unused) in the file system that is being scanned. If GPFS is unable to create a temporary file of the required size, the **mmfsck** command will fail with the following error message:

```
Error creating internal storage
```

This failure could be caused by:

- The lack of sufficient disk space in the temporary directory on the file system manager
- The lack of sufficient page pool space on the file system manager as shown in **mmlsconfig pagepool** output
- Insufficiently high **filesize** limit set for the **root** user by the operating system
- The lack of support for large files in the file system that is being used for temporary storage. Some file systems limit the maximum file size because of architectural constraints. For example, JFS on AIX does not support files larger than 2 GB, unless the **Large file support** option has been specified when the file system was created. Check local operating system documentation for maximum file size limitations.

Why do I get timeout executing function error message?

If any of the commands fails due to timeout while executing mmccr, rerun the command to fix the issue. This timeout issue is likely related to an increased workload of the system.

Questions related to active file management

Issues and explanations pertaining to active file management.

The following questions are related to active file management (AFM).

How can I change the mode of a fileset?

The mode of an AFM client cache fileset cannot be changed from local-update mode to any other mode; however, it can be changed from read-only to single-writer (and vice versa), and from either read-only or single-writer to local-update.

To change the mode, do the following:

1. Ensure that fileset status is active and that the gateway is available.
2. Unmount the file system.
3. Unlink the fileset.
4. Run the **mmchfileset** command to change the mode.
5. Mount the file system again.
6. Link the fileset again.

Why are setuid/setgid bits in a single-writer cache reset at home after data is appended?

The setuid/setgid bits in a single-writer cache are reset at home after data is appended to files on which those bits were previously set and synced. This is because over NFS, a write operation to a setuid file resets the setuid bit.

How can I traverse a directory that has not been cached?

On a fileset whose metadata in all subdirectories is not cached, any application that optimizes by assuming that directories contain two fewer subdirectories than their hard link count will not traverse the last subdirectory. One such example is **find**; on Linux, a workaround for this is to use **find -noleaf** to correctly traverse a directory that has not been cached.

What extended attribute size is supported?

For an operating system in the gateway whose Linux kernel version is below 2.6.32, the NFS max rsize is 32K, so AFM would not support an extended attribute size of more than 32K on that gateway.

What should I do when my file system or fileset is getting full?

The `.ptrash` directory is present in cache and home. In some cases, where there is a conflict that AFM cannot resolve automatically, the file is moved to `.ptrash` at cache or home. In cache the `.ptrash` gets cleaned up when eviction is triggered. At home, it is not cleared automatically. When the administrator is looking to clear some space, the `.ptrash` should be cleaned up first.

Chapter 12. Understanding the system limitations

It is important to understand the system limitations to analyze whether you are facing a real issue in the IBM Spectrum Scale system.

The following topics list the IBM Spectrum Scale system limitations:

AFM limitations
See *AFM limitations* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

AFM-based DR limitations
See *AFM-based DR limitations* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Authentication limitations
See *Authentication limitations* in *IBM Spectrum Scale: Administration Guide*.

File authorization limitations
See *Authorization limitations* in *IBM Spectrum Scale: Administration Guide*.

File compression limitations
See *File compression* in *IBM Spectrum Scale: Administration Guide*.

FPO limitations
See *Restrictions* in *IBM Spectrum Scale: Administration Guide*.

General NFS V4 Linux Exceptions and Limitations
See *General NFS V4 Linux exceptions and limitations* in *IBM Spectrum Scale: Administration Guide*.

GPFS exceptions and limitations to NFSv4 ACLs
See *GPFS exceptions and limitations to NFS V4 ACLs* in *IBM Spectrum Scale: Administration Guide*.

GUI limitations
See *GUI limitations.* in *IBM Spectrum Scale: Administration Guide*.

HDFS transparency limitations
See *Configuration that differs from native HDFS in IBM Spectrum Scale* in *IBM Spectrum Scale: Big Data and Analytics Guide*.

HDFS transparency federation limitations
See *Known limitations* in *IBM Spectrum Scale: Big Data and Analytics Guide*.

Installation toolkit limitations
See *Limitations of the spectrumscale installation toolkit* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

mmuserauth service create command limitations
See *Limitations of the mmuserauth service create command while configuring AD with RFC2307* in *IBM Spectrum Scale: Administration Guide*.

Multiprotocol export limitations
See *Multiprotocol export considerations* in *IBM Spectrum Scale: Administration Guide*.

Performance monitoring limitations
See *Performance monitoring limitations* in *IBM Spectrum Scale: Administration Guide*.

Protocol cluster disaster recovery limitations
See *Protocols cluster disaster recovery limitations* in *IBM Spectrum Scale: Administration Guide*.

Protocol data security limitations
See *Data security limitations* in *IBM Spectrum Scale: Administration Guide*.

S3 API support limitations
See *Managing OpenStack access control lists using S3 API* in *IBM Spectrum Scale: Administration Guide*.

SMB limitations
See *SMB limitations* topic in *IBM Spectrum Scale: Administration Guide*.

Transparent cloud tiering limitations
See *Known limitations of Transparent cloud tiering* in *IBM Spectrum Scale: Administration Guide*.

Unified file and object access limitations
See *Limitations of unified file and object access* in *IBM Spectrum Scale: Administration Guide*.

Chapter 13. Collecting details of the issues

You need to collect the details of the issues that are reported in the system to start the troubleshooting process.

The IBM Spectrum Scale system provides the following options to collect the details of the issues reported in the system:

- Logs
- Dumps
- Traces
- Diagnostic data collection through CLI
- Diagnostic data collection through GUI

Collecting details of issues by using logs, dumps, and traces

The problem determination tools that are provided with IBM Spectrum Scale are intended to be used by experienced system administrators who know how to collect data and run debugging routines.

You can collect various types of logs such as GPFS logs, protocol service logs, operating system logs, and transparent cloud tiering logs. The GPFS™ log is a repository of error conditions that are detected on each node, as well as operational events such as file system mounts. The operating system error log is also useful because it contains information about hardware failures and operating system or other software failures that can affect the IBM Spectrum Scale system.

Note: The GPFS error logs and messages contain the MMFS prefix to distinguish it from the components of the IBM Multi-Media LAN Server, a related licensed program.

The IBM Spectrum Scale system also provides a system snapshot dump, trace, and other utilities that can be used to obtain detailed information about specific problems.

The information is organized as follows:

- "GPFS logs" on page 196
- "Operating system error logs" on page 214
- "Using the **gpfs.snap** command" on page 236
- "mmdumpperfdata command" on page 247
- "mmfsadm command" on page 249
- "Trace facility" on page 221

Time stamp in GPFS log entries

The time stamp in a GPFS log entry indicates the time of an event.

In IBM Spectrum Scale v4.2.2 and later, you can select either the earlier time stamp format for log entries or the ISO 8601 time stamp format. To select a format, use the **mmfsLogTimeStampISO8601** attribute of the **mmchconfig** command. The default setting is the ISO 8601 log time stamp format.

When you migrate to IBM Spectrum Scale v4.2.2, the time stamp format for the GPFS log is automatically set to the ISO 8601 format. You can prevent this action by including the **mmfsLogTimeStampISO8601** attribute when you complete the migration. For more information, see *Completing the migration to a new level of IBM Spectrum Scale* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Earlier time stamp format

In IBM Spectrum Scale v4.2.1 and earlier, the time stamp in the GPFS log has the following format:

Www Mmm DD hh:mm:ss.sss YYYY

where

Www
Is a three-character abbreviation for the day of the week.

Mmm
Is a three-character abbreviation for the month.

DD Is the day of the month.

hh:mm:sec
Is the hours (24-hour clock), minutes, seconds, and milliseconds.

YYYY
Is the year.

The following examples show the earlier time stamp format:

```
Mon May 09 15:12:20.603 2016
Sun Aug 15 07:04:33.078 2016
```

ISO 8601 time stamp format

In IBM Spectrum Scale v4.2.2 and later, by default, the time stamp in logs and traces follows a format similar to the ISO 8601 standard:

YYYY-MM-DD_hh:mm:ss.sss±hhmm

where

YYYY-MM-DD
Is the year, month, and day.

_ Is a separator character.

hh:mm:ss.sss
Is the hours (24-hour clock), minutes, seconds, and milliseconds.

±hhmm
Is the time zone designator, in hours and minutes offset from UTC.

The following examples show the ISO 8601 format:

```
2016-05-09_15:12:20.603-0500
2016-08-15_07:04:33.078+0200
```

Logs

This topic describes various logs that are generated in the IBM Spectrum Scale.

GPFS logs

The GPFS log is a repository of error conditions that are detected on each node, as well as operational events such as file system mounts. The GPFS log is the first place to look when you start debugging the abnormal events. As GPFS is a cluster file system, events that occur on one node might affect system behavior on other nodes, and all GPFS logs can have relevant data.

The GPFS log can be found in the **/var/adm/ras** directory on each node. The GPFS log file is named **mmfs.log.***date***.***nodeName*, where *date* is the time stamp when the instance of GPFS started on the node and *nodeName* is the name of the node. The latest GPFS log file can be found by using the symbolic file name **/var/adm/ras/mmfs.log.latest**.

The GPFS log from the prior startup of GPFS can be found by using the symbolic file name **/var/adm/ras/mmfs.log.previous**. All other files have a time stamp and node name appended to the file name.

At GPFS startup, log files that are not accessed during the last 10 days are deleted. If you want to save old log files, copy them elsewhere.

Many GPFS log messages can be sent to **syslog** on Linux. The **systemLogLevel** attribute of the **mmchconfig** command determines the GPFS log messages to be sent to the **syslog**. For more information, see the **mmchconfig** command in the *IBM Spectrum Scale: Command and Programming Reference*.

This example shows normal operational messages that appear in the GPFS log file on Linux node:

```
2017-08-29_15:53:04.196-0400: runmmfs starting
Removing old /var/adm/ras/mmfs.log.* files:
Unloading modules from /lib/modules/3.10.0-123.el7.x86_64/extra
Unloading module tracedev
Loading modules from /lib/modules/3.10.0-123.el7.x86_64/extra
runmmfs: Module tracedev is already loaded.
Module                  Size  Used by
mmfs26               2560839  0
mmfslinux             786518  1 mmfs26
tracedev               48579  3 mmfs26,mmfslinux
2017-08-29_15:53:05.825-0400: [I] CLI root root [EXIT, CHANGE] 'mmstartup -a' RC=0
2017-08-29_15:53:06.039-0400: [I] This node has a valid advanced license
2017-08-29_15:53:06.038-0400: [I] Initializing the fast condition variables at 0x7F1199C29540 ...
2017-08-29_15:53:06.039-0400: [I] mmfsd initializing. {Version: 5.0.0.0 Built: Aug 29 2017 10:50:41} ...
2017-08-29_15:53:06.089-0400: [I] Tracing in overwrite mode
2017-08-29_15:53:06.089-0400: [I] Cleaning old shared memory ...
2017-08-29_15:53:06.089-0400: [I] First pass parsing mmfs.cfg ...
2017-08-29_15:53:06.089-0400: [I] Enabled automated deadlock detection.
2017-08-29_15:53:06.089-0400: [I] Enabled automated deadlock debug data collection.
2017-08-29_15:53:06.089-0400: [I] Enabled automated expel debug data collection.
2017-08-29_15:53:06.089-0400: [I] Please see https://ibm.biz/Bd4bNK for more information on deadlock amelioration.
2017-08-29_15:53:06.089-0400: [I] Initializing the main process ...
2017-08-29_15:53:06.098-0400: [I] Second pass parsing mmfs.cfg ...
2017-08-29_15:53:06.098-0400: [I] Initializing NUMA support ...
2017-08-29_15:53:06.098-0400: [I] NUMA discover nNumaNodes 2 nCpus 16 maxNumaNode 1
2017-08-29_15:53:06.099-0400: [I] NUMA discover NUMA node 0 with CPUs 0-4,8-11
2017-08-29_15:53:06.099-0400: [I] NUMA discover NUMA node 1 with CPUs 4-8,12-15
2017-08-29_15:53:06.099-0400: [I] Initializing the page pool ...
2017-08-29_15:53:06.193-0400: [I] Initializing the mailbox message system ...
2017-08-29_15:53:06.195-0400: [I] Initializing encryption ...
2017-08-29_15:53:06.242-0400: [I] Encryption: loaded crypto library: GSKit Non-FIPS context (ver: 8.5.30.0).
2017-08-29_15:53:06.242-0400: [I] Initializing the thread system ...
2017-08-29_15:53:06.242-0400: [I] Creating threads ...
2017-08-29_15:53:06.251-0400: [I] Initializing inter-node communication ...
2017-08-29_15:53:06.251-0400: [I] Creating the main SDR server object ...
2017-08-29_15:53:06.251-0400: [I] Initializing the sdrServ library ...
2017-08-29_15:53:06.256-0400: [I] Initializing the ccrServ library ...allowRemoteConnections=1
2017-08-29_15:53:06.267-0400: [I] Initializing the cluster manager ...
2017-08-29_15:53:07.078-0400: [I] Initializing the token manager ...
2017-08-29_15:53:07.080-0400: [I] Initializing network shared disks ...
2017-08-29_15:53:08.654-0400: [I] Starting the ccrServ ...
2017-08-29_15:53:08.671-0400: [D] PFD load: mostRecent: 1 seq: 58904 (4096)
2017-08-29_15:53:08.671-0400: [D] PFD load: nextToWriteIdx: 0 seq: 58903 (4096)
2017-08-29_15:53:08.721-0400: [I] Initialize compression libraries handlers...
2017-08-29_15:53:08.722-0400: [I] Found 2 compression libraries
2017-08-29_15:53:09.223-0400: [N] Connecting to 192.168.115.89 node1 <c0p2>
2017-08-29_15:53:09.224-0400: [N] Connecting to 192.168.200.63 node2 <c0p0>
2017-08-29_15:53:09.226-0400: [N] Connecting to 192.168.200.64 node3 <c0p1>
```

```
2017-08-29_15:53:14.749-0400: [I] Connected to 192.168.200.63 node1 <c0p0>
2017-08-29_15:53:16.295-0400: [I] Accepted and connected to 192.168.200.64 node2 <c0p1>
2017-08-29_15:53:25.016-0400: [I] Node 192.168.200.63 (node1) is now the Group Leader.
2017-08-29_15:53:25.092-0400: [I] Calling user exit script mmClusterManagerRoleChange: event clusterManagerTakeOver,
Async command /usr/lpp/mmfs/bin/mmsysmonc.
2017-08-29_15:53:25.103-0400: [I] Calling user exit script clusterManagerRole: event clusterManagerTakeOver,
Async command /usr/lpp/mmfs/bin/mmsysmonc.
2017-08-29_15:53:25.318-0400: [N] mmfsd ready
2017-08-29_15:53:25.389-0400: mmcommon mmfsup invoked. Parameters: 192.168.116.133 192.168.200.63 all
2017-08-29_15:53:25.910-0400: [I] Calling user exit script mmSysMonGpfsStartup: event startup,
Async command /usr/lpp/mmfs/bin/mmsysmoncontrol.
2017-08-29_15:53:25.922-0400: [I] Calling user exit script mmSinceShutdownRoleChange: event startup,
Async command /usr/lpp/mmfs/bin/mmsysmonc.
```

The **mmcommon logRotate** command can be used to rotate the GPFS log without shutting down and restarting the daemon. After the **mmcommon logRotate** command is issued, **/var/adm/ras/mmfs.log.previous** will contain the messages that occurred since the previous startup of GPFS or the last run of **mmcommon logRotate**. The **/var/adm/ras/mmfs.log.latest** file starts over at the point in time that **mmcommon logRotate** was run.

Depending on the size and complexity of your system configuration, the amount of time to start GPFS varies. If you cannot access a file system that is mounted, examine the log file for error messages.

Creating a master GPFS log file:

The GPFS log frequently shows problems on one node that actually originated on another node.

GPFS is a file system that runs on multiple nodes of a cluster. This means that problems originating on one node of a cluster often have effects that are visible on other nodes. It is often valuable to merge the GPFS logs in pursuit of a problem. Having accurate time stamps aids the analysis of the sequence of events.

Before following any of the debug steps, IBM suggests that you:

1. Synchronize all clocks of all nodes in the GPFS cluster. If this is not done, and clocks on different nodes are out of sync, there is no way to establish the real time line of events occurring on multiple nodes. Therefore, a merged error log is less useful for determining the origin of a problem and tracking its effects.
2. Merge and chronologically sort all of the GPFS log entries from each node in the cluster. The **--gather-logs** option of the **gpfs.snap** command can be used to achieve this:

   ```
   gpfs.snap --gather-logs -d /tmp/logs -N all
   ```

   The system displays information similar to:

   ```
   gpfs.snap: Gathering mmfs logs ...
   gpfs.snap: The sorted and unsorted mmfs.log files are in /tmp/logs
   ```

   If the **--gather-logs** option is not available on your system, you can create your own script to achieve the same task; use **/usr/lpp/mmfs/samples/gatherlogs.samples.sh** as an example.

Audit messages for cluster configuration changes

As an aid to troubleshooting and to improve cluster security, IBM Spectrum Scale can send an audit message to syslog and the GPFS log whenever a GPFS command changes the configuration of the cluster.

You can use the features of syslog to mine, process, or redirect the audit messages.

Restriction: Audit messages are not available on Windows operating systems.

Configuring syslog

On Linux operating systems, syslog typically is enabled by default. On AIX, syslog must be set up and configured. See the corresponding operating system documentation for details.

Configuring audit messages

By default, audit messages are enabled and messages are sent to syslog but not to the GPFS log. You can control audit messages with the **commandAudit** attribute of the **mmchconfig** command. For more information, see the topic *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference* guide.

Audit messages are not affected by the **systemLogLevel** attribute of the **mmchconfig** command.

If audit logs are enabled, the GUI receives the updates on configuration changes that you made through CLI and updates its configuration cache to reflect the changes in the GUI. You can also disable audit logging with the **mmchconfig** command. If the audit logs are disabled, the GUI does not show the configuration changes immediately. It might be as much as an hour late in reflecting configuration changes that are made through the CLI.

Message format

For security, sensitive information such as a password is replaced with asterisks (*) in the audit message.

Audit messages are sent to syslog with an identity of **mmfs**, a facility code of **user**, and a severity level of **informational**. For more information about the meaning of these terms, see the syslog documentation.

The format of the message depends on the source of the GPFS command:

- Messages about GPFS commands that are entered at the command line have the following format:

  CLI *user_name user_name* [*AUDIT_TYPE1*,*AUDIT_TYPE2*] '*command*' **RC**=*return_code*

  where:

  CLI The source of the command. Indicates that the command was entered from the command line.

  user_name user_name
  The name of the user who entered the command, such as `root`. The same name appears twice.

  AUDIT_TYPE1
  The point in the command when the message was sent to syslog. Always **EXIT**.

  AUDIT_TYPE2
  The action taken by the command. Always **CHANGE**.

  command
  The text of the command.

  return_code
  The return code of the GPFS command.

- Messages about GPFS commands that are issued by GUI commands have a similar format:

  GUI-CLI *user_name GUI_user_name* [*AUDIT_TYPE1*,*AUDIT_TYPE2*] '*command*' **RC**=*return_code*

  where:

  GUI-CLI
  The source of the command. Indicates that the command was called by a GUI command.

  user_name
  The name of the user, such as `root`.

GUI_user_name
The name of the user who logged on to the GUI.

The remaining fields are the same as in the **CLI** message.

The following lines are examples from a syslog:

```
Apr 24 13:56:26 c12c3apv12 mmfs[63655]: CLI root root [EXIT, CHANGE] 'mmchconfig
autoload=yes' RC=0
Apr 24 13:58:42 c12c3apv12 mmfs[65315]: CLI root root [EXIT, CHANGE] 'mmchconfig
deadlockBreakupDelay=300' RC=0
Apr 24 14:04:47 c12c3apv12 mmfs[67384]: CLI root root [EXIT, CHANGE] 'mmchconfig
FIPS1402mode=no' RC=0
```

The following lines are examples from a syslog where GUI is the originator:

```
Apr 24 13:56:26 c12c3apv12 mmfs[63655]: GUI-CLI root admin [EXIT, CHANGE] 'mmchconfig autoload=yes' RC=0
```

Commands

IBM Spectrum Scale sends audit messages to syslog for the following commands and options:

mmaddcallback
mmadddisk
mmaddnode
mmafmconfig add
mmafmconfig delete
mmafmconfig disable
mmafmconfig enable
mmafmconfig update
mmafmctl
mmapplypolicy
mmaudit
mmauth add
mmauth delete
mmauth deny
mmauth gencert
mmauth genkey
mmauth grant
mmauth update
mmbackup
mmbackupconfig
mmces address add
mmces address change
mmces address move
mmces address remove
mmces log
mmces node resume
mmces node suspend
mmces service disable
mmces service enable
mmces service start

mmces service stop
mmcesdr
mmcesmonitor
mmchcluster
mmchconfig
mmchdisk
mmchfileset
mmchfs
mmchlicense
mmchmgr
mmchnode
mmchnodeclass
mmchnsd
mmchpolicy
mmchpool
mmchqos
mmcloudgateway account create
mmcloudgateway account delete
mmcloudgateway account update
mmcloudgateway config set
mmcloudgateway config unset
mmcloudgateway files delete
mmcloudgateway files migrate
mmcloudgateway files recall
mmcloudgateway files reconcile
mmcloudgateway files restore
mmcloudgateway filesystem create
mmcloudgateway filesystem delete
mmcloudgateway service start
mmcloudgateway service stop
mmcrcluster
mmcrfileset
mmcrfs
mmcrnodeclass
mmcrnsd
mmcrsnapshot
mmdefedquota
mmdefquotaoff
mmdefquotaon
mmdefragfs
mmdelcallback
mmdeldisk
mmdelfileset
mmdelfs
mmdelnode

mmdelnodeclass
mmdelnsd
mmdelsnapshot
mmedquota
mmexpelnode
mmexportfs
mmfsctl
mmimgbackup
mmimgrestore
mmimportfs
mmkeyserv
mmlinkfileset
mmmigratefs
mmmsgqueue
mmnfs config change
mmnfs export add
mmnfs export change
mmnfs export load
mmnfs export remove
mmnsddiscover
mmobj config change
mmobj file access
mmobj multiregion enable
mmobj multiregion export
mmobj multiregion import
mmobj multiregion remove
mmobj policy change
mmobj policy create
mmobj policy deprecate
mmobj swift base
mmperfmon config add
mmperfmon config delete
mmperfmon config generate
mmperfmon config update
mmpsnap create
mmpsnap delete
mmquotaoff
mmquotaon
mmremotecluster add
mmremotecluster delete
mmremotecluster update
mmremotefs add
mmremotefs delete
mmremotefs update
mmrestoreconfig

mmrestorefs
mmrestripefile
mmrestripefs
mmrpldisk
mmsdrrestore
mmsetquota
mmshutdown
mmsmb config change
mmsmb export add
mmsmb export change
mmsmb export remove
mmsmb exportacl add
mmsmb exportacl change
mmsmb exportacl delete
mmsmb exportacl remove
mmsmb exportacl replace
mmsnapdir
mmstartup
mmumount
mmumount
mmunlinkfileset
mmuserauth service create
mmuserauth service remove
mmwinservctl

Protocol services logs

The protocol service logs contains the information that helps you to troubleshoot the issues related to the NFS, SMB, and Object services.

By default, the NFS and SMB protocol logs are stored at: `/var/log/messages`. The Object protocol logs are stored in the directories for the service: `/var/log/swift`, `/var/log/keystone/`, and `/var/log/httpd`.

For more information on logs of the installation toolkit, see *Logging and debugging for installation toolkit* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

SMB logs:

The SMB services write the most important messages to syslog.

The SMB service in IBM Spectrum Scale writes its log message into the syslog of the CES nodes. Thus, it needs a working syslog daemon and configuration. An SMB snap expects syslog on CES nodes to be found in a file in the distribution's default paths. If syslog gets redirected to another location, the customer should provide the logs in case of support.

With the standard syslog configuration, you can search for the terms such as `ctdbd` or `smbd` in the `/var/log/messages` file to see the relevant logs. For example:

```
grep ctdbd /var/log/messages
```

The system displays output similar to the following example:

```
May 31 09:11:23 prt002st001 ctdbd: Updated hot key database=locking.tdb key=0x2795c3b1 id=0 hop_count=1
May 31 09:27:33 prt002st001 ctdbd: Updated hot key database=smbXsrv_open_global.tdb key=0x0d0d4abe id=0 hop_count=1
May 31 09:37:17 prt002st001 ctdbd: Updated hot key database=brlock.tdb key=0xc37fe57c id=0 hop_count=1
```

grep smbd /var/log/messages

The system displays output similar to the following example:

```
May 31 09:40:58 prt002st001 smbd[19614]: [2015/05/31 09:40:58.357418,  0] ../source3
/lib/dbwrap/dbwrap_ctdb.c:962(db_ctdb_record_destr)
May 31 09:40:58 prt002st001 smbd[19614]: tdb_chainunlock on db /var/lib/ctdb/locking.tdb.2,
key FF5B87B2A3FF862E96EFB40000000000000000000000000 took 5.261000 milliseconds
May 31 09:55:26 prt002st001 smbd[1431]: [2015/05/31 09:55:26.703422,  0] ../source3
/lib/dbwrap/dbwrap_ctdb.c:962(db_ctdb_record_destr)
May 31 09:55:26 prt002st001 smbd[1431]: tdb_chainunlock on db /var/lib/ctdb/locking.tdb.2,
key FF5B87B2A3FF862EE5073801000000000000000000000000 took 17.844000 milliseconds
```

Additional SMB service logs are available in following folders:

- `/var/adm/ras/log.smbd`
- `/var/adm/ras/log.smbd.old`

When the size of the `log.smbd` file becomes 100 MB, the system changes the file as `log.smbd.old`. To capture more detailed traces for problem determination, use the **mmprotocoltrace** command.

Some of the issues with SMB services are related to winbind service also. For more information about winbind tracing, see "Winbind logs" on page 209.

Related concepts:

"Determining the health of integrated SMB server" on page 383
There are some IBM Spectrum Scale commands to determine the health of the SMB server.

NFS logs:

The clustered export services (CES) NFS server writes log messages in the `/var/log/ganesha.log` file at runtime.

Operating system's log rotation facility is used to manage NFS logs. The NFS logs are configured and enabled during the NFS server packages installation.

The following example shows a sample log file:

```
# tail -f /var/log/ganesha.log
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs_Init_admin_thread :NFS CB
:EVENT :Admin thread initialized
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs4_start_grace
:STATE :EVENT :NFS Server Now IN GRACE,
duration 59
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_rpc_cb_init_ccache :NFS STARTUP :EVENT
 :Callback creds directory (/var/run/ganesha) already exists
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs_rpc_cb_init_ccache
:NFS STARTUP :WARN :gssd_refresh_krb5_machine_credential failed (-1765328378:0)
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs_Start_threads :THREAD :EVENT :Starting delayed executor.
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs_Start_threads :THREAD :EVENT :gsh_dbusthread was started successfully
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_Start_threads :THREAD :EVENT :admin thread was started successfully
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_Start_threads :THREAD :EVENT :reaper thread was started successfully
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_Start_threads :THREAD :EVENT :General fridge was started successfully
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[reaper]
 nfs_in_grace :STATE :EVENT :NFS Server Now IN GRACE
```

```
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_start :NFS STARTUP :EVENT :-------------------------------------------------
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
 nfs_start :NFS STARTUP :EVENT : NFS SERVER INITIALIZED
2018-04-09 11:28:18 : epoch 000100a2 : rh424a : gpfs.ganesha.nfsd-20924[main]
nfs_start :NFS STARTUP :EVENT :-------------------------------------------------
```

Log levels can be displayed by using the **mmnfs config list | grep LOG_LEVEL** command. For example:

```
mmnfs config list | grep LOG_LEVEL
```

The system displays output similar to the following example:

```
LOG_LEVEL: EVENT
```

By default, the log level is EVENT. Additionally, the following NFS log levels can also be used; starting from lowest to highest verbosity:

- FATAL
- MAJ
- CRIT
- WARN
- INFO
- DEBUG
- MID_DEBUG
- FULL_DEBUG

  Note: The FULL_DEBUG level increases the size of the log file. Use it in the production mode only if instructed by the IBM Support.

Increasing the verbosity of the NFS server log impacts the overall NFS I/O performance.

To change the logging to the verbose log level INFO, use the following command:

mmnfs config change LOG_LEVEL=INFO

The system displays output similar to the following example:

```
NFS Configuration successfully changed. NFS server restarted on all NFS nodes on which NFS
server is running.
```

This change is cluster-wide and restarts all NFS instances to activate this setting. The log file now displays more informational messages, for example:

```
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_rpc_dispatch_threads
:THREAD :INFO :5 rpc dispatcher threads were started successfully
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[disp] rpc_dispatcher_thread
:DISP :INFO :Entering nfs/rpc dispatcher
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[disp] rpc_dispatcher_thread
:DISP :INFO :Entering nfs/rpc dispatcher
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[disp] rpc_dispatcher_thread
:DISP :INFO :Entering nfs/rpc dispatcher
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[disp] rpc_dispatcher_thread
:DISP :INFO :Entering nfs/rpc dispatcher
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_Start_threads
:THREAD :EVENT :gsh_dbusthread was started successfully
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_Start_threads
:THREAD :EVENT :admin thread was started successfully
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_Start_threads
:THREAD :EVENT :reaper thread was started successfully
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_Start_threads
:THREAD :EVENT :General fridge was started successfully
2015-06-03 12:49:31 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[reaper] nfs_in_grace
:STATE :EVENT :NFS Server Now IN GRACE
```

```
2015-06-03 12:49:32 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_start
:NFS STARTUP :EVENT :---------------------------------------------
2015-06-03 12:49:32 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_start
:NFS STARTUP :EVENT :             NFS SERVER INITIALIZED
2015-06-03 12:49:32 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[main] nfs_start
:NFS STARTUP :EVENT :---------------------------------------------
2015-06-03 12:50:32 : epoch 556edba9 : cluster1.ibm.com : ganesha.nfsd-21582[reaper] nfs_in_grace
:STATE :EVENT :NFS Server Now NOT IN GRACE
```

To display the currently configured CES log level, use the following command:

mmces log level

The system displays output similar to the following example:

```
CES log level is currently set to 0
```

The log file is `/var/adm/ras/mmfs.log.latest`. By default, the log level is `0` and other possible values are `1`, `2`, and `3`. To increase the log level, use the following command:

mmces log level 1

NFS-related log information is written to the standard GPFS log files as part of the overall CES infrastructure. This information relates to the NFS service management and recovery orchestration within CES.

Object logs:

There are a number of locations where messages are logged with the object protocol.

The core object services, proxy, account, container, and object server have their own logging level sets in their respective configuration files. By default, unified file and object access logging is set to show messages at or above the `ERROR` level, but can be changed to `INFO` or `DEBUG` levels if more detailed logging information is required.

By default, the messages logged by these services are saved in the `/var/log/swift` directory.

You can also configure these services to use separate syslog facilities by the **log_facility** parameter in one or all of the object service configuration files and by updating the rsyslog configuration. These parameters are described in the Swift Deployment Guide(docs.openstack.org/developer/swift/deployment_guide.html) that is available in the OpenStack documentation.

An example of how to set up this configuration can be found in the SAIO - Swift All In One documentation(docs.openstack.org/developer/swift/development_saio.html#optional-setting-up-rsyslog-for-individual-logging) that is available in the OpenStack documentation.

Note: To configure rsyslog for unique log facilities in the protocol nodes, the administrator needs to ensure that the manual steps mentioned in the preceding link are carried out on each of those protocol nodes.

The Keystone authentication service writes its logging messages to `/var/log/keystone/keystone.log` file. By default, Keystone logging is set to show messages at or above the `WARNING` level.

For information on how to view or change log levels on any of the object related services, see *CES tracing and debug data collection* in *IBM Spectrum Scale: Problem Determination Guide*.

The following commands can be used to determine the health of object services:

- To see whether there are any nodes in an active (failed) state, run the following command:

mmces state cluster OBJ

The system displays output similar to this:

```
NODE         COMPONENT   STATE     EVENTS
prt001st001  OBJECT      HEALTHY
prt002st001  OBJECT      HEALTHY
prt003st001  OBJECT      HEALTHY
prt004st001  OBJECT      HEALTHY
prt005st001  OBJECT      HEALTHY
prt006st001  OBJECT      HEALTHY
prt007st001  OBJECT      HEALTHY
```

In this example, all nodes are healthy so no active events are shown.

- To display the history of events generated by the monitoring framework, run the following command:

  mmces events list OBJ

  The system displays output similar to this:

```
Node  Timestamp                          Event Name                  Severity   Details
node1 2015-06-03 13:30:27.478725+08:08PDT  proxy-server_ok             INFO       proxy process as expected
node1 2015-06-03 14:26:30.567245+08:08PDT  object-server_ok            INFO       object process as expected
node1 2015-06-03 14:26:30.720534+08:08PDT  proxy-server_ok             INFO       proxy process as expected
node1 2015-06-03 14:28:30.689257+08:08PDT  account-server_ok           INFO       account process as expected
node1 2015-06-03 14:28:30.853518+08:08PDT  container-server_ok         INFO       container process as expected
node1 2015-06-03 14:28:31.015307+08:08PDT  object-server_ok            INFO       object process as expected
node1 2015-06-03 14:28:31.177589+08:08PDT  proxy-server_ok             INFO       proxy process as expected
node1 2015-06-03 14:28:49.025021+08:08PDT  postIpChange_info           INFO       IP addresses modified 192.167.12.21_0-_1.
node1 2015-06-03 14:28:49.194499+08:08PDT  enable_Address_database_node INFO       Enable Address Database Node
node1 2015-06-03 14:29:16.483623+08:08PDT  postIpChange_info           INFO       IP addresses modified 192.167.12.22_0-_2.
node1 2015-06-03 14:29:25.274924+08:08PDT  postIpChange_info           INFO       IP addresses modified 192.167.12.23_0-_3.
node1 2015-06-03 14:29:30.844626+08:08PDT  postIpChange_info           INFO       IP addresses modified 192.167.12.24_0-_4.
```

- To retrieve the OBJ related event entries, query the monitor client and grep for the name of the component you want to filter on, either object, proxy, account, container, keystone or postgres. For example, to see proxy-server related events, run the following command:

  mmces events list | grep proxy

  The system displays output similar to this:

```
node1 2015-06-01 14:39:49.120912+08:08PDT  proxy-server_failed      ERROR      proxy process should be started but is stopped
node1 2015-06-01 14:44:49.277940+08:08PDT  proxy-server_ok          INFO       proxy process as expected
node1 2015-06-01 16:27:37.923696+08:08PDT  proxy-server_failed      ERROR      proxy process should be started but is stopped
node1 2015-06-01 16:40:39.789920+08:08PDT  proxy-server_ok          INFO       proxy process as expected
node1 2015-06-03 13:28:18.875566+08:08PDT  proxy-server_failed      ERROR      proxy process should be started but is stopped
node1 2015-06-03 13:30:27.478725+08:08PDT  proxy-server_ok          INFO       proxy process as expected
node1 2015-06-03 13:30:57.482977+08:08PDT  proxy-server_failed      ERROR      proxy process should be started but is stopped
node1 2015-06-03 14:26:30.720534+08:08PDT  proxy-server_ok          INFO       proxy process as expected
node1 2015-06-03 14:27:00.759696+08:08PDT  proxy-server_failed      ERROR      proxy process should be started but is stopped
node1 2015-06-03 14:28:31.177589+08:08PDT  proxy-server_ok          INFO       proxy process as expected
```

- To check the monitor log, grep for the component you want to filter on, either object, proxy, account, container, keystone or postgres. For example, to see object-server related log messages:

  grep object /var/adm/ras/mmsysmonitor.log | head -n 10

  The system displays output similar to this:

```
2015-06-03T13:59:28.805-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJ running command
'systemctl status openstack-swift-proxy'
2015-06-03T13:59:28.916-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJ command result
ret:3 sout:openstack-swift-proxy.service - OpenStack Object Storage (swift) - Proxy Server
2015-06-03T13:59:28.916-08:00 util5.sonasad.almaden.ibm.com I:522632:Thread-9:object:OBJ openstack-swift-proxy is not started, ret3
2015-06-03T13:59:28.916-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJProcessMonitor openstack-swift-proxy failed:
2015-06-03T13:59:28.916-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJProcessMonitor memcached started
2015-06-03T13:59:28.917-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJ running command
'systemctl status memcached'
2015-06-03T13:59:29.018-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJ command result
ret:0 sout:memcached.service - Memcached
2015-06-03T13:59:29.018-08:00 util5.sonasad.almaden.ibm.com I:522632:Thread-9:object:OBJ memcached is started and active running
2015-06-03T13:59:29.018-08:00 util5.sonasad.almaden.ibm.com D:522632:Thread-9:object:OBJProcessMonitor memcached succeeded
2015-06-03T13:59:29.018-08:00 util5.sonasad.almaden.ibm.com I:522632:Thread-9:object:OBJ service started checks
after monitor loop, event count:6
```

The following tables list the IBM Spectrum Scale for object storage log files.

Table 51. Core object log files in /var/log/swift

Log file	Component	Configuration file
account-auditor.log account-auditor.error	Account auditor Swift service	account-server.conf
account-reaper.log account-reaper.error	Account reaper Swift service	account-server.conf
account-replicator.log account-replicator.error	Account replicator Swift service	account-server.conf
account-server.log account-server.error	Account server Swift service	account-server.conf
container-auditor.log container-auditor.error	Container auditor Swift service	container-server.conf
container-replicator.log container-replicator.error	Container replicator Swift service	container-server.conf
container-server.log container-server.error	Container server Swift service	container-server.conf
container-updater.log container-updater.error	Container updater Swift service	container-server.conf
object-auditor.log object-auditor.error	Object auditor Swift service	object-server.conf
object-expirer.log object-expirer.error	Object expirer Swift service	object-expirer.conf
object-replicator.log object-replicator.error	Object replicator Swift service	object-server.conf
object-server.log object-server.error	Object server Swift service	object-server.conf object-server-sof.conf
object-updater.log object-updater.error	Object updater Swift service	object-server.conf
proxy-server.log proxy-server.error	Proxy server Swift service	proxy-server.conf

Table 52. Additional object log files in /var/log/swift

Log file	Component	Configuration file
ibmobjectizer.log ibmobjectizer.error	Unified file and object access objectizer service	spectrum-scale-objectizer.conf spectrum-scale-object.conf
policyscheduler.log policyscheduler.error	Object storage policies	spectrum-scale-object-policies.conf

Table 52. Additional object log files in /var/log/swift (continued)

Log file	Component	Configuration file
swift.log swift.error	Performance metric collector (pmswift)	

Table 53. General system log files in /var/adm/ras

Log file	Component
mmsysmonitor.log	Includes everything that is monitored in the monitoring framework
mmfs.log	Various IBM Spectrum Scale command logging

Winbind logs:

The winbind services write the most important messages to syslog.

When using Active Directory, the most important messages are written to syslog, similar to the logs in SMB protocol. For example:

```
grep winbindd /var/log/messages
```

The system displays output similar to the following example:

```
Jun  3 12:04:34 prt001st001 winbindd[14656]: [2015/06/03 12:04:34.271459,  0] ../lib/util/become_daemon.c:124(daemon_ready)
Jun  3 12:04:34 prt001st001 winbindd[14656]: STATUS=daemon 'winbindd' finished starting up and ready to serve connections
```

Additional logs are available in /var/adm/ras/log.winbindd* and /var/adm/ras/log.wb*. There are multiple files that get rotated with the "old" suffix, when the size becomes 100 MB.

To capture debug traces for Active Directory authentication, use **mmprotocoltrace** command for the **winbind** component. To start the tracing of **winbind** component, issue this command:

```
mmprotocoltrace start winbind
```

After performing all steps, relevant for the trace, issue this command to stop tracing **winbind** component and collect tracing data from all participating nodes:

```
mmprotocoltrace stop winbind
```

Related concepts:

"Determining the health of integrated SMB server" on page 383
There are some IBM Spectrum Scale commands to determine the health of the SMB server.

The IBM Spectrum Scale HDFS transparency log:

In IBM Spectrum Scale HDFS transparency, all logs are recorded using log4j. The **log4j.properties** file is under the /usr/lpp/mmfs/hadoop/etc/hadoop directory.

By default, the logs are written under the /usr/lpp/mmfs/hadoop/logs directory.

The following entries can be added into the **log4j.properties** file to turn on the debugging information:

```
log4j.logger.org.apache.hadoop.yarn=DEBUG
log4j.logger.org.apache.hadoop.hdfs=DEBUG
log4j.logger.org.apache.hadoop.gpfs=DEBUG
log4j.logger.org.apache.hadoop.security=DEBUG
```

Protocol authentication log files:

The log files pertaining to protocol authentication are described here.

Table 54. Authentication log files

Service name	Log configuration file	Log files	Logging levels

Keystone	`/etc/keystone/keystone.conf`
	
	`/etc/keystone/logging.conf`
		`/var/log/keystone/keystone.log`
		
		`/var/log/keystone/httpd-error.log`
		
		`/var/log/keystone/httpd-access.log`
			In `keystone.conf` change
			1. `debug = true`- for getting debugging information in log file.
			2. `verbose = true` - for getting Info messages in log file .
			
			By default, these values are false and only warning messages are logged.
			
			Finer grained control of keystone logging levels can be specified by updating the keystones `logging.conf` file. For information on the logging levels in the `logging.conf` file, see OpenStack logging.conf documentation (docs.openstack.org/kilo/config-reference/content/section_keystone-logging.conf.html).

Table 54. Authentication log files (continued)

Service name	Log configuration file	Log files	Logging levels

SSSD	`/etc/sssd/sssd.conf`	`/var/log/sssd/sssd.log`
		
		`/var/log/sssd/sssd_nss.log`
		
		`/var/log/sssd/sssd_LDAPDOMAIN.log` (depends upon configuration)
		
		`/var/log/sssd/sssd_NISDOMAIN.log` (depends upon configuration)
		Note: For more information on SSSD log files, see Red Hat Linux documentation.
			0x0010: Fatal failures. Issue with invoking or running SSSD.
			
			0x0020: Critical failures. SSSD does not stop functioning. However, this error indicates that at least one major feature of SSSD is not to work properly.
			
			0x0040: Serious failures. A particular request or operation has failed.
			
			0x0080: Minor failures. These are the errors that would percolate down to cause the operation failure of 2.
			
			0x0100: Configuration settings.
			
			0x0200: Function data.
			
			0x0400: Trace messages for operation functions.
			
			0x1000: Trace messages for internal control functions.
			
			0x2000: Contents of function-internal variables that might be interesting.
			
			0x4000: Extremely low-level tracing information.
			Note: For more information on SSSD log levels, see Troubleshooting SSSD in Red Hat Enterprise Linux documentation.

Winbind	`/var/mmfs/ces/smb.conf`	`/var/adm/ras/log.wb-<DOMAIN>`
		
		[Depends upon available domains]
		
		`/var/adm/ras/log.winbindd-dc-connect`
		
		`/var/adm/ras/log.winbindd-idmap`
		
		`/var/adm/ras/log.winbindd`
			Log level is an integer. The value can be from 0-10.
			
			The default value for log level is 1.

Note: Some of the authentication modules like keystone services log information also in `/var/log/messages`.

If you change the log levels, the respective authentication service must be restarted manually on each protocol node. Restarting authentication services might result in disruption of protocol I/O.

CES monitoring and troubleshooting:

You can monitor system health, query events, and perform maintenance and troubleshooting tasks in Cluster Export Services (CES).

System health monitoring

Each CES node runs a separate GPFS process that monitors the network address configuration of the node. If a conflict between the network interface configuration of the node and the current assignments of the CES address pool is found, corrective action is taken. If the node is unable to detect an address that is assigned to it, the address is reassigned to another node.

Additional monitors check the state of the services that are implementing the enabled protocols on the node. These monitors cover NFS, SMB, Object, and Authentication services that monitor, for example, daemon liveliness and port responsiveness. If it is determined that any enabled service is not functioning correctly, the node is marked as `failed` and its CES addresses are reassigned. When the node returns to normal operation, it returns to the normal (healthy) state and is available to host addresses in the CES address pool.

An additional monitor runs on each protocol node if Microsoft Active Directory (AD), Lightweight Directory Access Protocol (LDAP), or Network Information Service (NIS) user authentication is configured. If a configured authentication server does not respond to test requests, GPFS marks the affected node as `failed`.

Querying state and events

Aside from the automatic failover and recovery of CES addresses, two additional outputs are provided by the monitoring that can be queried: events and state.

State can be queried by entering the **`mmces state show`** command, which shows you the state of each of the CES components. The possible states for a component follow:

HEALTHY
The component is working as expected.

DISABLED
The component has not been enabled.

SUSPENDED
When a CES node is in the suspended state, most components also report suspended.

STARTING
The component (or monitor) recently started. This state is a transient state that is updated after the startup is complete.

UNKNOWN
Something is preventing the monitoring from determining the state of the component.

STOPPED
The component was intentionally stopped. This situation might happen briefly if a service is being restarted due to a configuration change. It might also happen because a user ran the `mmces service stop` *`protocol`* command for a node.

DEGRADED
There is a problem with the component but not a complete failure. This state does not cause the CES addresses to be reassigned.

FAILED
The monitoring detected a significant problem with the component that means it is unable to function correctly. This state causes the CES addresses of the node to be reassigned.

DEPENDENCY_FAILED
This state implies that a component has a dependency that is in a failed state. An example would be NFS or SMB reporting `DEPENDENCY_FAILED` because the authentication failed.

Looking at the states themselves can be useful to find out which component is causing a node to fail and have its CES addresses reassigned. To find out why the component is being reported as failed, you can look at events.

The **mmces events** command can be used to show you either events that are currently causing a component to be unhealthy or a list of historical events for the node. If you want to know why a component on a node is in a failed state, use the **mmces events active** invocation. This command gives you a list of any currently active events that are affecting the state of a component, along with a message that describes the problem. This information should provide a place to start when you are trying to find and fix the problem that is causing the failure.

If you want to get a complete idea of what is happening with a node over a longer time period, use the **mmces events list** invocation. By default, this command prints a list of all events that occurred on this node, with a time stamp. This information can be narrowed down by component, time period, and severity. As well as being viewable with the command, all events are also pushed to the syslog.

Maintenance and troubleshooting

A CES node can be marked as unavailable by the monitoring process. The command **mmces node list** can be used to show the nodes and the current state flags that are associated with it. When unavailable (one of the following node flags are set), the node does not accept CES address assignments. The following possible node states can be displayed:

Suspended
Indicates that the node is suspended with the **mmces node suspend** command. When suspended, health monitoring on the node is discontinued. The node remains in the suspended state until it is resumed with the **mmces node resume** command.

Network-down
Indicates that monitoring found a problem that prevents the node from bringing up the CES addresses in the address pool. The state reverts to normal when the problem is corrected. Possible causes for this state are missing or non-functioning network interfaces and network interfaces that are reconfigured so that the node can no longer host the addresses in the CES address pool.

No-shared-root
Indicates that the CES shared root directory cannot be accessed by the node. The state reverts to normal when the shared root directory becomes available. Possible cause for this state is that the file system that contains the CES shared root directory is not mounted.

Failed Indicates that monitoring found a problem with one of the enabled protocol servers. The state reverts to normal when the server returns to normal operation or when the service is disabled.

Starting up
Indicates that the node is starting the processes that are required to implement the CES services that are enabled in the cluster. The state reverts to normal when the protocol servers are functioning.

Additionally, events that affect the availability and configuration of CES nodes are logged in the GPFS log file `/var/adm/ras/mmfs.log.latest`. The verbosity of the CES logging can be changed with the **mmces log level n** command, where *n* is a number from 0 (less logging) to 4 (more logging). The current log level can be viewed with the **mmlscluster --ces** command.

For more information about CES troubleshooting, see the IBM Spectrum Scale Wiki (www.ibm.com/developerworks/community/wikis/home/wiki/General Parallel File System (GPFS)).

Operating system error logs

GPFS records file system or disk failures using the error logging facility provided by the operating system: **syslog** facility on Linux, **errpt** facility on AIX, and Event Viewer on Windows.

The error logging facility is referred to as *the error log* regardless of operating-system specific error log facility naming conventions.

Note: Most logs use the UNIX command **logrotate** to tidy up older logs. Not all options of the command are supported on some older operating systems. This could lead to unnecessary log entries. However, it does not interfere with the script. While using **logrotate** you might come across the following errors:

- `error opening /var/adm/ras/mmsysmonitor.log:Too many levels of symbolic links.`
- `unknown option 'maxsize' -- ignoring line.`

This is the expected behavior and the error can be ignored.

Failures in the error log can be viewed by issuing this command on an AIX node:

```
errpt -a
```

and this command on a Linux node:

```
grep "mmfs:" /var/log/messages
```

You can also grep the appropriate filename where syslog messages are redirected to. For example, in Ubuntu, after the Natty release, this file will be at `/var/log/syslog`

On Windows, use the Event Viewer and look for events with a source label of **GPFS** in the **Application** event category.

On Linux, **syslog** may include GPFS log messages and the error logs described in this section. The **systemLogLevel** attribute of the **mmchconfig** command controls which GPFS log messages are sent to **syslog**. It is recommended that some kind of monitoring for GPFS log messages be implemented, particularly MMFS_FSSTRUCT errors. For more information, see the **mmchconfig** command in the *IBM Spectrum Scale: Command and Programming Reference*.

The error log contains information about several classes of events or errors. These classes are:

- "MMFS_ABNORMAL_SHUTDOWN"
- "MMFS_DISKFAIL"
- "MMFS_ENVIRON" on page 215
- "MMFS_FSSTRUCT" on page 215
- "MMFS_GENERIC" on page 215
- "MMFS_LONGDISKIO" on page 216
- "MMFS_QUOTA" on page 216
- "MMFS_SYSTEM_UNMOUNT" on page 217
- "MMFS_SYSTEM_WARNING" on page 217

MMFS_ABNORMAL_SHUTDOWN: The **MMFS_ABNORMAL_SHUTDOWN** error log entry means that GPFS has determined that it must shutdown all operations on this node because of a problem. Insufficient memory on the node to handle critical recovery situations can cause this error. In general there will be other error log entries from GPFS or some other component associated with this error log entry.

MMFS_DISKFAIL:

This topic describes the MMFS_DISKFAIL error log available in IBM Spectrum Scale.

The **MMFS_DISKFAIL** error log entry indicates that GPFS has detected the failure of a disk and forced the disk to the stopped state. This is ordinarily not a GPFS error but a failure in the disk subsystem or the path to the disk subsystem.

MMFS_ENVIRON:

This topic describes the MMFS_ENVIRON error log available in IBM Spectrum Scale.

MMFS_ENVIRON error log entry records are associated with other records of the **MMFS_GENERIC** or **MMFS_SYSTEM_UNMOUNT** types. They indicate that the root cause of the error is external to GPFS and usually in the network that supports GPFS. Check the network and its physical connections. The data portion of this record supplies the return code provided by the communications code.

MMFS_FSSTRUCT:

This topic describes the MMFS_FSSTRUCT error log available in IBM Spectrum Scale.

The **MMFS_FSSTRUCT** error log entry indicates that GPFS has detected a problem with the on-disk structure of the file system. The severity of these errors depends on the exact nature of the inconsistent data structure. If it is limited to a single file, **EIO** errors will be reported to the application and operation will continue. If the inconsistency affects vital metadata structures, operation will cease on this file system. These errors are often associated with an **MMFS_SYSTEM_UNMOUNT** error log entry and will probably occur on all nodes. If the error occurs on all nodes, some critical piece of the file system is inconsistent. This can occur as a result of a GPFS error or an error in the disk system.

Note: When an `fsstruct` error is show in mmhealth, you are asked to run a filesystem check. Once the problem is solved, you need to clear the `fsstruct` error from mmhealth manually by running the following command:

```
mmsysmonc event filesystem fsstruct_fixed <filesystem_name>
```

If the file system is severely damaged, the best course of action is to follow the procedures in "Additional information to collect for file system corruption or MMFS_FSSTRUCT errors" on page 470, and then contact the IBM Support Center.

MMFS_GENERIC:

This topic describes MMFS_GENERIC error logs available in IBM Spectrum Scale.

The **MMFS_GENERIC** error log entry means that GPFS self diagnostics have detected an internal error, or that additional information is being provided with an **MMFS_SYSTEM_UNMOUNT** report. If the record is associated with an **MMFS_SYSTEM_UNMOUNT** report, the event code fields in the records will be the same. The error code and return code fields might describe the error. See "Messages" on page 561 for a listing of codes generated by GPFS.

If the error is generated by the self diagnostic routines, service personnel should interpret the return and error code fields since the use of these fields varies by the specific error. Errors caused by the self checking logic will result in the shutdown of GPFS on this node.

MMFS_GENERIC errors can result from an inability to reach a critical disk resource. These errors might look different depending on the specific disk resource that has become unavailable, like logs and allocation maps. This type of error will usually be associated with other error indications. Other errors generated by disk subsystems, high availability components, and communications components at the same time as, or immediately preceding, the GPFS error should be pursued first because they might be the cause of these errors. **MMFS_GENERIC** error indications without an associated error of those types represent a GPFS problem that requires the IBM Support Center.

Before you contact IBM support center, see "Information to be collected before contacting the IBM Support Center" on page 469.

MMFS_LONGDISKIO:

This topic describes the MMFS_LONGDISKIO error log available in IBM Spectrum Scale.

The **MMFS_LONGDISKIO** error log entry indicates that GPFS is experiencing very long response time for disk requests. This is a warning message and can indicate that your disk system is overloaded or that a failing disk is requiring many I/O retries. Follow your operating system's instructions for monitoring the performance of your I/O subsystem on this node and on any disk server nodes that might be involved. The data portion of this error record specifies the disk involved. There might be related error log entries from the disk subsystems that will pinpoint the actual cause of the problem. If the disk is attached to an AIX node, refer to AIX in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/ssw_aix/welcome) and search for *performance management*. To enable or disable, use the **mmchfs -w** command. For more details, contact the IBM Support Center.

The **mmpmon** command can be used to analyze I/O performance on a per-node basis. For more information, see "Monitoring GPFS I/O performance with the mmpmon command" on page 4 and "Failures using the mmpmon command" on page 343.

MMFS_QUOTA:

This topic describes the MMFS_QUOTA error log available in IBM Spectrum Scale.

The **MMFS_QUOTA** error log entry is used when GPFS detects a problem in the handling of quota information. This entry is created when the quota manager has a problem reading or writing the quota file. If the quota manager cannot read all entries in the quota file when mounting a file system with quotas enabled, the quota manager shuts down but file system manager initialization continues. Mounts will not succeed and will return an appropriate error message (see "File system forced unmount" on page 322).

Quota accounting depends on a consistent mapping between user names and their numeric identifiers. This means that a single user accessing a quota enabled file system from different nodes should map to the same numeric user identifier from each node. Within a local cluster this is usually achieved by ensuring that **/etc/passwd** and **/etc/group** are identical across the cluster.

When accessing quota enabled file systems from other clusters, you need to either ensure individual accessing users have equivalent entries in **/etc/passwd** and **/etc/group**, or use the user identity mapping facility as outlined in the IBM white paper entitled *UID Mapping for GPFS in a Multi-cluster Environment* in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/SSFKCN/com.ibm.cluster.gpfs.doc/gpfs_uid/uid_gpfs.html).

It might be necessary to run an offline quota check (**mmcheckquota**) to repair or recreate the quota file. If the quota file is corrupted, **mmcheckquota** will not restore it. The file must be restored from the backup copy. If there is no backup copy, an empty file can be set as the new quota file. This is equivalent to recreating the quota file. To set an empty file or use the backup file, issue the **mmcheckquota** command with the appropriate operand:

- **-u** *UserQuotaFilename* for the user quota file
- **-g** *GroupQuotaFilename* for the group quota file
- **-j** *FilesetQuotaFilename* for the fileset quota file

After replacing the appropriate quota file, reissue the **mmcheckquota** command to check the file system inode and space usage.

For information about running the **mmcheckquota** command, see "The mmcheckquota command" on page 262.

MMFS_SYSTEM_UNMOUNT:

This topic describes the MMFS_SYSTEM_UNMOUNT error log available in IBM Spectrum Scale.

The **MMFS_SYSTEM_UNMOUNT** error log entry means that GPFS has discovered a condition that might result in data corruption if operation with this file system continues from this node. GPFS has marked the file system as disconnected and applications accessing files within the file system will receive **ESTALE** errors. This can be the result of:

- The loss of a path to all disks containing a critical data structure.

  If you are using SAN attachment of your storage, consult the problem determination guides provided by your SAN switch vendor and your storage subsystem vendor.
- An internal processing error within the file system.

See "File system forced unmount" on page 322. Follow the problem determination and repair actions specified.

MMFS_SYSTEM_WARNING:

This topic describes the MMFS_SYSTEM_WARNING error log available in IBM Spectrum Scale.

The **MMFS_SYSTEM_WARNING** error log entry means that GPFS has detected a system level value approaching its maximum limit. This might occur as a result of the number of inodes (files) reaching its limit. If so, issue the **mmchfs** command to increase the number of inodes for the file system so there is at least a minimum of 5% free.

Error log entry example:

This topic describes an example of an error log entry in IBM Spectrum Scale.

This is an example of an error log entry that indicates a failure in either the storage subsystem or communication subsystem:

```
LABEL: MMFS_SYSTEM_UNMOUNT
IDENTIFIER: C954F85D

Date/Time: Thu Jul 8 10:17:10 CDT
Sequence Number: 25426
Machine Id: 000024994C00
Node Id: nos6
Class: S
Type: PERM
Resource Name: mmfs

Description
STORAGE SUBSYSTEM FAILURE

Probable Causes
STORAGE SUBSYSTEM
COMMUNICATIONS SUBSYSTEM

Failure Causes
STORAGE SUBSYSTEM
COMMUNICATIONS SUBSYSTEM

Recommended Actions
CONTACT APPROPRIATE SERVICE REPRESENTATIVE
```

```
Detail Data
EVENT CODE
15558007
STATUS CODE
212
VOLUME
gpfsd
```

Transparent cloud tiering logs

This topic describes how to collect logs that are associated with Transparent cloud tiering.

To collect details of issues specific to Transparent cloud tiering, issue this command:

```
gpfs.snap [--cloud-gateway {BASIC | FULL}]
```

With the BASIC option, the Transparent cloud tiering service debugs information such as logs, traces, Java™ cores, along with minimal system and IBM Spectrum Scale cluster information is collected. No customer sensitive information is collected.

With the FULL option, extra details such as Java Heap dump are collected, along with the information captured with the BASIC option.

Successful invocation of this command generates a new .tar file at a specified location, and the file can be shared with IBM support team to debug a field issue.

Performance monitoring tool logs

The performance monitoring tool logs can be found in the `/var/log/zimon` directory on each node configured for performance monitoring.

The nodes that are configured as **Collector** have two files in this directory: `ZIMonCollector.log` and `ZIMonSensors.log`. For nodes configured as **Sensor**, only the `ZIMonSensors.log` file is present. These log files contain information, warning, and error messages for the collector service **pmcollector**, and the sensor service **pmsensors**.

Both log files are rotated every day. The previous logs are compressed and saved in the same `/var/log/zimon` directory.

During installation, the log level is set to `info`. Issue the **mmperfmon config show** command to see the current log level as shown in the following sample output:

```
# mmperfmon config show
cephMon = "/opt/IBM/zimon/CephMonProxy"
cephRados = "/opt/IBM/zimon/CephRadosProxy"
colCandidates = "nsd003st001", "nsd004st001"
colRedundancy = 2
collectors = {
host =""
port = "4739"
}
config = "/opt/IBM/zimon/ZIMonSensors.cfg"
ctdbstat = ""
daemonize = T
hostname = ""
ipfixinterface = "0.0.0.0"
logfile = "/var/log/zimon/ZIMonSensors.log"
loglevel = "info"
```

File audit logging logs

All major actions performed in configuring, starting, and stopping the message queue and file audit logging produce messages in multiple logs at varying degrees of granularity.

Some logs are specifically associated with the commands used in IBM Spectrum Scale (**mmmsgqueue** and **mmaudit**) and the actions by the producers and consumers that are specific to IBM Spectrum Scale. Other logs are directly associated with the underlying Kafka infrastructure. Because some logs might grow quickly, log rotation is used for all logs. Therefore, it is important to gather logs as soon as an issue is found and to look for logs that are not by default captured with **gpfs.snap** (no gzip versions of old logs are gathered by default via **gpfs.snap**). The following is a list of the types of logs that are useful for problem determination:

- Kafka logs: These are logs associated with the Kafka servers and brokers. They log items such as error messages from the brokers and state transitions. In general, these logs are located at `/opt/kafka/<kafka_version>/log`. For the first release, the value of <kafk_version> is `kafka_2.11-0.10.1.1`. A subset of these log files are collected by **gpfs.snap**.
- **mmmsgqueue** log: This log contains information regarding the set up and configuration operations that affect the message queue. Information is put into this log on any node containing a broker and/or ZooKeeper. This log is located at `/var/adm/ras/mmmsgqueue.log`. This log is collected via **gpfs.snap**.
- **mmaudit** log: This log contains information regarding the set up and configuration operations that affect file audit logging. Information is put into this log on any node running the file audit logging command or location where the subcommand might be run (such as a consumer). This log is located at `/var/adm/ras/mmaudit.log`. This log is collected via **gpfs.snap**.
- `mmfs.log.latest`: This is the most current version of the IBM Spectrum Scale daemon log. It contains entries from when major message queue or file audit logging activity occurs. Types of activity within this log file are enable or disable of the message queue, enable or disable of file audit logging for a given device, error is encountered when attempting to enable or disable the message queue or file audit logging for given device, etc. This log file is collected via **gpfs.snap**.
- `/var/log/messages`: This is a standard Linux log and many processes put information in this log. This log will have messages from Kafka components as well as the producer and consumers that are running on a node. This log is collected via **gpfs.snap**.

The **gpfs.snap** command gathers log files from multiple components including the message queue and file audit logging. The following files are collected as part of the message queue and file audit logging:

- `/opt/kafka/kafka_2.11-0.10.1.1/logs/controller.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/kafka-authorizer.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/kafka-request.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/kafkaServer-gc.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/log-cleaner.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/server.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/state-change.log`
- `/opt/kafka/kafka_2.11-0.10.1.1/logs/zookeeper-gc.log`
- `/var/adm/ras/mmmsgqueue.log`
- `/var/adm/ras/mmaudit.log`

Additionally, three files held in the CCR are saved when the **gpfs.snap** command is run:

- `kafka.broker.server.properties`: This CCR file contains the base configuration for all Kafka broker servers.
- `kafka.zookeeper.server.properties`: This CCR file contains the base configuration for all ZooKeeper servers.
- `spectrum-scale-file-audit.conf`: This CCR file contains the file audit logging configuration for all devices in the local cluster that is being audited.

Setting up core dumps on a client RHEL system

No core dump configuration is set up by IBM Spectrum Scale by default. Core dumps can be configured in a few ways.

core_pattern + ulimit

The simplest way is to change the core_pattern file at `/proc/sys/kernel/core_pattern` and to enable core dumps using the command 'ulimit -c unlimited'. Setting it to something like `/var/log/cores/core.%e.%t.%h.%p` will produce core dumps similar to `core.bash.1236975953.node01.2344` in `/var/log/cores`. This will create core dumps for Linux binaries but will not produce information for Java or Python exceptions.

ABRT

ABRT can be used to produce more detailed output as well as output for Java and Python exceptions.

The following packages should be installed to configure ABRT:

- abrt (Core package)
- abrt-cli (CLI tools)
- abrt-libs (Libraries)
- abrt-addon-ccpp (C/C++ crash handler)
- abrt-addon-python (Python unhandled exception handler)
- abrt-java-connector (Java crash handler)

This overwrites the values stored in core_pattern to pass core dumps to abrt. It then writes this information to the abrt directory configured in `/etc/abrt/abrt.conf`. Python exceptions is caught by the python interpreter automatically importing the abrt.pth file installed in `/usr/lib64/python2.7/site-packages/`. If some custom configuration has changed this behavior, Python dumps may not be created.

To get Java runtimes to report unhandled exceptions through abrt, they must be executed with the command line argument '-agentpath=/usr/lib64/libabrt-java-connector.so'.

Note: Passing exception information to ABRT by using the ABRT library will cause a decrease in the performance of the application.

ABRT Config files

The ability to collect core dumps has been added to gpfs.snap using the '--protocol core' option.

This attempts to gather core dumps from a number of locations:

- If core_pattern is set to dump to a file it will attempt to get dumps from the absolute path or from the root directory (the CWD for all IBM Spectrum Scale processes)
- If core_pattern is set to redirect to abrt it will try to read the `/etc/abrt/abrt.conf` file and read the 'DumpLocation' variable. All files and folders under this directory will be gathered.
- If the 'DumpLocation' value cannot be read then a default of '/var/tmp/abrt' is used.
- If core_pattern is set to use something other than abrt or a file path, core dumps will not be collected for the OS.

Samba can dump to the directory `'/var/adm/ras/cores/'`. Any files in this directory will be gathered.

Verification steps for RHEL: After the setup is complete, please check if the contents of the `/proc/sys/kernel/core_pattern` file are starting with `|/usr/libexec/abrt-hook-ccpp`.

Configuration changes required on protocol nodes to collect core dump data

To collect core dumps for debugging programs in provided packages, these system configuration changes need to be made on all protocol nodes in the cluster.

1. Install the `abrt-cli` RPM if not already installed. For example, run **rpm -qa | grep abrt-cli** to check if it is already installed, or **yum install abrt-cli** to install the RPM.
2. Set **OpenGPGCheck=no** in the **/etc/abrt/abrt-action-save-package-data.conf** file.
3. Set **MaxCrashReportsSize = 0** in the **/etc/abrt/abrt.conf** file.
4. Start (or restart) the abort daemon (for example, run **systemctl start abrtd** to start the abort daemon after a new install, or **systemctl restart abrtd** if the daemon was already running and the values in steps 2 and 3 were changed).

For additional details, see the Documentation about ABRT-specific configuration(https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Deployment_Guide/sect-abrt-configuration-abrt.html).

Additional setup steps applicable for NFS

A core dump might not be generated for code areas where the CES NFS process has changed credentials. To avoid this, do the following steps:

1. Insert the following entry into the `/etc/sysctl.conf` file:
   ```
   fs.suid_dumpable = 2
   ```
2. Issue the following command to refresh with the new configuration:
   ```
   sysctl -p
   ```
3. Verify that `/proc/sys/fs/suid_dumpable` is correctly set:
   ```
   cat /proc/sys/fs/suid_dumpable
   ```

   Note: The system displays the following output if it is correctly set:
   ```
   2
   ```

Setting up an Ubuntu system to capture crash files

This is the procedure for setting up an Ubuntu system for capturing crash files and debugging CES NFS core dump.

This setup is IBM Spectrum Scale version independent and applies to Ubuntu 16.04.1 and 16.04.2.

1. Install apport. For more information, see https://wiki.ubuntu.com/Apport.
2. Modify the `/etc/apport/crashdb.conf` file and comment this line `'problem_types': ['Bug', 'Package'],` as follows:
   ```
   # 'problem_types': ['Bug', 'Package'],
   ```

   Note: After these steps are performed, crash files will be saved to the `/var/crash/` folder.
   Verification steps for Ubuntu: After the setup is completed, verify the following:
   a. Run the **`systemctl status apport.service`** command to check if Apport service is running.
   b. If Apport service is not running, then start it with the **`systemctl start apport.service`** command and again verify if it has started successfully by running the **`systemctl status apport.service`** command.

      Note: Apport modifies the `/proc/sys/kernel/core_pattern` file. Verify the `core_pattern` file content starts with `|/usr/share/apport/apport.`

Trace facility

The IBM Spectrum Scale system includes many different trace points to facilitate rapid problem determination of failures.

IBM Spectrum Scale tracing is based on the kernel trace facility on AIX, embedded GPFS trace subsystem on Linux, and the Windows ETL subsystem on Windows. The level of detail that is gathered by the trace facility is controlled by setting the trace levels using the **mmtracectl** command.

The **mmtracectl** command sets up and enables tracing using default settings for various common problem situations. Using this command improves the probability of gathering accurate and reliable problem determination information. For more information about the **mmtracectl** command, see the *IBM Spectrum Scale: Command and Programming Reference*.

Generating GPFS trace reports

Use the **mmtracectl** command to configure trace-related configuration variables and to start and stop the trace facility on any range of nodes in the GPFS cluster.

To configure and use the trace properly:

1. Issue the **mmlsconfig dataStructureDump** command to verify that a directory for dumps was created when the cluster was configured. The default location for trace and problem determination data is **/tmp/mmfs**. Use **mmtracectl**, as instructed by the IBM Support Center, to set trace configuration parameters as required if the default parameters are insufficient. For example, if the problem results in GPFS shutting down, set the *traceRecycle* variable with **--trace-recycle** as described in the **mmtracectl** command in order to ensure that GPFS traces are performed at the time the error occurs.

   If desired, specify another location for trace and problem determination data by issuing this command:

   ```
   mmchconfig dataStructureDump=path_for_storage_of_dumps
   ```
2. To start the tracing facility on all nodes, issue this command:

   ```
   mmtracectl --start
   ```
3. Re-create the problem.
4. When the event to be captured occurs, stop the trace as soon as possible by issuing this command:

   ```
   mmtracectl --stop
   ```
5. The output of the GPFS trace facility is stored in **/tmp/mmfs**, unless the location was changed using the **mmchconfig** command in Step 1. Save this output.
6. If the problem results in a shutdown and restart of the GPFS daemon, set the *traceRecycle* variable as necessary to start tracing automatically on daemon startup and stop the trace automatically on daemon shutdown.

If the problem requires more detailed tracing, the IBM Support Center might ask you to modify the GPFS trace levels. Use the **mmtracectl** command to establish the required trace classes and levels of tracing. The syntax to modify trace classes and levels is as follows:

```
mmtracectl --set --trace={io | all | def | "Class Level [Class Level ...]"}
```

For example, to tailor the trace level for I/O, issue the following command:

```
mmtracectl --set --trace=io
```

Once the trace levels are established, start the tracing by issuing:

```
mmtracectl --start
```

After the trace data has been gathered, stop the tracing by issuing:

```
mmtracectl --stop
```

To clear the trace settings and make sure tracing is turned off, issue:

```
mmtracectl --off
```

Other possible values that can be specified for the trace *Class* include:

afm
active file management

alloc
disk space allocation

allocmgr
allocation manager

basic
'basic' classes

brl
byte range locks

cksum
checksum services

cleanup
cleanup routines

cmd
ts commands

defrag
defragmentation

dentry
dentry operations

dentryexit
daemon routine entry/exit

disk
physical disk I/O

disklease
disk lease

dmapi
Data Management API

ds data shipping

errlog
error logging

eventsExporter
events exporter

file
file operations

fs file system

fsck
online multinode fsck

ialloc
inode allocation

io physical I/O

kentryexit
kernel routine entry/exit

kernel
kernel operations

klockl
low-level vfs locking

ksvfs
generic kernel vfs information

lock
interprocess locking

log
recovery log

malloc
malloc and free in shared segment

mb mailbox message handling

mmpmon
mmpmon command

mnode
mnode operations

msg
call to routines in SharkMsg.h

mutex
mutexes and condition variables

nsd
network shared disk

perfmon
performance monitors

pgalloc
page allocator tracing

pin
pinning to real memory

pit
parallel inode tracing

quota
quota management

rdma
rdma

sanergy
SANergy

scsi
scsi services

sec
cluster security

shared
shared segments

smb
SMB locks

sp SP message handling

super
super_operations

tasking
tasking system but not Thread operations

thread
operations in Thread class

tm token manager

ts daemon specific code

user1
miscellaneous tracing and debugging

user2
miscellaneous tracing and debugging

vbhvl
behaviorals

vnode
vnode layer of VFS kernel support

vnop
one line per VNOP with all important information

Values that can be specified for the trace *Class*, relating to vdisks, include:

vdb
vdisk debugger

vdisk
vdisk

vhosp
vdisk hospital

For more information about vdisks and IBM Spectrum Scale RAID, see *IBM Spectrum Scale RAID: Administration*.

The trace *Level* can be set to a value from 0 through 14, which represents an increasing level of detail. A value of 0 turns tracing off. To display the trace level in use, issue the **mmfsadm showtrace** command.

On AIX, the **–aix-trace-buffer-size** option can be used to control the size of the trace buffer in memory.

On Linux nodes only, use the **mmtracectl** command to change the following:

- The trace buffer size in blocking mode.
  For example, to set the trace buffer size in blocking mode to 8K, issue:
  ```
  mmtracectl --set --tracedev-buffer-size=8K
  ```
- The raw data compression level.
  For example, to set the trace raw data compression level to the best ratio, issue:
  ```
  mmtracectl --set --tracedev-compression-level=9
  ```
- The trace buffer size in overwrite mode.
  For example, to set the trace buffer size in overwrite mode to 500M, issue:
  ```
  mmtracectl --set --tracedev-overwrite-buffer-size=500M
  ```
- When to overwrite the old data.

For example, to wait to overwrite the data until the trace data is written to the local disk and the buffer is available again, issue:

```
mmtracectl --set --tracedev-write-mode=blocking
```

`--tracedev-write-mode=blocking` specifies that if the trace buffer is full, wait until the trace data is written to the local disk and the buffer becomes available again to overwrite the old data. This is the default. `--tracedev-write-mode=overwrite` specifies that if the trace buffer is full, overwrite the old data.

Note: Before switching between **--tracedev-write-mode=overwrite** and **--tracedev-write-mode=blocking**, or vice versa, run the **mmtracectl --stop** command first. Next, run the **mmtracectl --set --tracedev-write-mode** command to switch to the desired mode. Finally, restart tracing with the **mmtracectl --start** command.

For more information about the **mmtracectl** command, see the *IBM Spectrum Scale: Command and Programming Reference*.

CES tracing and debug data collection

You can collect debugging information in Cluster Export Services.

Data collection (First Time Data Collection): To diagnose the cause of an issue, it might be necessary to gather some extra information from the cluster. This information can then be used to determine the root cause of an issue.

Collection of debugging information, such as configuration files and logs, can be gathered by using the **gpfs.snap** command. This command gathers data about GPFS, operating system information, and information for each of the protocols. Following services can be traced by **gpfs.snap** command:

GPFS + OS
: GPFS configuration and logs plus operating system information such as network configuration or connected drives.

CES Generic protocol information such as configured CES nodes.

NFS CES NFS configuration and logs.

SMB SMB and CTDB configuration and logs.

OBJECT
: Openstack Swift and Keystone configuration and logs.

AUTHENTICATION
: Authentication configuration and logs.

PERFORMANCE
: Dump of the performance monitor database.

Information for each of the enabled protocols is gathered automatically when the **gpfs.snap** command is run. If any protocol is enabled, then information for CES and authentication is gathered.

To gather performance data, add the **--performance** option. The **--performance** option causes **gpfs.snap** to try to collect performance information.

Note: Because this process can take up to 30 minutes to run, gather performance data only if necessary.

If data is only required for one protocol or area, the automatic collection can be bypassed. Provided one or more of the following options to the **--protocol** argument: `smb,nfs,object,ces,auth,none`

If the **--protocol** command is provided, automatic data collection is disabled. If **--protocol smb,nfs** is provided to **gpfs.snap**, only NFS and SMB information is gathered and no CES or Authentication data is collected. To disable all protocol data collection, use the argument **--protocol none**.

Types of tracing:
Tracing is logging at a high level. The command for starting and stopping tracing (**mmprotocoltrace**) supports SMB, Winbind, Network and Object tracing. NFS tracing can be done with a combination of commands.

SMB To start SMB tracing, use the **mmprotocoltrace start smb -c <clientIP>** command. The output looks similar to this example:

```
Trace 'fcb7cb07-c45e-43f8-8f1f-2de50cf15062' created successfully for 'smb'
```

To see the status of the trace command, use the **mmprotocoltrace status smb** command. The output looks similar to this example:

```
Trace ID:    fcb7cb07-c45e-43f8-8f1f-2de50cf15062
State:       Active
User ID:     root
Protocol:    smb
Start Time:  10:57:43 04/03/2016
End Time:    11:07:43 04/03/2016
Client IPs:  10.0.100.42, 10.0.100.43
Origin Node: ch-42.localnet.com
Syscall:      False
Syscall Only: False
Nodes:
     Node Name:      ch-41.localnet.com
     State:          ACTIVE
     Trace Location:  /tmp/mmfs/smb.20160304_105742.trc

     Node Name:      ch-42.localnet.com
     State:          ACTIVE
     Trace Location: /tmp/mmfs/smb.20160304_105742.trc

     Node Name:      ch-43.localnet.com
     State:          ACTIVE
     Trace Location: /tmp/mmfs/smb.20160304_105742.trc
```

To stop the trace the command, use the **mmprotocoltrace stop smb** command:

```
Stopping traces
Trace 'fcb7cb07-c45e-43f8-8f1f-2de50cf15062' stopped for smb
Waiting for traces to complete
Waiting for node 'node1'
Waiting for node 'node2'
Waiting for node 'node3'
Finishing trace 'fcb7cb07-c45e-43f8-8f1f-2de50cf15062'
Successfully copied file from 'node1:/tmp/mmfs/smb.20160304_105742.trc'
Successfully copied file from 'node2:/tmp/mmfs/smb.20160304_105742.trc'
Successfully copied file from 'node3:/tmp/mmfs/smb.20160304_105742.trc'
Trace tar file has been written to '/tmp/mmfs/smb.trace.20160304_105845.tar.gz
```

The `tar` file then includes the log files that contain top-level logs and configuration details of SMB for each node and every connected client for the time period the trace was running for.

Traces time out after a certain amount of time. By default, this time is 10 minutes. The timeout can be changed by using the **-d** argument when you start the trace. When a trace times out, the first node with the timeout ends the trace and writes the location of the collected data into the `mmprotocoltrace` logs. Each other node writes an information message that states that another node ended the trace.

A full usage message for the **mmprotocoltrace** command is printable by using the **-h** argument.

NFS NFS tracing is achieved by increasing the log level, repeating the issue, capturing the log file, and then restoring the log level.

To increase the log level, use the command `mmnfs config change` *`LOG_LEVEL=FULL_DEBUG`*. `mmnfs config change` restarts server on all nodes. You can increase the log level by using **ganesha_mgr**. This command takes effect without restart, only on the node on which the command is run.

You can set the log level to the following values: NULL, FATAL, MAJ, CRIT, WARN, EVENT, INFO, DEBUG, MID_DEBUG, and FULL_DEBUG.

FULL_DEBUG is the most useful for debugging purposes. This command collects large amount of data, straining disk usage and affecting performance.

After the issue is recreated by running the **gpfs.snap** command either with no arguments or with the **--protocol nfs** argument, the NFS logs are captured. The logs can then be used to diagnose any issues.

To return the log level to normal, use the same command but with a lower logging level (the default is EVENT).

Object

The process for tracing the object protocol is similar to NFS. The Object service consists of multiple processes that can be controlled individually.

The Object services use these logging levels, at increasing severity: DEBUG, INFO, AUDIT, WARNING, ERROR, CRITICAL, and TRACE.

Keystone and Authentication

```
mmobj config change --ccrfile keystone.conf --section DEFAULT --property debug
--value True
```

Finer grained control of Keystone logging levels can be specified by updating the Keystone's `logging.conf` file. For information on the logging levels in the `logging.conf` file, see the OpenStack logging.conf documentation (docs.openstack.org/kilo/config-reference/content/section_keystone-logging.conf.html).

Swift Proxy Server

```
mmobj config change --ccrfile proxy-server.conf --section DEFAULT --property
log_level --value DEBUG
```

Swift Account Server

```
mmobj config change --ccrfile account-server.conf --section DEFAULT --property
log_level --value DEBUG
```

Swift Container Server

```
mmobj config change --ccrfile container-server.conf --section DEFAULT --property
log_level --value DEBUG
```

Swift Object Server

```
mmobj config change --ccrfile object-server.conf --section DEFAULT --property
log_level --value DEBUG
```

These commands increase the log level for the particular process to the debug level. After you have re-created the problem, run the **gpfs.snap** command with no arguments or with the **--protocol object** argument.

Then, decrease the log levels again by using the commands that are shown previously but with `--value ERROR` instead of `--value DEBUG`.

Winbind

The Winbind tracing process is similar to SMB tracing. To start Winbind tracing, use the **mmprotocoltrace start winbind** command. The output looks similar to this example:

```
Setting up traces
Trace '05c53397-2783-49e7-aaba-31451375cd6c' created successfully for 'winbind'
```

To see the status of the trace command, use the **mmprotocoltrace status winbind** command. The output looks similar to this example:

```
Trace ID:     05c53397-2783-49e7-aaba-31451375cd6c
State:        ACTIVE
User ID:      root
Protocol:     winbind
Start Time:   11:28:40 17/08/2016
End Time:     11:38:40 17/08/2016
Client IPs:
Origin Node:  ch-41.localnet.com
Syscall:      False
Syscall Only: False
Nodes:
      Node Name:            ch-42.localnet.com
      State:                ACTIVE
      Trace Location:       /tmp/mmfs/winbind.20160817_112840.trc

      Node Name:            ch-41.localnet.com
      State:                ACTIVE
      Trace Location:       /tmp/mmfs/winbind.20160817_112840.trc

      Node Name:            ch-43.localnet.com
      State:                ACTIVE
      Trace Location:       /tmp/mmfs/winbind.20160817_112840.trc
```

To stop the trace the command, use the **mmprotocoltrace stop winbind** command:

```
Stopping traces
Trace '05c53397-2783-49e7-aaba-31451375cd6c' stopped for winbind
Waiting for traces to complete
Waiting for node 'ch-41.localnet.com'
Waiting for node 'ch-42.localnet.com'
Waiting for node 'ch-43.localnet.com'
Finishing trace '05c53397-2783-49e7-aaba-31451375cd6c'
Successfully copied file from 'ch-41.localnet.com:/tmp/mmfs/winbind.20160817_112840.trc'
Successfully copied file from 'ch-42.localnet.com:/tmp/mmfs/winbind.20160817_112840.trc'
Successfully copied file from 'ch-43.localnet.com:/tmp/mmfs/winbind.20160817_112840.trc'
Trace tar file has been written to '/tmp/mmfs/winbind.trace.20160817_112913.tar.gz'
```

Winbind has an integrated logger, which writes important messages during its execution into a specified log file. The logger traces the detailed logging information (level 10) for protocol authentication and times out after 10 minutes. The timeout can be changed by using the **-d** argument when you start the trace.

Collecting trace information: Use the **mmprotocoltrace** command to collect trace information for debugging system problems or performance issues. For more information, see the **mmprotocoltrace** command in the *IBM Spectrum Scale: Command and Programming Reference*.

- "Running a typical trace"
- "Trace timeout" on page 231
- "Trace log files" on page 231
- "Trace configuration file" on page 231
- "Resetting the trace system " on page 233
- "Using advanced options" on page 233

Running a typical trace

The following steps describe how to run a typical trace. It is assumed that the trace system is reset for the type of trace that you want to run: SMB, Network, or Object. The examples use the SMB trace.

1. Before you start the trace, you can check the configuration settings for the type of trace that you plan to run:

   ```
   mmprotocoltrace config smb
   ```

   The response to this command displays the current settings from the trace configuration file. For more information about this file, see the "Trace configuration file" on page 231 subtopic.
2. Clear the trace records from the previous trace of the same type:

   ```
   mmprotocoltrace clear smb
   ```

   This command responds with an error message if the previous state of a trace node is something other than **DONE** or **FAILED**. If this error occurs, follow the instructions in the "Resetting the trace system " on page 233 subtopic.
3. Start the new trace:

   ```
   mmprotocoltrace start smb -c <clientIP>
   ```

   The following response is typical:

   ```
   Trace '3f36dbed-b567-4566-9beb-63b6420bbb2d' created successfully for 'smb'
   ```
4. Check the status of the trace to verify that tracing is active on all the configured nodes:

   ```
   mmprotocoltrace status smb
   ```

   The following response is typical:

   ```
   Trace ID:    fcb7cb07-c45e-43f8-8f1f-2de50cf15062
   State:       Active
   User ID:     root
   Protocol:    smb
   Start Time:  10:57:43 04/03/2016
   End Time:    11:07:43 04/03/2016
   Client IPs:  10.0.100.42, 10.0.100.43
   Origin Node: ch-42.localnet.com
   Syscall:     False
   Syscall Only:False
   Nodes:
        Node Name:      ch-41.localnet.com
        State:          ACTIVE
        Trace Location:  /tmp/mmfs/smb.20160304_105742.trc

        Node Name:      ch-42.localnet.com
        State:          ACTIVE
        Trace Location: /tmp/mmfs/smb.20160304_105742.trc

        Node Name:      ch-43.localnet.com
        State:          ACTIVE
        Trace Location: /tmp/mmfs/smb.20160304_105742.trc
   ```

   To display more status information, add the **-v** (verbose) option:

   ```
    mmprotocoltrace -v status smb
   ```

   If the status of a node is **FAILED**, the node did not start successfully. Look at the logs for the node to determine the problem. After you fix the problem, reset the trace system by following the steps in the "Resetting the trace system " on page 233 subtopic.
5. If all the nodes started successfully, perform the actions that you want to trace. For example, if you are tracing a client IP address, enter commands that create traffic on that client.
6. Stop the trace:

   ```
   mmprotocoltrace stop smb
   ```

   The following response is typical. The last line gives the location of the trace log file:

   ```
   Stopping traces
   Trace '01239483-be84-wev9-a2d390i9ow02' stopped for smb
   Waiting for traces to complete
   ```

```
Waiting for node 'node1'
Waiting for node 'node2'
Finishing trace '01239483-be84-wev9-a2d390i9ow02'
Trace tar file has been written to '/tmp/mmfs/smb.20150513_162322.trc/smb.trace.20150513_162542.tar.gz'
```

If you do not stop the trace, it continues until the trace duration expires. For more information, see the "Trace timeout" subtopic.

7. Look in the trace log files for the results of the trace. For more information, see the "Trace log files" subtopic.

Trace timeout

If you do not stop a trace manually, the trace runs until its trace duration expires. The default trace duration is 10 minutes, but you can set a different value in the **mmprotocoltrace** command. Each node that participates in a trace starts a timeout process that is set to the trace duration. When a timeout occurs, the process checks the trace status. If the trace is active, the process stops the trace, writes the file location to the log file, and exits. If the trace is not active, the timeout process exits.

If a trace stops because of a timeout, look in the log file of each node to find the location of the trace log file. The log entry is similar to the following entry:

```
2015-08-26T16:53:35.885 W:14150:MainThread:TIMEOUT:
         Trace 'd4643ccf-96c1-467d-93f8-9c71db7333b2' tar file located at
         '/tmp/mmfs/smb.20150826_164328.trc/smb.trace.20150826_165334.tar.gz'
```

Trace log files

Trace log files are compressed files in the `/var/adm/ras` directory. The contents of a trace log file depends on the type of trace. The product supports four types of tracing: SMB, Network, Object, and Winbind.

SMB SMB tracing captures System Message Block information. The resulting trace log file contains an `smbd.log` file for each node for which information has been collected and for each client that is connected to this node. A trace captures information for all clients with the specified IP address.

Network
Network tracing calls Wireshark's dumpcap utility to capture network packets. The resulting trace log file contains a `pcappng` file that is readable by Wireshark and other programs. The file name is similar to `bfn22-10g_all_00001_20150907125015.pcap`.

If the **mmprotocoltrace** command specifies a client IP address, the trace captures traffic between that client and the server. If no IP address is specified, the trace captures traffic across all network interfaces of each participating node.

Object
The trace log file contains log files for each node, one for each of the object services.

Object tracing sets the log location in the rsyslog configuration file. For more information about this file, see the description of the *rsyslogconflocation* configuration parameter in the "Trace configuration file" subtopic.

It is not possible to configure an Object trace by clients so that information for all connections is recorded.

Winbind
Winbind tracing collects detailed logging information (level 10) for the winbind component when using it for protocol authentication.

Trace configuration file

Each node in the cluster has its own trace configuration file, which is stored in the `/var/mmfs/ces` directory. The configuration file contains settings for logging and for each type of tracing:

[logging]

filename
The name of the log file.

level The current logging level, which can be `debug`, `info`, `warning`, `error`, or `critical`.

[smb]

defaultloglocation
The default log location that is used by the reset command or when current information is not retrievable.

defaultloglevel
The default log level that is used by the reset command or when current information is not retrievable.

traceloglevel
The log level for tracing.

maxlogsize
The maximum size of the log file in kilobytes.

esttracesize
The estimated trace size in kilobytes.

[network]

numoflogfiles
The maximum number of log files.

logfilesize
The maximum size of the log file in kilobytes.

esttracesize
The estimated trace size in kilobytes.

[object]

defaultloglocation
The default log location that is used by the reset command or when current information is not retrievable.

defaultloglevel
The default log level that is used by the reset command or when current information is not retrievable.

traceloglevel
The log level for tracing.

rsyslogconflocation
The location of the rsyslog configuration file. Rsyslog is a service that is provided by Red Hat, Inc. that redirects log output. The default location is `/etc/rsyslog.d/00-swift.conf`..

esttracesize
The estimated trace size in kilobytes.

[winbind]

defaultlogfiles
The location of the winbind log files. The default location is `/var/adm/ras/log.w*`.

defaultloglevel
The default log level that is used by the reset command or when current information is not retrievable. The value of defaultloglevel is set to 1.

traceloglevel
The log level for tracing. The value for traceloglevel is set to 10.

esttracesize
The estimated trace size in kilobytes. The value of esttracesize is set to 500000.

[syscalls]

args The CLI arguments, used while executing the strace_executable. By default: -T -tt -C.

Resetting the trace system

Before you run a new trace, verify that the trace system is reset for the type of trace that you want to run: SMB, Network, or Object. The examples in the following instructions use the SMB trace system. To reset the trace system, follow these steps:

1. Stop the trace if it is still running.
   a. Check the trace status to see whether the current trace is stopped on all the nodes:
      ```
      mmprotocoltrace status smb
      ```
      If the trace is still running, stop it:
      ```
      mmprotocoltrace stop smb
      ```
2. Clear the trace records:
   ```
   mmprotocoltrace clear smb
   ```
   If the command is successful, then you have successfully reset the trace system. Skip to the last step in these instructions.
   If the command returns an error message, go to the next step.

   Note: The command responds with an error message if the trace state of a node is something other than **DONE** or **FAILED**. You can verify the trace state of the nodes by running the **status** command:
   ```
   mmprotocoltrace status smb
   ```
3. Run the clear command again with the **-f** (force) option.
   ```
   mmprotocoltrace -f clear smb
   ```
4. After a forced clear, the trace system might still be in an invalid state. Run the reset command. For more information about the command, see the "Using advanced options."
   ```
   mmprotocoltrace reset smb
   ```
5. Check the default values in the trace configuration file to verify that they are correct. To display the values in the trace configuration file, run the config command. For more information about the file, see the "Trace configuration file" on page 231 subtopic.
   ```
   mmprotocoltrace config smb
   ```
6. The trace system is ready. You can now start a new trace.

Using advanced options

The **reset** command restores the trace system to the default values that are set in the trace configuration file. The command also performs special actions for each type of trace:

- For an SMB trace, the reset removes any IP-specific configuration files and sets the log level and log location to the default values.
- For a Network trace, the reset stops all dumpcap processes.
- For an Object trace, the reset sets the log level to the default value. It then sets the log location to the default location in the rsyslog configuration file, and restarts the rsyslog service.

The following command resets the SMB trace:

```
mmprotocoltrace reset smb
```

The **status** command with the **-v** (verbose) option provides more trace information, including the values of trace variables. The following command returns verbose trace information for the SMB trace:

```
mmprotocoltrace -v status smb
```

Tips for using mmprotocoltrace

Follow these tips for **mmprotocoltrace**.

Specifying nodes with the -N and -c parameters.

It is important to understand the difference between the **-N** and **-c** parameters of the **mmprotocoltrace** command:

- The **-N** parameter specifies the CES nodes where you want tracing to be done. The default value is all CES nodes.
- The **-c** parameter specifies the IP addresses of clients whose incoming connections are to be traced. Where these clients are connected to the CES nodes that are specified in the **-N** parameter, those CES nodes trace the connections with the clients.

For example, in the SMB trace started by the following command, the CES node `10.40.72.105` traces incoming connections from clients `192.168.4.1`, `192.168.4.26`, and `192.168.4.22`. The command is all on one line:

```
mmprotocoltrace start smb -c 192.168.4.1,192.168.4.26,192.168.4.22
          -N 10.40.72.105
```

Discovering client IP addresses for an smb trace

If you have only a few clients that you want to trace, you can list their IP addresses by running the system command **smbstatus** on a CES node. This command lists the IP addresses of all smb clients that are connected to the node.

However, if many clients are connected to the CES node, running **smbstatus** on the node to discover client IP addresses might not be practical. The command sets a global lock on the node for the entire duration of the command, which might be a long time if many clients are connected.

Instead, run the system command **ip** on each client that you are interested in and filter the results according to the type of device that you are looking for. In the following example, the command is run on client `ch-41` and lists the IP address `10.0.100.41` for that client:

```
[root@ch-41 ~]# ip a | grep "inet "
    inet 127.0.0.1/8 scope host lo
    inet 10.0.100.41/24 brd 10.255.255.255 scope global eth0
```

A client might have more than one IP address, as in the following example where the command **ip** is run on client `ch-44`:

```
[root@ch-44 ~]# ip a | grep "inet "
    inet 127.0.0.1/8 scope host lo
    inet 10.0.100.44/24 brd 10.255.255.255 scope global eth0
    inet 192.168.4.1/16 brd 192.168.255.255 scope global eth1
    inet 192.168.4.26/16 brd 192.168.255.255 scope global secondary eth1:0
    inet 192.168.4.22/16 brd 192.168.255.255 scope global secondary eth1:1
```

In such a case, specify all the possible IP addresses in the **mmprotocoltrace** command because you cannot be sure which IP address the client will use. The following example specifies all the IP addresses that the previous example listed for client `ch-44`, and by default all CES nodes will trace incoming connections from any of these IP addresses:

```
mmprotocoltrace start smb -c 10.0.100.44,192.168.4.1,192.168.4.26,192.168.4.22
```

Collecting diagnostic data through GUI

IBM Support might ask you to collect logs, trace files, and dump files from the system to help them resolve a problem. You can perform this task from the management GUI or by using the **gpfs.snap** command. Use the **Settings** > **Diagnostic Data** page in the IBM Spectrum Scale GUI to collect details of the issues reported in the system.

The entire set of diagnostic data available in the system helps to analyze all kinds of IBM Spectrum Scale issues. Depending on the data selection criteria, these files can be large (gigabytes) and might take an hour to download. The diagnostic data is collected from each individual node in a cluster. In a cluster with hundreds of nodes, downloading the diagnostic data might take a long time and the downloaded file can be large in size.

It is always better to reduce the size of the log file as you might need to send it to IBM Support to help fix the issues. You can reduce the size of the diagnostic data file by reducing the scope. The following options are available to reduce the scope of the diagnostic data:

- Include only affected functional areas
- Include only affected nodes
- Reduce the number of days for which the diagnostic data needs to be collected

The following three modes are available in the GUI to select the functional areas of the diagnostic data:

1. Standard diagnostics

   The data that is collected in the standard diagnostics consists of the configuration, status, log files, dumps, and traces in the following functional areas:

   - Core IBM Spectrum Scale
   - Network
   - GUI
   - NFS
   - SMB
   - Object
   - Authentication
   - Cluster export services (CES)
   - Crash dumps

   You can download the diagnostic data for the above functional areas at the following levels:

   - All nodes
   - Specific nodes
   - All nodes within one or more node classes
2. Deadlock diagnostics

   The data that is collected in this category consists of the minimum amount of data that is needed to investigate a deadlock problem.
3. Performance diagnostics

   The data that is collected in this category consists of the system performance details collected from performance monitoring tools. You can only use this option if it is requested by the IBM Support.

The GUI log files contain the issues that are related to GUI and it is smaller in size as well. The GUI log consists of the following types of information:

- Traces from the GUI that contain the information about errors occurred inside GUI code
- Several configuration files of GUI and postgreSQL
- Dump of postgreSQL database that contains IBM Spectrum Scale configuration data and events

- Output of most **mmls*** commands
- Logs from the performance collector

Note: Instead of collecting the diagnostic data again, you can also utilize the diagnostic data that Is collected in the past. You can analyze the relevance of the historic data based on the date on which the issue is reported in the system. Ensure to delete the diagnostic data that is no longer needed to save disk space.

Sharing the diagnostic data with the IBM Support using call home

The call home shares support information and your contact information with IBM on a scheduled basis. The IBM Support monitors the details that are shared by the call home and takes necessary action in case of any issues or potential issues. Enabling call home reduces the response time for the IBM Support to address the issues.

You can also manually upload the diagnostic data that is collected through the **Settings** > **Diagnostic Data** page in the GUI to share the diagnostic data to resolve a Problem Management Record (PMR). To upload data manually, perform the following steps:

1. Go to **Settings** > **Diagnostic Data**.
2. Collect diagnostic data based on the requirement. You can also use the previously collected data for the upload.
3. Select the relevant data set from the **Previously Collected Diagnostic Data** section and then right-click and select **Upload to PMR**.
4. Select the PMR to which the data must be uploaded and then click **Upload**.

CLI commands for collecting issue details

You can issue several CLI commands to collect details of the issues that you might encounter while using IBM Spectrum Scale.

Using the gpfs.snap command

This topic describes the usage of **gpfs.snap** command in IBM Spectrum Scale.

Running the **gpfs.snap** command with no options is similar to running the **gpfs.snap -a** command. It collects data from all nodes in the cluster. This invocation creates a file that is made up of multiple gpfs.snap snapshots. The file that is created includes a master snapshot of the node from which the **gpfs.snap** command was invoked, and non-master snapshots of each of other nodes in the cluster.

If the node on which the **gpfs.snap** command is run is not a file system manager node, the **gpfs.snap** creates a non-master snapshot on the file system manager nodes.

The difference between a master snapshot and a non-master snapshot is the data that is gathered. A master snapshot gathers information from nodes in the cluster. A master snapshot contains all data that a non-master snapshot has. There are two categories of data that is collected:

1. Data that is always gathered by the **gpfs.snap** command for master snapshots and non-master snapshots:
   - "Data gathered by gpfs.snap on all platforms" on page 237
   - "Data gathered by gpfs.snap on AIX" on page 238
   - "Data gathered by gpfs.snap on Linux" on page 238
   - "Data gathered by gpfs.snap on Windows" on page 239
2. Data that is gathered by the **gpfs.snap** command in the case of only a master snapshot. For more information, see "Data gathered by gpfs.snap for a master snapshot" on page 239.

When the **gpfs.snap** command runs with no options, data is collected for each of the enabled protocols. You can turn off the collection of all protocol data and specify the type of protocol information to be collected using the **--protocol** option. For more information, see **gpfs.snap command** in *IBM Spectrum Scale: Command and Programming Reference*.

The following data is gathered by gpfs.snap on Linux for protocols:

- "Data gathered for SMB on Linux" on page 240
- "Data gathered for NFS on Linux" on page 241
- "Data gathered for Object on Linux" on page 241
- "Data gathered for CES on Linux" on page 243
- "Data gathered for authentication on Linux" on page 244
- "Data gathered for performance on Linux" on page 246
-

Data gathered by gpfs.snap on all platforms

These items are always obtained by the **gpfs.snap** command when gathering data for an AIX, Linux, or Windows node:

1. The output of these commands:
   - **ls -l /user/lpp/mmfs/bin**
   - **mmdevdiscover**
   - **tspreparedisk -S**
   - **mmfsadm dump malloc**
   - **mmfsadm dump fs**
   - **df -k**
   - **ifconfig** *interface*
   - **ipcs -a**
   - **ls -l /dev**
   - **mmfsadm dump alloc hist**
   - **mmfsadm dump alloc stats**
   - **mmfsadm dump allocmgr**
   - **mmfsadm dump allocmgr hist**
   - **mmfsadm dump allocmgr stats**
   - **mmfsadm dump cfgmgr**
   - **mmfsadm dump config**
   - **mmfsadm dump dealloc stats**
   - **mmfsadm dump disk**
   - **mmfsadm dump mmap**
   - **mmfsadm dump mutex**
   - **mmfsadm dump nsd**
   - **mmfsadm dump rpc**
   - **mmfsadm dump sgmgr**
   - **mmfsadm dump stripe**
   - **mmfsadm dump tscomm**
   - **mmfsadm dump version**
   - **mmfsadm dump waiters**
   - **netstat** with the **-i**, **-r**, **-rn**, **-s**, and **-v** options

- **ps -edf**
- **vmstat**

2. The contents of these files:
   - **/etc/syslog.conf** or **/etc/syslog-ng.conf**
   - **/tmp/mmfs/internal***
   - **/tmp/mmfs/trcrpt***
   - **/var/adm/ras/mmfs.log.***
   - **/var/mmfs/gen/***
   - **/var/mmfs/etc/***
   - **/var/mmfs/tmp/***
   - **/var/mmfs/ssl/*** except for **complete.map** and **id_rsa** files

Data gathered by gpfs.snap on AIX

This topic describes the type of data that is always gathered by the **gpfs.snap** command on the AIX platform.

These items are always obtained by the **gpfs.snap** command when gathering data for an AIX node:

1. The output of these commands:
   - **errpt -a**
   - **lssrc -a**
   - **lslpp -hac**
   - **no -a**
2. The contents of these files:
   - **/etc/filesystems**
   - **/etc/trcfmt**

Data gathered by gpfs.snap on Linux

This topic describes the type of data that is always gathered by the **gpfs.snap** command on the Linux platform.

Note: The **gpfs.snap** command does not collect installation toolkit logs. You can collect these logs by using the **installer.snap.py** script that is located in the same directory as the installation toolkit. For more information, see *Logging and debugging for installation toolkit* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

These items are always obtained by the **gpfs.snap** command when gathering data for a Linux node:

1. The output of these commands:
   - **dmesg**
   - **fdisk -l**
   - **lsmod**
   - **lspci**
   - **rpm -qa**
   - **rpm --verify gpfs.base**
   - **rpm --verify gpfs.docs**
   - **rpm --verify gpfs.gpl**
   - **rpm --verify gpfs.msg.en_US**
2. The contents of these files:
   - **/etc/filesystems**

- **/etc/fstab**
- **/etc/*release**
- **/proc/cpuinfo**
- **/proc/version**
- **/usr/lpp/mmfs/src/config/site.mcr**
- **/var/log/messages***

Performance monitoring data

The following data is collected to enable performance monitoring diagnosis:

1. The output of these commands:
   - **mmperfmon config show**
   - **ps auxw | grep ZIMon**
   - **service pmsensors status**
   - **service pmcollector status**
   - **du -h /opt/IBM/zimon**
   - **ls -laR /opt/IBM/zimon/data**
2. The contents of these files:
   - **/var/log/zimon/***
   - **/opt/IBM/zimon/*.cfg**

The following data is also collected on Linux on Z:

1. The output of the **dbginfo.sh** tool.
   If s390-tools are installed, then the output of **dbginfo.sh** is captured.
2. The content of these files:
   - **/boot/config-$(active-kernel)** (for example: **/boot/config-3.10.0-123.6.3.el7.s390x**)

Data gathered by gpfs.snap on Windows

This topic describes the type of data that is always gathered by the **gpfs.snap** command on the Windows platform.

These items are always obtained by the **gpfs.snap** command when gathering data for a Windows node:

1. The output from **systeminfo.exe**
2. Any raw trace files ***.tmf** and **mmfs.trc***
3. The ***.pdb** symbols from **/usr/lpp/mmfs/bin/symbols**

Data gathered by gpfs.snap for a master snapshot

This topic describes the type of data that is always gathered by the **gpfs.snap** command for a master snapshot.

When the **gpfs.snap** command is specified with no options, a master snapshot is taken on the node where the command was issued. All of the information from "Data gathered by gpfs.snap on all platforms" on page 237, "Data gathered by gpfs.snap on AIX" on page 238, "Data gathered by gpfs.snap on Linux" on page 238, and "Data gathered by gpfs.snap on Windows" is obtained, as well as this data:

1. The output of these commands:
   - **mmauth**
   - **mmgetstate -a**
   - **mmlscluster**
   - **mmlsconfig**

- **mmlsdisk**
- **mmlsfileset**
- **mmlsfs**
- **mmlspolicy**
- **mmlsmgr**
- **mmlsnode -a**
- **mmlsnsd**
- **mmlssnapshot**
- **mmremotecluster**
- **mmremotefs**
- **tsstatus**

2. The contents of the **/var/adm/ras/mmfs.log.*** file (on all nodes in the cluster)

Performance monitoring data

The master snapshot, when taken on a Linux node, collects the following data:

1. The output of these commands:
   - **mmlscluster**
   - **mmdiag --waiters --iohist --threads --stats --memory**
   - **mmfsadm eventsExporter mmpmon chms**
   - **mmfsadm dump nsd**
   - **mmfsadm dump mb**

Note: The performance monitoring data is only collected if the master node is a Linux node.

Data gathered by gpfs.snap on Linux for protocols

When the **gpfs.snap** command runs with no options, data is collected for each of the enabled protocols.

You can turn off the collection of all protocol data and specify the type of protocol information to be collected using the **--protocol** option..

Data gathered for SMB on Linux:

The following data is always obtained by the `gpfs.snap` command for the server message block (SMB).

1. The output of these commands:
   - **ctdb status**
   - **ctdb scriptstatus**
   - **ctdb ip**
   - **ctdb statistics**
   - **ctdb uptime**
   - **wbinfo -P**
   - **rpm -q gpfs.smb (or dpkg-query on Ubuntu)**
   - **rpm -q samba (or dpkg-query on Ubuntu)**
   - **rpm --verify gpfs.smb (or dpkg-query on Ubuntu)**
   - **net conf list**
   - **sharesec --view-all**
   - **- ps -ef**
   - **- ls -lR /var/lib/samba**

- **mmlsperfdata smb2Throughput -n 1440 -b 60**
- **mmlsperfdata smb2IORate -n 1440 -b 60**
- **mmlsperfdata smb2IOLatency -n 1440 -b 60**
- **mmlsperfdata smbConnections -n 1440 -b 60**
- **ls -l /var/ctdb/CTDB_DBDIR**
- **ls -l /var/ctdb/persistent**

2. The content of these files:
   - /var/adm/ras/log.smbd*
   - /var/adm/ras/log.wb-*
   - /var/var/ras/log.winbindd*
   - /var/adm/ras/cores/smbd/* (Only files from the last 60 days)
   - /var/adm/ras/cores/winbindd/* (Only files from the last 60 days.)
   - /var/lib/samba/*.tdb
   - /var/lib/samba/msg/*
   - /etc/sysconfig/gpfs-ctdb/* (or /etc/default/ctdb on Ubuntu)
   - /var/mmfs/ces/smb.conf
   - /var/mmfs/ces/smb.ctdb.nodes
   - /var/lib/ctdb/persistent/*.tdb* # except of secrets.tdb
   - /etc/sysconfig/ctdb

Data gathered for NFS on Linux:

The following data is always obtained by the gpfs.snap command for NFS.

1. The output of these commands:
   - **mmnfs export list**
   - **mmnfs config list**
   - **rpm -qi - for all installed ganesha packages (or dpkg-query on Ubuntu)**
2. The content of these files:
   - /var/mmfs/ces/nfs-config/*
   - /var/log/ganesha.log*
   - /var/tmp/abrt/* for all sub-directories, not older than 60 days
   - /etc/sysconfig/ganesha

   Files stored in the CCR:
   - gpfs.ganesha.exports.conf
   - gpfs.ganesha.main.conf
   - gpfs.ganesha.nfsd.conf
   - gpfs.ganesha.log.conf
   - gpfs.ganesha.statdargs.conf

Data gathered for Object on Linux:

The following data is always obtained by the gpfs.snap command for Object protocol.

1. The output of these commands:
   - **curl -i http://localhost:8080/info -X GET**
   - **rpm -qi - for all installed openstack rpms (or dpkg-query on Ubuntu)**
   - **ps aux | grep keystone**
2. The content of these files:

- /var/log/swift/account-reaper.log*
- /var/log/swift/account-reaper.error*
- /var/log/swift/account-replicator.log*
- /var/log/swift/account-replicator.error*
- /var/log/swift/account-server.log*
- /var/log/swift/account-server.error*
- /var/log/swift/container-replicator.log*
- /var/log/swift/container-replicator.error*
- /var/log/swift/container-server.log*
- /var/log/swift/container-server.error*
- /var/log/swift/container-updater.log*
- /var/log/swift/container-updater.error*
- /var/log/swift/ibmobjectizer.log*
- /var/log/swift/object-expirer.log*
- /var/log/swift/object-expirer.error*
- /var/log/swift/object-replicator.log*
- /var/log/swift/object-replicator.error*
- /var/log/swift/object-server.log*
- /var/log/swift/object-server.error*
- /var/log/swift/object-updater.log*
- /var/log/swift/object-updater.error*
- /var/log/swift/policyscheduler.log*
- /var/log/swift/proxy-server.log*
- /var/log/swift/proxy-server.error*
- /var/log/swift/swift.log*
- /var/log/swift/swift.error*
- /var/log/keystone/keystone.log*
- /var/log/keystone/httpd-error.log*
- /var/log/keystone/httpd-access.log*
- /var/log/secure/*
- /var/log/httpd/access_log*
- /var/log/httpd/error_log*
- /var/log/httpd/ssl_access_log*
- /var/log/httpd/ssl_error_log*
- /var/log/httpd/ssl_request_log*
- /var/log/messages
- /etc/httpd/conf/httpd.conf
- /etc/httpd/conf.d/ssl.conf
- /etc/keystone/keystone-paste.ini
- /etc/keystone/logging.conf
- /etc/keystone/policy.json
- /etc/keystone/ssl/certs/*

All files stored in the directory specified in the `spectrum-scale-objectizer.conf` CCR file in the **`objectization_tmp_dir`** parameter.

The following files are collected under /var/mmfs/tmp/object.snap by stripping any sensitive information:

- /etc/swift/proxy-server.conf
- /etc/swift/swift.conf
- /etc/keystone/keystone.conf

Files stored in the CCR:

- account-server.conf
- account.builder
- account.ring.gz
- container-server.conf
- container.builder
- container.ring.gz
- object-server.conf
- object*.builder
- object*.ring.gz
- container-reconciler.conf
- spectrum-scale-compression-scheduler.conf
- spectrum-scale-object-policies.conf
- spectrum-scale-objectizer.conf
- spectrum-scale-object.conf
- object-server-sof.conf
- object-expirer.conf
- keystone-paste.ini
- policy*.json
- sso/certs/ldap_cacert.pem
- spectrum-scale-compression-status.stat
- wsgi-keystone.conf

Data gathered for CES on Linux:

The following data is always obtained by the gpfs.snap command for any enabled protocols.

The following data is collected by the gpfs.snap command from by default if any protocols are enabled:

- Information collected for each relevant node:
  1. The output of these commands:
     - sqlite3 -header -csv /var/adm/ras/ras.db
       'SELECT event_time, time_zone, component, name, code, internal_component,
        identifier, severity, event_type, state, message, details FROM events2;'
       # contents of ras.db in the csv format
     - **mmces service list -Y**
     - **mmces service list --verbose -Y**
     - **mmces state show -Y**
     - **mmces events active -Y**
     - **mmhealth node eventlog -Y**
     - **tsctl shownodes up**
  2. The content of these files:
     - /var/adm/ras/mmcesdr.log*
     - /var/adm/ras/mmsysmonitor.*.log*

- /var/adm/ras/mmprotocoltrace.log*

- Information collected once for the cluster:
  1. The output of these commands:
     - **mmces node list**
     - **mmces address list**
     - ls -l <cesSharedRoot>/ces/addrs/*
     - **mmces service list -a**
     - **mmccr flist**
     - **mmlscluster --ces**
  2. The content of the following file:
     <cesSharedRoot>/ces/connections/*
  3. The content of these CCR files:
     - cesiplist
     - ccr.nodes
     - ccr.disks

Data gathered for authentication on Linux:

The following data is always obtained by the gpfs.snap command for any enabled protocol.

1. The output of these commands:
   - **mmcesuserauthlsservice**
   - **mmcesuserauthckservice --data-access-method all --nodes cesNodes**
   - **mmcesuserauthckservice --data-access-method all --nodes cesNodes --server-reachability**
   - **systemctl status ypbind**
   - **systemctl status sssd**
   - **lsof -i**
   - **sestatus**
   - **systemctl status firewalld**
   - **systemstl status iptables**
   - **net ads info**
2. The content of these files:
   - /etc/nsswitch.conf
   - /etc/ypbind.conf
   - /etc/idmapd.conf
   - /etc/krb5.conf
   - /etc/krb5.keytab
   - /etc/firewalld/*
   - /var/log/sssd/*
   - /var/log/secure/*

   Files stored in the CCR:
   - NSSWITCH_CONF
   - YP_CONF
   - LDAP_TLS_CACERT
   - authccr

Data gathered for hadoop on Linux:

The following data is gathered when running **gpfs.snap** with the --hadoop core argument:

1. The output of these commands:
   - **ps -elf**
   - **netstat --nap**
2. The content of these files:
   - /var/log/hadoop
   - /var/log/flume
   - /var/log/hadoop-hdfs
   - /var/log/hadoop-httpfs
   - /var/log/hadoop-mapreduce
   - /var/log/hadoop-yarn
   - /var/log/hbase
   - /var/log/hive
   - /var/log/hive-hcatalog
   - /var/log/kafka
   - /var/log/knox
   - /var/log/oozie
   - /var/log/ranger
   - /var/log/solr
   - /var/log/spark
   - /var/log/sqoop
   - /var/log/zookeeper
   - /usr/lpp/mmfs/hadoop/etc/hadoop
   - /usr/lpp/mmfs/hadoop/logs

   The user can customize hadoop.snap.py to include the user defined files and directories into the snap, by listing these custom files and directories in the environment variable HADOOP_LOG_DIRS. This helps users to set up the hadoop.snap for using custom paths for the hadoop-installation or for including some special files.

   In this case the syntax of the contents of the environment variable HADOOP_LOG_DIRS is:

   ```
   pathname1[;pathname2[;pathname3[...]]]
   ```

   where pathname1..pathnameN are file path names (wildcard usage allowed)/directory path names. For directory path names all files in these directories are collected recursively.

Limitations of customizations when using sudo wrapper

If the sudo wrapper is in use, persistent environment variables, saved in the $HOME/.bashrc in /root/.bashrc, $HOME/.kshrc, /root/.kshrc and similar paths are not initialized when the current non-root gpfsadmin user elevates his rights with the sudo command. Thus gpfs.snap will not be able to detect any customization options for the Hadoop data collection. Starting from IBM Spectrum Scale version 5.0, if you want to apply your customization to the Hadoop debugging data with an active sudo wrapper, you must create symlinks from the actual log files to the corresponding locations mentioned in the list of collected log files.

Data gathered for core dumps on Linux:

The following data is gathered when running **gpfs.snap** with the --protocol core argument:

- If core_pattern is set to dump to a file it will gather files matching that pattern.
- If core_pattern is set to redirect to abrt then everything is gathered from the directory specified in the abrt.conf file under DumpLocation. If this is not set then '/var/tmp/abrt' is used.
- Other core dump mechanisms are not supported by the script.
- Any files in the directory '/var/adm/ras/cores/' will also be gathered.

Data gathered for performance on Linux

The following data is obtained by the `gpfs.snap` command for any enabled protocols, if the option `--performance` is provided.

1. The output of these commands:
   - **top -n 1 -b**
   - **mmdiag --waiters --iohist --threads --stats --memory**
   - **mmfsadm eventsExporter mmpmon chms**
   - **mmfsadm dump nsd**
   - **mmfsadm dump mb**
   - **mmdumpperfdata -r 86400**
2. The content of these files:
   - `/opt/IBM/zimon/*`
   - `/var/log/cnlog/zimon/*`

Note: The performance data collection will work only if the `gpfs.ext` package is installed.

mmdumpperfdata command

Collects and archives the performance metric information.

Synopsis

mmdumpperfdata [**--remove-tree**] [*StartTime EndTime* | *Duration*]

Availability

Available with IBM Spectrum Scale Standard Edition or higher.

Description

The **mmdumpperfdata** command runs all named queries and computed metrics used in the **mmperfmon query** command for each cluster node, writes the output into CSV files, and archives all the files in a single .tgz file. The file name is in the `iss_perfdump_YYYYMMDD_hhmmss.tgz` format.

The tar archive file contains a folder for each cluster node and within that folder there is a text file with the output of each named query and computed metric.

If the start and end time, or duration are not given, then by default the last four hours of metrics information is collected and archived.

Parameters

--remove-tree or -r
Removes the folder structure that was created for the TAR archive file.

StartTime
Specifies the start timestamp for query in the YYYY-MM-DD[-hh:mm:ss] format.

EndTime
Specifies the end timestamp for query in the YYYY-MM-DD[-hh:mm:ss] format.

Duration
Specifies the duration in seconds

Exit status

0 Successful completion.

nonzero
A failure has occurred.

Security

You must have root authority to run the **mmdumpperfdata** command.

The node on which the command is issued must be able to execute remote shell commands on any other node in the cluster without the use of a password and without producing any extraneous messages. For more information, see *Requirements for administering a GPFS file system* in *IBM Spectrum Scale: Administration Guide*.

Examples

1. To archive the performance metric information collected for the default time period of last four hours and also delete the folder structure that the command creates, issue this command:

   ```
   mmdumpperfdata --remove-tree
   ```

   The system displays output similar to this:

```
Using the following options:
tstart  :
tend    :
duration: 14400
rem tree: True
Target folder: ./iss_perfdump_20150513_142420
[1/120] Dumping data for node=fscc-hs21-22 and query q=swiftAccThroughput
  file: ./iss_perfdump_20150513_142420/fscc-hs21-22/swiftAccThroughput
[2/120] Dumping data for node=fscc-hs21-22 and query q=NetDetails
  file: ./iss_perfdump_20150513_142420/fscc-hs21-22/NetDetails
[3/120] Dumping data for node=fscc-hs21-22 and query q=ctdbCallLatency
  file: ./iss_perfdump_20150513_142420/fscc-hs21-22/ctdbCallLatency
[4/120] Dumping data for node=fscc-hs21-22 and query q=usage
  file: ./iss_perfdump_20150513_142420/fscc-hs21-22/usage
```

2. To archive the performance metric information collected for a specific time period, issue this command:

```
mmdumpperfdata --remove-tree 2015-01-25-04:04:04 2015-01-26-04:04:04
```

The system displays output similar to this:

```
Using the following options:
tstart  : 2015-01-25 04:04:04
tend    : 2015-01-26 04:04:04
duration:
rem tree: True
Target folder: ./iss_perfdump_20150513_144344
[1/120] Dumping data for node=fscc-hs21-22 and query q=swiftAccThroughput
  file: ./iss_perfdump_20150513_144344/fscc-hs21-22/swiftAccThroughput
[2/120] Dumping data for node=fscc-hs21-22 and query q=NetDetails
  file: ./iss_perfdump_20150513_144344/fscc-hs21-22/NetDetails
```

3. To archive the performance metric information collected in the last 200 seconds, issue this command:

```
mmdumpperfdata --remove-tree 200
```

The system displays output similar to this:

```
Using the following options:
tstart  :
tend    :
duration: 200
rem tree: True
Target folder: ./iss_perfdump_20150513_144426
[1/120] Dumping data for node=fscc-hs21-22 and query q=swiftAccThroughput
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/swiftAccThroughput
[2/120] Dumping data for node=fscc-hs21-22 and query q=NetDetails
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/NetDetails
[3/120] Dumping data for node=fscc-hs21-22 and query q=ctdbCallLatency
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/ctdbCallLatency
[4/120] Dumping data for node=fscc-hs21-22 and query q=usage
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/usage
[5/120] Dumping data for node=fscc-hs21-22 and query q=smb2IORate
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/smb2IORate
[6/120] Dumping data for node=fscc-hs21-22 and query q=swiftConLatency
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/swiftConLatency
[7/120] Dumping data for node=fscc-hs21-22 and query q=swiftCon
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/swiftCon
[8/120] Dumping data for node=fscc-hs21-22 and query q=gpfsNSDWaits
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/gpfsNSDWaits
[9/120] Dumping data for node=fscc-hs21-22 and query q=smb2Throughput
  file: ./iss_perfdump_20150513_144426/fscc-hs21-22/smb2Throughput
```

See also

For more information, see **mmperfmon** *command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Location

```
/usr/lpp/mmfs/bin
```

mmfsadm command

The **mmfsadm** command is intended for use by trained service personnel. IBM suggests you do not run this command except under the direction of such personnel.

Note: The contents of **mmfsadm** output might vary from release to release, which could obsolete any user programs that depend on that output. Therefore, we suggest that you do not create user programs that invoke **mmfsadm**.

The **mmfsadm** command extracts data from GPFS without using locking, so that it can collect the data in the event of locking errors. In certain rare cases, this can cause GPFS or the node to fail. Several options of this command exist and might be required for use:

cleanup
: Delete shared segments left by a previously failed GPFS daemon without actually restarting the daemon.

dump *what*
: Dumps the state of a large number of internal state values that might be useful in determining the sequence of events. The *what* parameter can be set to **all**, indicating that all available data should be collected, or to another value, indicating more restricted collection of data. The output is presented to STDOUT and should be collected by redirecting STDOUT. For more information about internal GPFS™ states, see the **mmdiag** command in *IBM Spectrum Scale: Command and Programming Reference*.

showtrace
: Shows the current level for each subclass of tracing available in GPFS. Trace level 14 provides the highest level of tracing for the class and trace level 0 provides no tracing. Intermediate values exist for most classes. More tracing requires more storage and results in a higher probability of overlaying the required event.

trace *class n*
: Sets the trace class to the value specified by *n*. Actual trace gathering only occurs when the **mmtracectl** command has been issued.

Other options provide interactive GPFS debugging, but are not described here. Output from the **mmfsadm** command will be required in almost all cases where a GPFS problem is being reported. The **mmfsadm** command collects data only on the node where it is issued. Depending on the nature of the problem, **mmfsadm** output might be required from several or all nodes. The **mmfsadm** output from the file system manager is often required.

To determine where the file system manager is, issue the **mmlsmgr** command:

```
mmlsmgr
```

Output similar to this example is displayed:

```
file system      manager node
---------------- ------------------
fs3              9.114.94.65 (c154n01)
fs2              9.114.94.73 (c154n09)
fs1              9.114.94.81 (c155n01)

Cluster manager node: 9.114.94.65 (c154n01)
```

Commands for GPFS cluster state information

There are a number of GPFS commands used to obtain cluster state information.

The information is organized as follows:

The mmafmctl *Device* getstate command

The **mmafmctl** *Device* **getstate** command displays the status of active file management cache filesets and gateway nodes.

When this command displays a `NeedsResync` target/fileset state, inconsistencies between home and cache are being fixed automatically; however, unmount and mount operations are required to return the state to `Active`.

The **mmafmctl** *Device* **getstate** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

The mmhealth command

The **mmhealth** command monitors and displays the health status of services hosted on nodes and the health status of complete cluster in a single view.

Use the **mmhealth** command to monitor the health of the node and services hosted on the node in IBM Spectrum Scale. If the status of a service hosted on any node is failed, the **`mmhealth`** command allows the user to view the event log to analyze and determine the problem. The **`mmhealth`** command provides list of events responsible for the failure of any service. On detailed analysis of these events a set of troubleshooting steps might be followed to resume the failed service. For more details on troubleshooting, see "How to get started with troubleshooting" on page 185.

The **mmhealth** command is fully described in the *mmhealth command* section in the *IBM Spectrum Scale: Command and Programming Reference* and Chapter 3, "Monitoring system health by using the mmhealth command," on page 109.

The mmdiag command

The **mmdiag** command displays diagnostic information about the internal GPFS state on the current node.

Use the **mmdiag** command to query various aspects of the GPFS internal state for troubleshooting and tuning purposes. The **mmdiag** command displays information about the state of GPFS on the node where it is executed. The command obtains the required information by querying the GPFS daemon process (**mmfsd**), and thus will only function when the GPFS daemon is running.

The **mmdiag** command is fully described in the *Command reference* section in *IBM Spectrum Scale: Command and Programming Reference*.

The mmgetstate command

The **mmgetstate** command displays the state of the GPFS daemon on one or more nodes.

These flags are of interest for problem determination:

-a List all nodes in the GPFS cluster. The option does not display information for nodes that cannot be reached. You may obtain more information if you specify the **-v** option.

-L Additionally display quorum, number of nodes up, and total number of nodes.

The total number of nodes may sometimes be larger than the actual number of nodes in the cluster. This is the case when nodes from other clusters have established connections for the purposes of mounting a file system that belongs to your cluster.

-s Display summary information: number of local and remote nodes that have joined in the cluster, number of quorum nodes, and so forth.

-v Display intermediate error messages.

The remaining flags have the same meaning as in the **mmshutdown** command. They can be used to specify the nodes on which to get the state of the GPFS daemon.

The GPFS *states* recognized and displayed by this command are:

active
GPFS is ready for operations.

arbitrating
A node is trying to form quorum with the other available nodes.

down
GPFS daemon is not running on the node or is recovering from an internal error.

unknown
Unknown value. Node cannot be reached or some other error occurred.

For example, to display the quorum, the number of nodes up, and the total number of nodes, issue:

```
mmgetstate -L -a
```

The system displays output similar to:

```
Node number Node name Quorum Nodes up Total nodes GPFS state Remarks
------------------------------------------------------------------
  2         k154n06     1*     3        7         active     quorum node
  3         k155n05     1*     3        7         active     quorum node
  4         k155n06     1*     3        7         active     quorum node
  5         k155n07     1*     3        7         active
  6         k155n08     1*     3        7         active
  9         k156lnx02   1*     3        7         active
 11         k155n09     1*     3        7         active
```

where *, if present, indicates that tiebreaker disks are being used.

The **mmgetstate** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

The mmlscluster command

The **mmlscluster** command displays GPFS cluster configuration information.

The syntax of the **mmlscluster** command is:

```
mmlscluster
```

The system displays output similar to:

```
GPFS cluster information
========================
  GPFS cluster name:         cluster1.kgn.ibm.com
  GPFS cluster id:           680681562214606028
  GPFS UID domain:           cluster1.kgn.ibm.com
  Remote shell command:      /usr/bin/rsh
  Remote file copy command:  /usr/bin/rcp
  Repository type:           server-based

GPFS cluster configuration servers:
-----------------------------------
  Primary server:    k164n06.kgn.ibm.com
  Secondary server:  k164n05.kgn.ibm.com

 Node  Daemon node name        IP address      Admin node name       Designation
-------------------------------------------------------------------------------------
   1   k164n04.kgn.ibm.com     198.117.68.68   k164n04.kgn.ibm.com   quorum
   2   k164n05.kgn.ibm.com     198.117.68.71   k164n05.kgn.ibm.com   quorum
   3   k164n06.kgn.ibm.com     198.117.68.70   k164n06.kgn.ibm.com   quorum-manager
```

The **mmlscluster** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

The mmlsconfig command

The **mmlsconfig** command displays current configuration data for a GPFS cluster.

Depending on your configuration, additional information not documented in either the **mmcrcluster** command or the **mmchconfig** command may be displayed to assist in problem determination.

If a configuration parameter is not shown in the output of this command, the default value for that parameter, as documented in the **mmchconfig** command, is in effect.

The syntax of the **mmlsconfig** command is:

```
mmlsconfig
```

The system displays information similar to:

```
Configuration data for cluster cl1.cluster:
-------------------------------------------
clusterName cl1.cluster
clusterId 680752107138921233
autoload no
minReleaseLevel 4.1.0.0
pagepool 1G
maxblocksize 4m
[c5n97g]
pagepool 3500m
[common]
cipherList EXP-RC4-MD5

File systems in cluster cl1.cluster:
------------------------------------
/dev/fs2
```

The **mmlsconfig** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

The mmrefresh command

The **mmrefresh** command is intended for use by experienced system administrators who know how to collect data and run debugging routines.

Use the **mmrefresh** command only when you suspect that something is not working as expected and the reason for the malfunction is a problem with the GPFS configuration data. For example, a **mount** command fails with a `device not found` error, and you know that the file system exists. Another example is if any of the files in the **/var/mmfs/gen** directory were accidentally erased. Under normal circumstances, the GPFS command infrastructure maintains the cluster data files automatically and there is no need for user intervention.

The **mmrefresh** command places the most recent GPFS cluster configuration data files on the specified nodes. The syntax of this command is:

```
mmrefresh [-f] [ -a | -N {Node[,Node...] | NodeFile | NodeClass}]
```

The **-f** flag can be used to force the GPFS cluster configuration data files to be rebuilt whether they appear to be at the most current level or not. If no other option is specified, the command affects only the node on which it is run. The remaining flags have the same meaning as in the **mmshutdown** command, and are used to specify the nodes on which the refresh is to be performed.

For example, to place the GPFS cluster configuration data files at the latest level, on all nodes in the cluster, issue:

```
mmrefresh -a
```

The mmsdrrestore command

The **mmsdrrestore** command is intended for use by experienced system administrators.

The **mmsdrrestore** command restores the latest GPFS system files on the specified nodes. If no nodes are specified, the command restores the configuration information only on the node where it is invoked. If the local GPFS configuration file is missing, the file specified with the **-F** option from the node specified with the **-p** option is used instead.

This command works best when used in conjunction with the **mmsdrbackup** user exit, which is described in the *GPFS user exits* topic in the *IBM Spectrum Scale: Command and Programming Reference*.

For more information, see **mmsdrrestore command** in *IBM Spectrum Scale: Command and Programming Reference*.

The mmexpelnode command

The **mmexpelnode** command instructs the cluster manager to expel the target nodes and to run the normal recovery protocol.

The cluster manager keeps a list of the expelled nodes. Expelled nodes will not be allowed to rejoin the cluster until they are removed from the list using the **-r** or **--reset** option on the **mmexpelnode** command. The expelled nodes information will also be reset if the cluster manager node goes down or is changed with **mmchmgr -c**.

The syntax of the **mmexpelnode** command is:

```
mmexpelnode [-o | --once] [-f | --is-fenced] [-w | --wait] -N Node[,Node...]
```

Or,

```
mmexpelnode {-l | --list}
```

Or,

```
mmexpelnode {-r | --reset} -N {all | Node[,Node...]}
```

The flags used by this command are:

-o | --once
Specifies that the nodes should not be prevented from rejoining. After the recovery protocol completes, expelled nodes will be allowed to rejoin the cluster immediately, without the need to first invoke **mmexpelnode --reset**.

-f | --is-fenced
Specifies that the nodes are fenced out and precluded from accessing any GPFS disks without first rejoining the cluster (for example, the nodes were forced to reboot by turning off power). Using this flag allows GPFS to start log recovery immediately, skipping the normal 35-second wait.

Note: The `-f` option should only be used when the administrator is sure that the node being expelled can no longer write to any of the disks. This includes both the locally attached disks and the remote NSDs. The node in question must be down, and it must not have any disk I/O pending on any of the devices. Incorrect use of the `-f` option could lead to file system corruption.

-w | --wait
Instructs the **mmexpelnode** command to wait until GPFS recovery for the failed node has completed before it runs.

-l | --list
Lists all currently expelled nodes.

-r | --reset
Allows the specified nodes to rejoin the cluster (that is, resets the status of the nodes). To unexpel all of the expelled nodes, issue: **mmexpelnode -r -N all**.

-N {all | *Node*[,*Node*...]}
Specifies a list of host names or IP addresses that represent the nodes to be expelled or unexpelled. Specify the daemon interface host names or IP addresses as shown by the **mmlscluster** command. The **mmexpelnode** command does not support administration node names or node classes.

Note: -N all can *only* be used to unexpel nodes.

Examples of the mmexpelnode command

1. To expel node c100c1rp3, issue the command:
   ```
   mmexpelnode -N c100c1rp3
   ```
2. To show a list of expelled nodes, issue the command:
   ```
   mmexpelnode --list
   ```
   The system displays information similar to:
   ```
   Node List
   ---------------------
       192.168.100.35 (c100c1rp3.ppd.pok.ibm.com)
   ```
3. To allow node c100c1rp3 to rejoin the cluster, issue the command:
   ```
   mmexpelnode -r -N c100c1rp3
   ```

GPFS file system and disk information commands

The problem determination tools provided with GPFS for file system, disk and NSD problem determination are intended for use by experienced system administrators who know how to collect data and run debugging routines.

The information is organized as follows:

Restricted mode mount

GPFS provides a capability to mount a file system in a restricted mode when significant data structures have been destroyed by disk failures or other error conditions.

Restricted mode mount is not intended for normal operation, but may allow the recovery of some user data. Only data which is referenced by intact directories and metadata structures would be available.

Attention:

1. Follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center before using this capability.
2. Attempt this only after you have tried to repair the file system with the **mmfsck** command. (See "Why does the offline mmfsck command fail with "Error creating internal storage"?" on page 190.)
3. Use this procedure only if the failing disk is attached to an AIX or Linux node.

Some disk failures can result in the loss of enough metadata to render the entire file system unable to mount. In that event it might be possible to preserve some user data through a *restricted mode mount*. This facility should only be used if a normal mount does not succeed, and should be considered a last resort to save some data after a fatal disk failure.

Restricted mode mount is invoked by using the **mmmount** command with the **-o rs** flags. After a restricted mode mount is done, some data *may* be sufficiently accessible to allow copying to another file system. The success of this technique depends on the actual disk structures damaged.

Read-only mode mount

Some disk failures can result in the loss of enough metadata to make the entire file system unable to mount. In that event, it might be possible to preserve some user data through a *read-only mode mount*.

Attention: Attempt this only after you have tried to repair the file system with the **mmfsck** command.

This facility should be used only if a normal mount does not succeed, and should be considered a last resort to save some data after a fatal disk failure.

Read-only mode mount is invoked by using the **mmmount** command with the **-o ro** flags. After a read-only mode mount is done, some data *may* be sufficiently accessible to allow copying to another file system. The success of this technique depends on the actual disk structures damaged.

The lsof command

The **lsof** (list open files) command returns the user processes that are actively using a file system. It is sometimes helpful in determining why a file system remains in use and cannot be unmounted.

The **lsof** command is available in Linux distributions or by using anonymous ftp from **lsof.itap.purdue.edu** (**cd** to **/pub/tools/unix/lsof**). The inventor of the **lsof** command is Victor A. Abell (abe@purdue.edu), Purdue University Computing Center.

The mmlsmount command

The **mmlsmount** command lists the nodes that have a given GPFS file system mounted.

Use the **-L** option to see the node name and IP address of each node that has the file system in use. This command can be used for all file systems, all remotely mounted file systems, or file systems mounted on nodes of certain clusters.

While not specifically intended as a service aid, the **mmlsmount** command is useful in these situations:

1. When writing and debugging new file system administrative procedures, to determine which nodes have a file system mounted and which do not.
2. When mounting a file system on multiple nodes, to determine which nodes have successfully completed the mount and which have not.
3. When a file system is mounted, but appears to be inaccessible to some nodes but accessible to others, to determine the extent of the problem.
4. When a normal (not force) unmount has not completed, to determine the affected nodes.
5. When a file system has force unmounted on some nodes but not others, to determine the affected nodes.

For example, to list the nodes having all file systems mounted:

```
mmlsmount all -L
```

The system displays output similar to:

```
File system fs2 is mounted on 7 nodes:
  192.168.3.53    c25m3n12                c34.cluster
  192.168.110.73  c34f2n01                c34.cluster
  192.168.110.74  c34f2n02                c34.cluster
  192.168.148.77  c12c4apv7               c34.cluster
  192.168.132.123 c20m2n03                c34.cluster             (internal mount)
  192.168.115.28  js21n92                 c34.cluster             (internal mount)
  192.168.3.124   c3m3n14                 c3.cluster

File system fs3 is not mounted.

File system fs3 (c3.cluster:fs3) is mounted on 7 nodes:
  192.168.2.11    c2m3n01                 c3.cluster
  192.168.2.12    c2m3n02                 c3.cluster
  192.168.2.13    c2m3n03                 c3.cluster
  192.168.3.123   c3m3n13                 c3.cluster
  192.168.3.124   c3m3n14                 c3.cluster
  192.168.110.74  c34f2n02                c34.cluster
  192.168.80.20   c21f1n10                c21.cluster
```

The **mmlsmount** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

The mmapplypolicy -L command

Use the **-L** flag of the **mmapplypolicy** command when you are using policy files to manage storage resources and the data stored on those resources. This command has different levels of diagnostics to help debug and interpret the actions of a policy file.

The **-L** flag, used in conjunction with the **-I test** flag, allows you to display the actions that would be performed by a policy file without actually applying it. This way, potential errors and misunderstandings can be detected and corrected without actually making these mistakes.

These are the trace levels for the **mmapplypolicy -L** flag:

Value	Description
0	Displays only serious errors.
1	Displays some information as the command runs, but not for each file.

2 Displays each chosen file and the scheduled action.

3 Displays the information for each of the preceding trace levels, plus each candidate file and the applicable rule.

4 Displays the information for each of the preceding trace levels, plus each explicitly excluded file, and the applicable rule.

5 Displays the information for each of the preceding trace levels, plus the attributes of candidate and excluded files.

6 Displays the information for each of the preceding trace levels, plus files that are not candidate files, and their attributes.

These terms are used:

candidate file
A file that matches a policy rule.

chosen file
A candidate file that has been scheduled for an action.

This policy file is used in the examples that follow:

```
/* Exclusion rule */
RULE 'exclude *.save files' EXCLUDE WHERE NAME LIKE '%.save'
/* Deletion rule */
RULE 'delete' DELETE FROM POOL 'sp1' WHERE NAME LIKE '%tmp%'
/* Migration rule */
RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WHERE NAME LIKE '%file%'
/* Typo in rule : removed later */
RULE 'exclude 2' EXCULDE
/* List rule */
RULE EXTERNAL LIST 'tmpfiles' EXEC '/tmp/exec.list'
RULE 'all' LIST 'tmpfiles' where name like '%tmp%'
```

These are some of the files in file system **/fs1**:

```
. .. data1  file.tmp0  file.tmp1  file0  file1  file1.save  file2.save
```

The **mmapplypolicy** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*.

mmapplypolicy -L 0:

Use this option to display only serious errors.

In this example, there is an error in the policy file. This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 0
```

produces output similar to this:

```
[E:-1] Error while loading policy rules.
PCSQLERR: Unexpected SQL identifier token - 'EXCULDE'.
PCSQLCTX: at line 8 of 8: RULE 'exclude 2' {{{EXCULDE}}}
mmapplypolicy: Command failed.  Examine previous error messages to determine cause.
```

The error in the policy file is corrected by removing these lines:

```
/* Typo in rule */
RULE 'exclude 2' EXCULDE
```

Now rerun the command:

```
mmapplypolicy fs1 -P policyfile -I test -L 0
```

No messages are produced because no serious errors were detected.

mmapplypolicy -L 1:

Use this option to display all of the information (if any) from the previous level, plus some information as the command runs, but not for each file. This option also displays total numbers for file migration and deletion.

This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 1
```

produces output similar to this:

```
[I] GPFS Current Data Pool Utilization in KB and %
sp1     5120    19531264        0.026214%
system  102400  19531264        0.524288%
[I] Loaded policy rules from policyfile.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP = 2009-03-04@02:40:12 UTC
parsed 0 Placement Rules, 0 Restore Rules, 3 Migrate/Delete/Exclude Rules,
        1 List Rules, 1 External Pool/List Rules
/* Exclusion rule */
RULE 'exclude *.save files' EXCLUDE WHERE NAME LIKE '%.save'
/* Deletion rule */
RULE 'delete' DELETE FROM POOL 'sp1' WHERE NAME LIKE '%tmp%'
/* Migration rule */
RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WHERE NAME LIKE '%file%'
/* List rule */
RULE EXTERNAL LIST 'tmpfiles' EXEC '/tmp/exec.list'
RULE 'all' LIST 'tmpfiles' where name like '%tmp%'
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
[I] Inodes scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
[I] Summary of Rule Applicability and File Choices:
 Rule#  Hit_Cnt KB_Hit  Chosen  KB_Chosen       KB_Ill  Rule
  0     2       32      0       0       0       RULE 'exclude *.save files' EXCLUDE WHERE(.)
  1     2       16      2       16      0       RULE 'delete' DELETE FROM POOL 'sp1' WHERE(.)
  2     2       32      2       32      0       RULE 'migration to system pool' MIGRATE FROM POOL \
                                                 'sp1' TO POOL 'system' WHERE(.)
  3     2       16      2       16      0       RULE 'all' LIST 'tmpfiles' WHERE(.)

[I] Files with no applicable rules: 5.

[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 32KB: 2 of 2 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 16KB: 2 of 2 candidates;
 Chose to list 16KB: 2 of 2 candidates;
 0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
sp1     5072    19531264        0.025969%
system  102432  19531264        0.524451%
```

mmapplypolicy -L 2:

Use this option to display all of the information from the previous levels, plus each chosen file and the scheduled migration or deletion action.

This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 2
```

produces output similar to this:

```
[I] GPFS Current Data Pool Utilization in KB and %
sp1     5120    19531264        0.026214%
system  102400  19531264        0.524288%
[I] Loaded policy rules from policyfile.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP = 2009-03-04@02:43:10 UTC
parsed 0 Placement Rules, 0 Restore Rules, 3 Migrate/Delete/Exclude Rules,
        1 List Rules, 1 External Pool/List Rules
/* Exclusion rule */
RULE 'exclude *.save files' EXCLUDE WHERE NAME LIKE '%.save'
/* Deletion rule */
RULE 'delete' DELETE FROM POOL 'sp1' WHERE NAME LIKE '%tmp%'
/* Migration rule */
RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WHERE NAME LIKE '%file%'
/* List rule */
RULE EXTERNAL LIST 'tmpfiles' EXEC '/tmp/exec.list'
RULE 'all' LIST 'tmpfiles' where name like '%tmp%'
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
[I] Inodes scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp1 SHOW()
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp0 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp1 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp0 SHOW()
WEIGHT(INF) MIGRATE /fs1/file1 TO POOL system SHOW()
WEIGHT(INF) MIGRATE /fs1/file0 TO POOL system SHOW()
[I] Summary of Rule Applicability and File Choices:
 Rule#  Hit_Cnt KB_Hit  Chosen  KB_Chosen       KB_Ill  Rule
  0     2       32      0       0       0       RULE 'exclude *.save files' EXCLUDE WHERE(.)
  1     2       16      2       16      0       RULE 'delete' DELETE FROM POOL 'sp1' WHERE(.)
  2     2       32      2       32      0       RULE 'migration to system pool' MIGRATE FROM POOL \
                                                 'sp1' TO POOL 'system' WHERE(.)
  3     2       16      2       16      0       RULE 'all' LIST 'tmpfiles' WHERE(.)

[I] Files with no applicable rules: 5.

[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 32KB: 2 of 2 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 16KB: 2 of 2 candidates;
 Chose to list 16KB: 2 of 2 candidates;
 0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
sp1     5072    19531264        0.025969%
system  102432  19531264        0.524451%
```

where the lines:

```
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp1 SHOW()
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp0 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp1 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp0 SHOW()
WEIGHT(INF) MIGRATE /fs1/file1 TO POOL system SHOW()
WEIGHT(INF) MIGRATE /fs1/file0 TO POOL system SHOW()
```

show the chosen files and the scheduled action.

mmapplypolicy -L 3:

Use this option to display all of the information from the previous levels, plus each candidate file and the applicable rule.

This command:

```
mmapplypolicy fs1-P policyfile -I test -L 3
```

produces output similar to this:

```
[I] GPFS Current Data Pool Utilization in KB and %
sp1     5120    19531264        0.026214%
system  102400  19531264        0.524288%
[I] Loaded policy rules from policyfile.
Evaluating MIGRATE/DELETE/EXCLUDE rules with CURRENT_TIMESTAMP = 2009-03-04@02:32:16 UTC
parsed 0 Placement Rules, 0 Restore Rules, 3 Migrate/Delete/Exclude Rules,
        1 List Rules, 1 External Pool/List Rules
/* Exclusion rule */
RULE 'exclude *.save files' EXCLUDE WHERE NAME LIKE '%.save'
/* Deletion rule */
RULE 'delete' DELETE FROM POOL 'sp1' WHERE NAME LIKE '%tmp%'
/* Migration rule */
RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WHERE NAME LIKE '%file%'
/* List rule */
RULE EXTERNAL LIST 'tmpfiles' EXEC '/tmp/exec.list'
RULE 'all' LIST 'tmpfiles' where name like '%tmp%'
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
/fs1/file.tmp1   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp1   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file.tmp0   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp0   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file1       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
/fs1/file0       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
[I] Inodes scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp1 SHOW()
WEIGHT(INF) LIST 'tmpfiles' /fs1/file.tmp0 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp1 SHOW()
WEIGHT(INF) DELETE /fs1/file.tmp0 SHOW()
WEIGHT(INF) MIGRATE /fs1/file1 TO POOL system SHOW()
WEIGHT(INF) MIGRATE /fs1/file0 TO POOL system SHOW()
[I] Summary of Rule Applicability and File Choices:
 Rule#  Hit_Cnt KB_Hit  Chosen  KB_Chosen       KB_Ill  Rule
  0     2       32      0       0       0       RULE 'exclude *.save files' EXCLUDE WHERE(.)
  1     2       16      2       16      0       RULE 'delete' DELETE FROM POOL 'sp1' WHERE(.)
  2     2       32      2       32      0       RULE 'migration to system pool' MIGRATE FROM POOL \
                                                 'sp1' TO POOL 'system' WHERE(.)
  3     2       16      2       16      0       RULE 'all' LIST 'tmpfiles' WHERE(.)

[I] Files with no applicable rules: 5.

[I] GPFS Policy Decisions and File Choice Totals:
 Chose to migrate 32KB: 2 of 2 candidates;
 Chose to premigrate 0KB: 0 candidates;
 Already co-managed 0KB: 0 candidates;
 Chose to delete 16KB: 2 of 2 candidates;
 Chose to list 16KB: 2 of 2 candidates;
 0KB of chosen data is illplaced or illreplicated;
Predicted Data Pool Utilization in KB and %:
sp1     5072    19531264            0.025969%
system  102432  19531264        0.524451%
```

where the lines:

```
/fs1/file.tmp1   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp1   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file.tmp0   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp0   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file1       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
/fs1/file0       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
```

show the candidate files and the applicable rules.

mmapplypolicy -L 4:

Use this option to display all of the information from the previous levels, plus the name of each explicitly excluded file, and the applicable rule.

This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 4
```

produces the following additional information:

```
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
/fs1/file1.save  RULE 'exclude *.save files' EXCLUDE
/fs1/file2.save  RULE 'exclude *.save files' EXCLUDE
/fs1/file.tmp1   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp1   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file.tmp0   RULE 'delete' DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp0   RULE 'all' LIST 'tmpfiles' WEIGHT(INF)
/fs1/file1       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
/fs1/file0       RULE 'migration to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
```

where the lines:

```
/fs1/file1.save  RULE 'exclude *.save files' EXCLUDE
/fs1/file2.save  RULE 'exclude *.save files' EXCLUDE
```

indicate that there are two excluded files, **/fs1/file1.save** and **/fs1/file2.save**.

mmapplypolicy -L 5:

Use this option to display all of the information from the previous levels, plus the attributes of candidate and excluded files.

These attributes include:

- **MODIFICATION_TIME**
- **USER_ID**
- **GROUP_ID**
- **FILE_SIZE**
- **POOL_NAME**
- **ACCESS_TIME**
- **KB_ALLOCATED**
- **FILESET_NAME**

This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 5
```

produces the following additional information:

```
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
/fs1/file1.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'exclude \
  *.save files'  EXCLUDE
/fs1/file2.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-03@21:19:57 16 root] RULE 'exclude \
  *.save files' EXCLUDE
/fs1/file.tmp1  [2009-03-04@02:09:31 0 0 0 sp1 2009-03-04@02:09:31 0 root] RULE 'delete' DELETE \
  FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp1  [2009-03-04@02:09:31 0 0 0 sp1 2009-03-04@02:09:31 0 root] RULE 'all' LIST \
  'tmpfiles' WEIGHT(INF)
/fs1/file.tmp0  [2009-03-04@02:09:38 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'delete' \
  DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp0  [2009-03-04@02:09:38 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'all' \
  LIST 'tmpfiles' WEIGHT(INF)
/fs1/file1      [2009-03-03@21:32:41 0 0 16384 sp1 2009-03-03@21:32:41 16 root] RULE 'migration \
  to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
/fs1/file0      [2009-03-03@21:21:11 0 0 16384 sp1 2009-03-03@21:32:41 16 root] RULE 'migration \
  to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
```

where the lines:

```
/fs1/file1.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'exclude \
  *.save files' EXCLUDE
/fs1/file2.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-03@21:19:57 16 root] RULE 'exclude \
  *.save files' EXCLUDE
```

show the attributes of excluded files **/fs1/file1.save** and **/fs1/file2.save**.

mmapplypolicy -L 6:

Use this option to display all of the information from the previous levels, plus files that are not candidate files, and their attributes.

These attributes include:

- **MODIFICATION_TIME**
- **USER_ID**
- **GROUP_ID**
- **FILE_SIZE**
- **POOL_NAME**
- **ACCESS_TIME**
- **KB_ALLOCATED**
- **FILESET_NAME**

This command:

```
mmapplypolicy fs1 -P policyfile -I test -L 6
```

produces the following additional information:

```
[I] Directories scan: 10 files, 1 directories, 0 other objects, 0 'skipped' files and/or errors.
/fs1/.  [2009-03-04@02:10:43 0 0 8192 system 2009-03-04@02:17:43 8 root] NO RULE APPLIES
/fs1/file1.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE \
  'exclude *.save files' EXCLUDE
/fs1/file2.save [2009-03-03@21:19:57 0 0 16384 sp1 2009-03-03@21:19:57 16 root] RULE \
  'exclude *.save files' EXCLUDE
/fs1/file.tmp1  [2009-03-04@02:09:31 0 0 0 sp1 2009-03-04@02:09:31 0 root] RULE 'delete' \
  DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp1  [2009-03-04@02:09:31 0 0 0 sp1 2009-03-04@02:09:31 0 root] RULE 'all' LIST \
  'tmpfiles' WEIGHT(INF)
/fs1/data1      [2009-03-03@21:20:23 0 0 0 sp1 2009-03-04@02:09:31 0 root] NO RULE APPLIES
/fs1/file.tmp0  [2009-03-04@02:09:38 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'delete' \
  DELETE FROM POOL 'sp1' WEIGHT(INF)
/fs1/file.tmp0  [2009-03-04@02:09:38 0 0 16384 sp1 2009-03-04@02:09:38 16 root] RULE 'all' LIST \
  'tmpfiles' WEIGHT(INF)
/fs1/file1      [2009-03-03@21:32:41 0 0 16384 sp1 2009-03-03@21:32:41 16 root] RULE 'migration \
  to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
/fs1/file0      [2009-03-03@21:21:11 0 0 16384 sp1 2009-03-03@21:32:41 16 root] RULE 'migration \
  to system pool' MIGRATE FROM POOL 'sp1' TO POOL 'system' WEIGHT(INF)
```

where the line:

```
/fs1/data1      [2009-03-03@21:20:23 0 0 0 sp1 2009-03-04@02:09:31 0 root] NO RULE APPLIES
```

contains information about the **data1** file, which is not a candidate file.

The mmcheckquota command

The **mmcheckquota** command counts inode and space usage for a file system and writes the collected data into quota files.

Run the **mmcheckquota** command if:

- You have MMFS_QUOTA error log entries. This error log entry is created when the quota manager has a problem reading or writing the quota file.
- Quota information is lost due to node failure. Node failure could leave you unable to open files or it could deny you disk space that their quotas allow.
- The in-doubt value is approaching the quota limit. The sum of the in-doubt value and the current usage cannot exceed the hard limit. Therefore, the actual block space and number of files available to you might be constrained by the in-doubt value. If the in-doubt value approaches a significant percentage of the maximum quota amount, use the **mmcheckquota** command to account for the lost space and files.
- Any user, group, or fileset quota files are corrupted.

During the normal operation of file systems with quotas enabled (not running **mmcheckquota** online), the usage data reflects the actual usage of the blocks and inodes, which means that if you delete files you the usage amount decreases. The in-doubt value does not reflect this usage amount. Instead, it is the number of quotas that the quota server assigns to its clients. The quota server does not know whether the assigned amount is used or not.

The only situation in which the in-doubt value is important is when the sum of the usage and the in-doubt value is greater than the quota hard limit. In this case, you cannot allocate more blocks or inodes unless you reduce the usage amount.

Note: The **mmcheckquota** command is I/O sensitive and if you specify it your system workload might increase substantially. Specify it only when your workload is light.

The **mmcheckquota** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Administration Guide*.

The mmlsnsd command

The **mmlsnsd** command displays information about the currently defined disks in the cluster.

For example, if you issue **mmlsnsd**, your output is similar to this:

```
 File system   Disk name    NSD servers
------------------------------------------------------------------------
 fs2           hd3n97       c5n97g.ppd.pok.ibm.com,c5n98g.ppd.pok.ibm.com,c5n99g.ppd.pok.ibm.com
 fs2           hd4n97       c5n97g.ppd.pok.ibm.com,c5n98g.ppd.pok.ibm.com,c5n99g.ppd.pok.ibm.com
 fs2           hd5n98       c5n98g.ppd.pok.ibm.com,c5n97g.ppd.pok.ibm.com,c5n99g.ppd.pok.ibm.com
 fs2           hd6n98       c5n98g.ppd.pok.ibm.com,c5n97g.ppd.pok.ibm.com,c5n99g.ppd.pok.ibm.com
 fs2           sdbnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sdcnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sddnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sdensd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sdgnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sdfnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 fs2           sdhnsd       c5n94g.ppd.pok.ibm.com,c5n96g.ppd.pok.ibm.com
 (free disk)   hd2n97       c5n97g.ppd.pok.ibm.com,c5n98g.ppd.pok.ibm.com
```

To find out the local device names for these disks, use the **mmlsnsd** command with the **-m** option. For example, issuing **mmlsnsd -m** produces output similar to this:

```
 Disk name    NSD volume ID      Device         Node name                Remarks
---------------------------------------------------------------------------------------
 hd2n97       0972846145C8E924   /dev/hdisk2    c5n97g.ppd.pok.ibm.com   server node
 hd2n97       0972846145C8E924   /dev/hdisk2    c5n98g.ppd.pok.ibm.com   server node
 hd3n97       0972846145C8E927   /dev/hdisk3    c5n97g.ppd.pok.ibm.com   server node
 hd3n97       0972846145C8E927   /dev/hdisk3    c5n98g.ppd.pok.ibm.com   server node
 hd4n97       0972846145C8E92A   /dev/hdisk4    c5n97g.ppd.pok.ibm.com   server node
 hd4n97       0972846145C8E92A   /dev/hdisk4    c5n98g.ppd.pok.ibm.com   server node
 hd5n98       0972846245EB501C   /dev/hdisk5    c5n97g.ppd.pok.ibm.com   server node
 hd5n98       0972846245EB501C   /dev/hdisk5    c5n98g.ppd.pok.ibm.com   server node
```

```
 hd6n98       0972846245DB3AD8   /dev/hdisk6    c5n97g.ppd.pok.ibm.com   server node
 hd6n98       0972846245DB3AD8   /dev/hdisk6    c5n98g.ppd.pok.ibm.com   server node
 hd7n97       0972846145C8E934   /dev/hd7n97    c5n97g.ppd.pok.ibm.com   server node
```

To obtain extended information for NSDs, use the **mmlsnsd** command with the **-X** option. For example, issuing **mmlsnsd -X** produces output similar to this:

```
 Disk name    NSD volume ID      Device         Devtype  Node name                Remarks
---------------------------------------------------------------------------------------------
 hd3n97       0972846145C8E927   /dev/hdisk3    hdisk    c5n97g.ppd.pok.ibm.com   server node,pr=no
 hd3n97       0972846145C8E927   /dev/hdisk3    hdisk    c5n98g.ppd.pok.ibm.com   server node,pr=no
 hd5n98       0972846245EB501C   /dev/hdisk5    hdisk    c5n97g.ppd.pok.ibm.com   server node,pr=no
 hd5n98       0972846245EB501C   /dev/hdisk5    hdisk    c5n98g.ppd.pok.ibm.com   server node,pr=no
 sdfnsd       0972845E45F02E81   /dev/sdf       generic  c5n94g.ppd.pok.ibm.com   server node
 sdfnsd       0972845E45F02E81   /dev/sdm       generic  c5n96g.ppd.pok.ibm.com   server node
```

The **mmlsnsd** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Administration Guide*.

The mmwindisk command

On Windows nodes, use the **mmwindisk** command to view all disks known to the operating system along with partitioning information relevant to GPFS.

For example, if you issue **mmwindisk list**, your output is similar to this:

```
Disk Avail Type    Status    Size     GPFS Partition ID
---- ----- ------- --------- -------- ------------------------------------
   0       BASIC   ONLINE     137 GiB
   1       GPFS    ONLINE      55 GiB 362DD84E-3D2E-4A59-B96B-BDE64E31ACCF
   2       GPFS    ONLINE     200 GiB BD5E64E4-32C8-44CE-8687-B14982848AD2
   3       GPFS    ONLINE      55 GiB B3EC846C-9C41-4EFD-940D-1AFA6E2D08FB
   4       GPFS    ONLINE      55 GiB 6023455C-353D-40D1-BCEB-FF8E73BF6C0F
   5       GPFS    ONLINE      55 GiB 2886391A-BB2D-4BDF-BE59-F33860441262
   6       GPFS    ONLINE      55 GiB 00845DCC-058B-4DEB-BD0A-17BAD5A54530
   7       GPFS    ONLINE      55 GiB 260BCAEB-6E8A-4504-874D-7E07E02E1817
   8       GPFS    ONLINE      55 GiB 863B6D80-2E15-457E-B2D5-FEA0BC41A5AC
   9 YES   UNALLOC OFFLINE     55 GiB
  10 YES   UNALLOC OFFLINE    200 GiB
```

Where:

Disk
: is the Windows disk number as shown in the Disk Management console and the DISKPART command-line utility.

Avail
: shows the value **YES** when the disk is available and in a state suitable for creating an NSD.

GPFS Partition ID
: is the unique ID for the GPFS partition on the disk.

The **mmwindisk** command does not provide the NSD volume ID. You can use **mmlsnsd -m** to find the relationship between NSDs and devices, which are disk numbers on Windows.

The mmfileid command

The **mmfileid** command identifies files that are on areas of a disk that are damaged or suspect.

Attention: Use this command only when the IBM Support Center directs you to do so.

Before you run **mmfileid**, you must run a disk analysis utility and obtain the disk sector numbers that are damaged or suspect. These sectors are input to the **mmfileid** command.

The command syntax is as follows:

```
mmfileid Device
 {-d DiskDesc | -F DescFile}
 [-o OutputFile] [-f NumThreads] [-t Directory]
 [-N {Node[,Node...] | NodeFile | NodeClass}] [--qos QOSClass]
```

The input parameters are as follows:

Device
The device name for the file system.

-d *DiskDesc*
A descriptor that identifies the disk to be scanned. *DiskDesc* has the following format:

```
NodeName:DiskName[:PhysAddr1[-PhysAddr2]]
```

It has the following alternative format:

```
:{NsdName|DiskNum|BROKEN}[:PhysAddr1[-PhysAddr2]]
```

NodeName
Specifies a node in the GPFS cluster that has access to the disk to scan. You must specify this value if the disk is identified with its physical volume name. Do not specify this value if the disk is identified with its NSD name or its GPFS disk ID number, or if the keyword **BROKEN** is used.

DiskName
Specifies the physical volume name of the disk to scan as known on node *NodeName*.

NsdName
Specifies the GPFS NSD name of the disk to scan.

DiskNum
Specifies the GPFS disk ID number of the disk to scan as displayed by the **mmlsdisk -L** command.

BROKEN
Tells the command to scan all the disks in the file system for files with broken addresses that result in lost data.

*PhysAddr1***[-***PhysAddr2***]**
Specifies the range of physical disk addresses to scan. The default value for *PhysAddr1* is zero. The default value for *PhysAddr2* is the value for *PhysAddr1*.

If both *PhysAddr1* and *PhysAddr2* are zero, the command searches the entire disk.

The following lines are examples of valid disk descriptors:

```
k148n07:hdisk9:2206310-2206810
:gpfs1008nsd:
:10:27645856
:BROKEN
```

-F *DescFile*
Specifies a file that contains a list of disk descriptors, one per line.

-f *NumThreads*
Specifies the number of worker threads to create. The default value is 16. The minimum value is 1. The maximum value is the maximum number allowed by the operating system function **pthread_create** for a single process. A suggested value is twice the number of disks in the file system.

-N {*Node***[,***Node***...] |** *NodeFile* **|** *NodeClass***}**
Specifies the list of nodes that participate in determining the disk addresses. This command supports all defined node classes. The default is **all** or the current value of the **defaultHelperNodes** configuration parameter of the **mmchconfig** command.

For general information on how to specify node names, see *Specifying nodes as input to GPFS commands* in the *IBM Spectrum Scale: Administration Guide*.

-o *OutputFile*
The path name of a file to which the result from the **mmfileid** command is to be written. If not specified, the result is sent to standard output.

-t *Directory*
Specifies the directory to use for temporary storage during **mmfileid** command processing. The default directory is **/tmp**.

--qos *QOSClass*
Specifies the Quality of Service for I/O operations (QoS) class to which the instance of the command is assigned. If you do not specify this parameter, the instance of the command is assigned by default to the **maintenance** QoS class. This parameter has no effect unless the QoS service is enabled. For more information, see the help topic on the **mmchqos** command in the *IBM Spectrum Scale: Command and Programming Reference*. Specify one of the following QoS classes:

maintenance
This QoS class is typically configured to have a smaller share of file system IOPS. Use this class for I/O-intensive, potentially long-running GPFS commands, so that they contribute less to reducing overall file system performance.

other This QoS class is typically configured to have a larger share of file system IOPS. Use this class for administration commands that are not I/O-intensive.

For more information, see the help topic on *Setting the Quality of Service for I/O operations (QoS)* in the *IBM Spectrum Scale: Administration Guide*.

You can redirect the output to a file with the **-o** flag and sort the output on the inode number with the **sort** command.

The **mmfileid** command output contains one line for each inode found to be on a corrupted disk sector. Each line of the command output has this format:

InodeNumber LogicalDiskAddress SnapshotId Filename

InodeNumber
Indicates the inode number of the file identified by **mmfileid**.

LogicalDiskAddress
Indicates the disk block (disk sector) number of the file identified by **mmfileid**.

SnapshotId
Indicates the snapshot identifier for the file. A *SnapshotId* of 0 means that the file is not a snapshot file.

Filename
Indicates the name of the file identified by **mmfileid**. File names are relative to the root of the file system in which they reside.

Assume that a disk analysis tool reports that disks `hdisk6, hdisk7, hdisk8,` and `hdisk9` contain bad sectors, and that the file `addr.in` has the following contents:

```
k148n07:hdisk9:2206310-2206810
k148n07:hdisk8:2211038-2211042
k148n07:hdisk8:2201800-2202800
k148n01:hdisk6:2921879-2926880
k148n09:hdisk7:1076208-1076610
```

You run the following command:

```
mmfileid /dev/gpfsB -F addr.in
```

The command output might be similar to the following example:

```
Address 2201958 is contained in the Block allocation map (inode 1)
Address 2206688 is contained in the ACL Data file (inode 4, snapId 0)
Address 2211038 is contained in the Log File (inode 7, snapId 0)
14336 1076256 0 /gpfsB/tesDir/testFile.out
14344 2922528 1 /gpfsB/x.img
```

The lines that begin with the word `Address` represent GPFS system metadata files or reserved disk areas. If your output contains any lines like these, do not attempt to replace or repair the indicated files. If you suspect that any of the special files are damaged, call the IBM Support Center for assistance.

The following line of output indicates that inode number 14336, disk address 1072256 contains file **/gpfsB/tesDir/testFile.out**. The 0 to the left of the name indicates that the file does not belong to a snapshot. This file is on a potentially bad disk sector area:

```
14336 1072256 0 /gpfsB/tesDir/testFile.out
```

The following line of output indicates that inode number 14344, disk address 2922528 contains file **/gpfsB/x.img**. The 1 to the left of the name indicates that the file belongs to snapshot number 1. This file is on a potentially bad disk sector area:

```
14344 2922528 1 /gpfsB/x.img
```

The SHA digest

The Secure Hash Algorithm (SHA) digest is relevant only when using GPFS in a multi-cluster environment.

The SHA digest is a short and convenient way to identify a key registered with either the **mmauth show** or **mmremotecluster** command. In theory, two keys may have the same SHA digest. In practice, this is extremely unlikely. The SHA digest can be used by the administrators of two GPFS clusters to determine if they each have received (and registered) the right key file from the other administrator.

An example is the situation of two administrators named **Admin1** and **Admin2** who have registered the others' respective key file, but find that mount attempts by **Admin1** for file systems owned by **Admin2** fail with the error message: `Authorization failed`. To determine which administrator has registered the wrong key, they each run **mmauth show** and send the local clusters SHA digest to the other administrator. **Admin1** then runs the **mmremotecluster** command and verifies that the SHA digest for **Admin2**'s cluster matches the SHA digest for the key that **Admin1** has registered. **Admin2** then runs the **mmauth show** command and verifies that the SHA digest for **Admin1**'s cluster matches the key that **Admin2** has authorized.

If **Admin1** finds that the SHA digests do not match, **Admin1** runs the **mmremotecluster update** command, passing the correct key file as input.

If **Admin2** finds that the SHA digests do not match, **Admin2** runs the **mmauth update** command, passing the correct key file as input.

This is an example of the output produced by the **mmauth show all** command:

```
Cluster name: fksdcm.pok.ibm.com
Cipher list: EXP1024-RC2-CBC-MD5
SHA digest: d5eb5241eda7d3ec345ece906bfcef0b6cd343bd
File system access: fs1 (rw, root allowed)


Cluster name: kremote.cluster
Cipher list: EXP1024-RC4-SHA
SHA digest: eb71a3aaa89c3979841b363fd6d0a36a2a460a8b
File system access: fs1 (rw, root allowed)
```

```
Cluster name: dkq.cluster (this cluster)
Cipher list: AUTHONLY
SHA digest: 090cd57a2e3b18ac163e5e9bd5f26ffabaa6aa25
File system access: (all rw)
```

Collecting details of the issues from performance monitoring tools

This topic describes how to collect details of issues that you might encounter in IBM Spectrum Scale by using performance monitoring tools.

With IBM Spectrum Scale, system administrators can monitor the performance of GPFS and the communications protocols that it uses. Issue the **mmperfmon query** command to query performance data.

Note: If you issue the **mmperfmon query** command without any additional parameters, you can see a list of options for querying performance-related information, as shown in the following sample output:

```
 Usage:
mmperfmon query Metric[,Metric...] | Key[,Key...] | NamedQuery [StartTime EndTime | Duration] [Options]
OR
mmperfmon query compareNodes ComparisonMetric [StartTime EndTime | Duration] [Options]
where
    Metric                  metric name
    Key                     a key consisting of node name, sensor group, optional additional filters,
                            metric name, separated by pipe symbol
                            e.g.: "cluster1.ibm.com|CTDBStats|locking|db_hop_count_bucket_00"
    NamedQuery              name of a pre-defined query
    ComparisonMetric        name of a metric to be compared if using CompareNodes
    StartTime               Start timestamp for query
                            Format: YYYY-MM-DD-hh:mm:ss
    EndTime                 End timestamp for query. Omitted means: execution time
                            Format: YYYY-MM-DD-hh:mm:ss
    Duration                Number of seconds into the past from today or <EndTime>


Options:
  -h, --help            show this help message and exit
  -N NodeName, --Node=NodeName
                        Defines the node that metrics should be retrieved from
  -b BucketSize, --bucket-size=BucketSize
                        Defines a bucket size (number of seconds), default is
                        1
  -n NumberBuckets, --number-buckets=NumberBuckets
                        Number of buckets ( records ) to show, default is 10
  --filter=Filter       Filter criteria for the query to run
  --format=Format       Common format for all columns
  --csv                 Provides output in csv format.
  --raw                 Provides output in raw format rather than a pretty
                        table format.
  --nice                Use colors and other text attributes for output.
  --resolve             Resolve computed metrics, show metrics used
  --short               Shorten column names if there are too many to fit into
                        one row.
  --list=List           Show list of specified values (overrides other
                        options). Values are all, metrics, computed, queries,
                        keys.
Possible named queries are:
            compareNodes - Compares a single metric across all nodes running sensors
                     cpu - Show CPU utilization in system and user space, and context switches
         ctdbCallLatency - Show CTDB call latency.
     ctdbHopCountDetails - Show CTDB hop count buckets 0 to 5 for one database.
           ctdbHopCounts - Show CTDB hop counts (bucket 00 = 1-3 hops) for all databases.
      gpfsCRUDopsLatency - Show GPFS CRUD operations latency
             gpfsFSWaits - Display max waits for read and write operations for all file systems
            gpfsNSDWaits - Display max waits for read and write operations for all disks
    gpfsNumberOperations - Get the number of operations to the GPFS file system.
```

```
   gpfsVFSOpCounts - Display VFS operation counts
        netDetails - Get details about the network.
         netErrors - Show network problems for all available networks: collisions, drops, errors
         nfsErrors - Get the NFS error count for read and write operations
      nfsIOLatency - Get the NFS IO Latency in nanoseconds per second
         nfsIORate - Get the NFS IOps per second
          nfsQueue - Get the NFS read and write queue size in bytes
    nfsThroughput - Get the NFS Throughput in bytes per second
nfsThroughputPerOp - Get the NFS read and write throughput per op in bytes
           objAcc - Object account overall performance.
         objAccIO - Object account IO details.
    objAccLatency - Object proxy Latency.
  objAccThroughput - Object account overall Throughput.
           objCon - Object container overall performance.
         objConIO - Object container IO details.
    objConLatency - Object container Latency.
  objConThroughput - Object container overall Throughput.
           objObj - Object overall performance.
         objObjIO - Object overall IO details.
    objObjLatency - Object Latency.
  objObjThroughput - Object overall Throughput.
           objPro - Object proxy overall performance.
         objProIO - Object proxy IO details.
  objProThroughput - Object proxy overall Throughput.
 protocolIOLatency - Compare latency per protocol (smb, nfs, object).
    protocolIORate - Get the percentage of total I/O rate per protocol (smb, nfs, object).
protocolThroughput - Get the percentage of total throughput per protocol (smb, nfs, object).
     smb2IOLatency - Get the SMB2 I/O latencies per bucket size ( default 1 sec )
        smb2IORate - Get the SMB2 I/O rate in number of operations per bucket size ( default 1 sec )
    smb2Throughput - Get the SMB2 Throughput in bytes per bucket size ( default 1 sec )
        smb2Writes - Count, # of idle calls, bytes in and out and operation time for smb2 writes
    smbConnections - Number of smb connections
             usage - Show CPU, memory, storage and network usage
```

For more information on monitoring performance and analyzing performance related issues, see "Performance monitoring tool overview" on page 44 and **mmperfmon** *command* in the *IBM Spectrum Scale: Command and Programming Reference*

Other problem determination tools

Other problem determination tools include the kernel debugging facilities and the **mmpmon** command.

If your problem occurs on the AIX operating system, see AIX in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/ssw_aix/welcome) and search for the appropriate kernel debugging documentation for information about the AIX **kdb** command.

If your problem occurs on the Linux operating system, see the documentation for your distribution vendor.

If your problem occurs on the Windows operating system, the following tools that are available from the Windows Sysinternals, might be useful in troubleshooting:

- Debugging Tools for Windows
- Process Monitor
- Process Explorer
- Microsoft Windows Driver Kit
- Microsoft Windows Software Development Kit

The **mmpmon** command is intended for system administrators to analyze their I/O on the node on which it is run. It is not primarily a diagnostic tool, but may be used as one for certain problems. For example, running **mmpmon** on several nodes may be used to detect nodes that are experiencing poor performance or connectivity problems.

The syntax of the **mmpmon** command is fully described in the *Command reference* section in the *IBM Spectrum Scale: Command and Programming Reference*. For details on the **mmpmon** command, see "Monitoring GPFS I/O performance with the mmpmon command" on page 4.

Chapter 14. Managing deadlocks

IBM Spectrum Scale provides functions for automatically detecting potential deadlocks, collecting deadlock debug data, and breaking up deadlocks.

The distributed nature of GPFS, the complexity of the locking infrastructure, the dependency on the proper operation of disks and networks, and the overall complexity of operating in a clustered environment all contribute to increasing the probability of a deadlock.

Deadlocks can be disruptive in certain situations, more so than other type of failure. A deadlock effectively represents a single point of failure that can render the entire cluster inoperable. When a deadlock is encountered on a production system, it can take a long time to debug. The typical approach to recovering from a deadlock involves rebooting all of the nodes in the cluster. Thus, deadlocks can lead to prolonged and complete outages of clusters.

To troubleshoot deadlocks, you must have specific types of debug data that must be collected while the deadlock is in progress. Data collection commands must be run manually before the deadlock is broken. Otherwise, determining the root cause of the deadlock after that is difficult. Also, deadlock detection requires some form of external action, for example, a complaint from a user. Waiting for a user complaint means that detecting a deadlock in progress might take many hours.

In GPFS V4.1 and later, automated deadlock detection, automated deadlock data collection, and deadlock breakup options are provided to make it easier to handle a deadlock situation.

Debug data for deadlocks

Debug data for potential deadlocks is automatically collected. System administrators must monitor and manage the file systems where debug data is stored.

Automated deadlock detection and automated deadlock data collection are enabled by default. Automated deadlock breakup is disabled by default.

At the start of the GPFS daemon, the `mmfs.log` file shows entries like the following:

```
Thu Jul 16 18:50:14.097 2015: [I] Enabled automated deadlock detection.
Thu Jul 16 18:50:14.098 2015: [I] Enabled automated deadlock debug data
collection.
Thu Jul 16 18:50:14.099 2015: [I] Enabled automated expel debug data collection.
Thu Jul 16 18:50:14.100 2015: [I] Please see https://ibm.biz/Bd4bNK for more
information on deadlock amelioration.
```

The short URL points to this help topic to make it easier to find the information later.

By default, debug data is put into the `/tmp/mmfs` directory, or the directory specified for the **dataStructureDump** configuration parameter, on each node. Plenty of disk space, typically many GBs, needs to be available. Debug data is not collected when the directory runs out of disk space.

Important: Before you change the value of **dataStructureDump**, stop the GPFS trace. Otherwise you will lose GPFS trace data. Restart the GPFS trace afterwards.

After a potential deadlock is detected and the relevant debug data is collected, IBM Service needs to be contacted to report the problem and to upload the debug data. Outdated debug data needs to be removed to make room for new debug data in case a new potential deadlock is detected.

It is the responsibility of system administrators to manage the disk space under the `/tmp/mmfs` directory or **dataStructureDump**. They know which set of debug data is still useful.

The "expel debug data" is similar to the "deadlock debug data", but it is collected when a node is expelled from a cluster for no apparent reasons.

Automated deadlock detection

Automated deadlock detection flags unexpected long waiters as potential deadlocks. Effective deadlock detection thresholds are self-tuned to reduce false positive detection. You can register a user program for the **deadlockDetected** event to receive automatic notification.

GPFS code uses waiters to track what a thread is waiting for and how long it is waiting. Many deadlocks involve long waiters. In a real deadlock, long waiters do not disappear naturally as the deadlock prevents the threads from getting what they are waiting for. With some exceptions, long waiters typically indicate that something in the system is not healthy. A deadlock might be in progress, some disk might be failing, or the entire system might be overloaded.

Automated deadlock detection monitors waiters to detect potential deadlocks. Some waiters can become long legitimately under normal operating conditions and such waiters are ignored by automated deadlock detection. Such waiters appear in the **mmdiag --waiters** output but never in the **mmdiag --deadlock** output. From now on in this topic, the word *waiters* refers only to those waiters that are monitored by automated deadlock detection.

Automated deadlock detection flags a waiter as a potential deadlock when the waiter length exceeds certain threshold for deadlock detection. For example, the following `mmfs.log` entry indicates that a waiter started on thread 8397 at 2015-07-18 09:36:58 passed 905 seconds at Jul 18 09:52:04.626 2015 and is suspected to be a deadlock waiter.

```
Sat Jul 18 09:52:04.626 2015: [A] Unexpected long waiter detected: Waiting 905.9380 sec since
2015-07-18 09:36:58, on node c33f2in01,
SharedHashTabFetchHandlerThread 8397: on MsgRecordCondvar,
reason 'RPC wait' for tmMsgTellAcquire1
```

The `/var/log/messages` file on Linux and the error log on AIX also log an entry for the deadlock detection, but the `mmfs.log` file has most details.

The **deadlockDetected** event is triggered on "Unexpected long waiter detected" and any user program that is registered for the event is invoked. The user program can be made for recording and notification purposes. See `/usr/lpp/mmfs/samples/deadlockdetected.sample` for an example and more information.

When the flagged waiter disappears, an entry like the following one might appear in the `mmfs.log` file:

```
Sat Jul 18 10:00:05.705 2015: [N] The unexpected long waiter on thread 8397 has disappeared in 1386 seconds.
```

The **mmdiag --deadlock** command shows the flagged waiter and possibly other waiters closely behind which also passed the threshold for deadlock detection

If the flagged waiter disappears on its own, without any deadlock breakup actions, then the flagged waiter is not a real deadlock, and the detection is a false positive. A reasonable threshold needs to be

established to reduce false positive deadlock detection. It is a good practice to consider the trade-off between waiting too long and not having a timely detection and not waiting long enough causing a false-positive detection.

A false positive deadlock detection and debug data collection are not necessarily a waste of resources. A long waiter, even if it eventually disappears on its own, likely indicates that something is not working well, and is worth looking into.

The configuration parameter **deadlockDetectionThreshold** is used to specify the initial threshold for deadlock detection. GPFS code adjusts the threshold on each node based on what's happening on the node and cluster. The adjusted threshold is the effective threshold used in automated deadlock detection.

An internal algorithm is used to evaluate whether a cluster is overloaded or not. Overload is a factor that influences the adjustment of the effective deadlock detection threshold. The effective deadlock detection threshold and the cluster overload index are shown in the output of the `mmdiag --deadlock`.

```
Effective deadlock detection threshold on c37f2n04 is 1000 seconds
Effective deadlock detection threshold on c37f2n04 is 430 seconds for short waiters
Cluster my.cluster is overloaded. The overload index on c40bbc2xn2 is 1.14547
```

If **deadlockDetectionThresholdForShortWaiters** is positive, and it is by default, certain waiters, including most of the mutex waiters, are considered short waiters that should not be long. These short waiters have a shorter effective deadlock detection threshold that is self-tuned separately.

Certain waiters, including most of the mutex waiters, are considered short waiters that should not be long. If **deadlockDetectionThresholdForShortWaiters** is positive, and it is by default, these short waiters are monitored separately. Their effective deadlock detection threshold is also self-tuned separately.

The overload index is the weighted average duration of all I/Os completed over a long time. Recent I/O durations count more than the ones in the past. The cluster overload detection affects deadlock amelioration functions only. The determination by GPFS that a cluster is overloaded is not necessarily the same as the determination by a customer. But customers might use the determination by GPFS as a reference and check the workload, hardware and network of the cluster to see whether anything needs correction or adjustment. An overloaded cluster with a workload far exceeding its resource capability is not healthy nor productive.

If the existing effective deadlock detection threshold value is no longer appropriate for the workload, run the **mmfsadm resetstats** command to restart the local adjustment.

To view the current value of **deadlockDetectionThreshold** and **deadlockDetectionThresholdForShortWaiters**, which are the initial thresholds for deadlock detection, enter the following command:

```
mmlsconfig deadlockDetectionThreshold
mmlsconfig deadlockDetectionThresholdForShortWaiters
```

The system displays output similar to the following:

```
deadlockDetectionThreshold 300
deadlockDetectionThresholdForShortWaiters 60
```

To disable automated deadlock detection, specify a value of 0 for **deadlockDetectionThreshold**. All deadlock amelioration functions, not just deadlock detection, are disabled by specifying 0 for **deadlockDetectionThreshold**. A positive value must be specified for **deadlockDetectionThreshold** to enable any part of the deadlock amelioration functions.

Automated deadlock data collection

Automated deadlock data collection gathers crucial debug data when a potential deadlock is detected.

Automated deadlock data collection helps gather crucial debug data on detection of a potential deadlock. Messages similar to the following ones are written to the `mmfs.log` file:

```
Sat Jul 18 09:52:04.626 2015: [A] Unexpected long waiter detected:
2015-07-18 09:36:58: waiting 905.938 seconds on node c33f2in01:
SharedHashTabFetchHandlerThread 8397: on MsgRecordCondvar,
reason 'RPC wait' for tmMsgTellAcquire1
Sat Jul 18 09:52:04.627 2015: [I] Initiate debug data collection from
this node.
Sat Jul 18 09:52:04.628 2015: [I] Calling User Exit Script
gpfsDebugDataCollection: event deadlockDebugData,
Async command /usr/lpp/mmfs/bin/mmcommon.
```

What debug data is collected depends on the value of the configuration parameter **debugDataControl**. The default value is **light** and a minimum amount of debug data, the data that is most frequently needed to debug a GPFS issue, is collected. The value **medium** gets more debug data collected. The value **heavy** is meant to be used routinely by internal test teams only. The value **verbose** needed only for troubleshooting special cases and can result in very large dumps. No debug data is collected when the value **none** is specified. You can set different values for the **debugDataControl** parameter across nodes in the cluster. For more information, see the topic *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Automated deadlock data collection is enabled by default and controlled by the configuration parameter **deadlockDataCollectionDailyLimit**. This parameter specifies the maximum number of times debug data can be collected in a 24-hour period by automated deadlock data collection

To view the current value of **deadlockDataCollectionDailyLimit**, enter the following command:

```
mmlsconfig deadlockDataCollectionDailyLimit
```

The system displays output similar to the following:

```
deadlockDataCollectionDailyLimit 3
```

To disable automated deadlock data collection, specify a value of 0 for **deadlockDataCollectionDailyLimit**.

Another configuration parameter, **deadlockDataCollectionMinInterval**, is used to control the minimum amount of time between consecutive debug data collections. The default is 3600 seconds or 1 hour.

Automated deadlock breakup

Automated deadlock breakup helps resolve a deadlock situation without human intervention. To break up a deadlock, less disruptive actions are tried first; for example, causing a file system panic. If necessary, more disruptive actions are then taken; for example, shutting down a GPFS **mmfsd** daemon.

If a system administrator prefers to control the deadlock breakup process, the **deadlockDetected** callback can be used to notify system administrators that a potential deadlock was detected. The information from the **mmdiag --deadlock** section can then be used to help determine what steps to take to resolve the deadlock.

Automated deadlock breakup is disabled by default and controlled with the **mmchconfig** attribute **deadlockBreakupDelay**. The **deadlockBreakupDelay** attribute specifies how long to wait after a deadlock is detected before attempting to break up the deadlock. Enough time must be provided to allow the debug data collection to complete. To view the current breakup delay, enter the following command:

```
mmlsconfig deadlockBreakupDelay
```

The system displays output similar to the following:

```
deadlockBreakupDelay 0
```

The value of 0 shows that automated deadlock breakup is disabled. To enable automated deadlock breakup, specify a positive value for **deadlockBreakupDelay**. If automated deadlock breakup is to be enabled, a delay of 300 seconds or longer is recommended.

Automated deadlock breakup is done on a node-by-node basis. If automated deadlock breakup is enabled, the breakup process is started when the suspected deadlock waiter is detected on a node. The process first waits for the **deadlockBreakupDelay**, and then goes through various phases until the deadlock waiters disappear. There is no central coordination on the deadlock breakup, so the time to take deadlock breakup actions may be different on each node. Breaking up a deadlock waiter on one node can cause some deadlock waiters on other nodes to disappear, so no breakup actions need to be taken on those other nodes.

If a suspected deadlock waiter disappears while waiting for the **deadlockBreakupDelay**, the automated deadlock breakup process stops immediately without taking any further action. To lessen the number of breakup actions that are taken in response to detecting a false-positive deadlock, increase the **deadlockBreakupDelay**. If you decide to increase the **deadlockBreakupDelay**, a deadlock can potentially exist for a longer period.

If your goal is to break up a deadlock as soon as possible, and your workload can afford an interruption at any time, then enable automated deadlock breakup from the beginning. Otherwise, keep automated deadlock breakup disabled to avoid unexpected interruptions to your workload. In this case, you can choose to break the deadlock manually, or use the function that is described in the "Deadlock breakup on demand" topic.

Due to the complexity of the GPFS code, asserts or segmentation faults might happen during a deadlock breakup action. That might cause unwanted disruptions to a customer workload still running normally on the cluster. A good reason to use deadlock breakup on demand is to not disturb a partially working cluster until it is safe to do so. Try not to break up a suspected deadlock prematurely to avoid unnecessary disruptions. If automated deadlock breakup is enabled all of the time, it is good to set **deadlockBreakupDelay** to a large value such as 3600 seconds. If using **mmcommon breakDeadlock**, it is better to wait until the longest deadlock waiter is an hour or longer. Much shorter times can be used if a customer prefers fast action in breaking a deadlock over assurance that a deadlock is real.

The following messages, related to deadlock breakup, might be found in the **mmfs.log** files:

```
[I] Enabled automated deadlock breakup.

[N] Deadlock breakup: starting in 300 seconds

[N] Deadlock breakup: aborting RPC on 1 pending nodes.

[N] Deadlock breakup: panicking fs fs1

[N] Deadlock breakup: shutting down this node.

[N] Deadlock breakup: the process has ended.
```

Deadlock breakup on demand

Deadlocks can be broken up on demand, which allows a system administrator to choose the appropriate time to start the breakup actions.

A deadlock can be localized, for example, it might involve only one of many file systems in a cluster. The other file systems in the cluster can still be used, and a mission critical workload might need to continue uninterrupted. In these cases, the best time to break up the deadlock is after the mission critical workload ends.

The **mmcommon** command can be used to break up an existing deadlock in a cluster when the deadlock was previously detected by deadlock amelioration. To start the breakup on demand, use the following syntax:

```
mmcommon breakDeadlock [-N {Node[,Node...] | NodeFile | NodeClass}]
```

If the **mmcommon breakDeadlock** command is issued without the **-N** parameter, then every node in the cluster receives a request to take action on any long waiter that is a suspected deadlock.

If the **mmcommon breakDeadlock** command is issued with the **-N** parameter, then only the nodes that are specified receive a request to take action on any long waiter that is a suspected deadlock. For example, assume that there are two nodes, called **node3** and **node6**, that require a deadlock breakup. To send the breakup request to just these nodes, issue the following command:

```
mmcommon breakDeadlock -N node3,node6
```

Shortly after running the **mmcommon breakDeadlock** command, issue the following command:

```
mmdsh -N all /usr/lpp/mmfs/bin/mmdiag --deadlock
```

The output of the **mmdsh** command can be used to determine if any deadlock waiters still exist and if any additional actions are needed.

The effect of the **mmcommon breakDeadlock** command only persists on a node until the longest deadlock waiter that was detected disappears. All actions that are taken by **mmcommon breakDeadlock** are recorded in the **mmfs.log** file. When **mmcommon breakDeadlock** is issued for a node that did not have a deadlock, no action is taken except for recording the following message in the **mmfs.log** file:

```
[N] Received deadlock breakup request from 192.168.40.72: No deadlock to break up.
```

The **mmcommon breakDeadlock** command provides more control over breaking up deadlocks, but multiple breakup requests might be required to achieve satisfactory results. All waiters that exceeded the **deadlockDetectionThreshold** might not disappear when **mmcommon breakDeadlock** completes on a node. In complicated deadlock scenarios, some long waiters can persist after the longest waiters disappear. Waiter length can grow to exceed the **deadlockDetectionThreshold** at any point, and waiters can disappear at any point as well. Examine the waiter situation after **mmcommon breakDeadlock** completes to determine whether the command must be repeated to break up the deadlock.

Another way to break up a deadlock on demand is to enable automated deadlock breakup by changing **deadlockBreakupDelay** to a positive value. By enabling automated deadlock breakup, breakup actions are initiated on existing deadlock waiters. The breakup actions repeat automatically if deadlock waiters are detected. Change **deadlockBreakupDelay** back to 0 when the results are satisfactory, or when you want to control the timing of deadlock breakup actions again. If automated deadlock breakup remains enabled, breakup actions start on any newly detected deadlocks without any intervention.

Chapter 15. Installation and configuration issues

You might encounter errors with GPFS installation, configuration, and operation. Use the information in this topic to help you identify and correct errors.

An IBM Spectrum Scale installation problem should be suspected when GPFS modules are not loaded successfully, commands do not work, either on the node that you are working on or on other nodes, new command operands added with a new release of IBM Spectrum Scale are not recognized, or there are problems with the kernel extension.

A GPFS configuration problem should be suspected when the GPFS daemon will not activate, it will not remain active, or it fails on some nodes but not on others. Suspect a configuration problem also if quorum is lost, certain nodes appear to hang or do not communicate properly with GPFS, nodes cannot be added to the cluster or are expelled, or GPFS performance is very noticeably degraded once a new release of GPFS is installed or configuration parameters have been changed.

These are some of the errors encountered with GPFS installation, configuration and operation:

• "Windows issues" on page 310

Resolving most frequent problems related to installation, deployment, and upgrade

Use the following information to resolve the most frequent problems related to installation, deployment, and upgrade.

Finding deployment related error messages more easily and using them for failure analysis

Use this information to find and analyze error messages related to installation, deployment, and upgrade from the respective logs when using the installation toolkit.

In case of any installation, deployment, and upgrade related error:

1. Go to the end of the corresponding log file and search upwards for the text `FATAL`.
2. Find the topmost occurrence of `FATAL` (or first FATAL error that occurred) and look above and below this error for further indications of the failure.

Error messages at the bottom of the installation, deployment, and upgrade related logs are specific to the Chef component which controls the entire activity and therefore they are not typically the first place to look during failure analysis. For more information, see the following examples:

- "Example 1 - Installation failed and the bottom of the log file contains the following Chef output, which is not indicative of the error"
- "Example 2 - Deployment failed and the bottom of the log file contains the following Chef output, which is not indicative of the error" on page 281

Example 1 - Installation failed and the bottom of the log file contains the following Chef output, which is not indicative of the error

```
2016-01-28 09:29:21,839 [ TRACE ] Stopping chef zero
2016-01-28 09:29:21,839 [ ERROR ] The following error was encountered:
Traceback (most recent call last):
 File "/usr/lpp/mmfs/4.2.0.1/installer/espylib/reporting.py", line 193, in log_to_file
   yield handler
 File "/usr/lpp/mmfs/4.2.0.1/installer/espylib/install.py", line 152, in _install
   setup.install(config)
 File "/usr/lpp/mmfs/4.2.0.1/installer/espylib/setup/gpfs.py", line 481, in install
   self.deploy(config.admin_nodes[0], recipe, attributes)
 File "/usr/lpp/mmfs/4.2.0.1/installer/espylib/connectionmanager.py", line 52, in deploy
   ssh_identity=self._get_ssh_identity()
 File "/usr/lpp/mmfs/4.2.0.1/installer/espylib/deploy.py", line 108, in deploy_nodes
   raise DeployError()
DeployError: Installation failed on one or more nodes. Check the log for more details.
2016-01-28 09:29:21,927 [ INFO  ] Detailed error log:
/usr/lpp/mmfs/4.2.0.1/installer/logs/INSTALL-28-01-2016_09:05:58.log
```

1. To find more details, go to the end of the log file and search upwards for the text `FATAL`.

   In this example, the first search hit is the last instance of the text `FATAL` in the log file that is being searched. The output typically shows what was printed to the screen and gives a general indication of where the failure occurred. It is also helpful to search for the terms `fail` and `error`. In this case, the failure occurred while creating the GPFS cluster with the default profile:

```
2016-01-28 09:28:52,994 [ FATAL ]
localhost.localdomain failure whilst: Creating GPFS cluster with default profile  (SS04)
```

2. Search further upwards for the text `FATAL` to find its first occurrence in the log file.

   In this example, the text `FATAL` is found 3 times, wherein the following is its first occurrence in the log file:

   Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain
[2016-01-28T09:28:52+00:00] ERROR: Running exception handlers
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain Running handlers complete
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain
```

```
#[0m[2016-01-28T09:28:52+00:00] ERROR: Exception handlers complete
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain
[2016-01-28T09:28:52+00:00] FATAL: Stacktrace dumped to /var/chef/cache/chef-stacktrace.out
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain Chef Client failed.
3 resources updated in 14.197169001 seconds#[0m
2016-01-28 09:28:52,826 [ TRACE ] localhost.localdomain
[2016-01-28T09:28:52+00:00] ERROR: execute[create_GPFS_cluster_default_profile]
(gpfs::gpfs_cluster_create line 20) had an error: Mixlib::ShellOut::ShellCommandFailed:
Expected process to exit with [0], but received '1'
```

This log snippet mentions the exact Chef recipe (`gpfs::gpfs_cluster_create`) that failed during install.

3. To find more information, visually search upwards within the log file.

   Root cause output is typically close to this first occurrence (time-wise) of the text `FATAL`. Following is a snippet of the log text above the first occurrence of `FATAL`. It shows the start of the **Creating GPFS cluster** portion and then shows where the first error occurred: a stanza encapsulated in "=======" symbols. Immediately following this is the command executed by the installation toolkit:

```
/usr/lpp/mmfs/bin/mmcrcluster -N /tmp/NodesDesc -r /usr/bin/ssh -R /usr/bin/scp \
-C spectrumscale.example.com --profile gpfsprotocoldefaults
```

   Following that is a `STDERR: Warning` when adding the host details to the list of known hosts. Because of this the installation has failed.

   Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-28 09:28:44,583 [ INFO  ] [localhost.localdomain 28-01-2016 09:28:44]
IBM SPECTRUM SCALE: Creating GPFS cluster with default profile (SS04)
2016-01-28 09:28:44,583 [ TRACE ] localhost.localdomain
#[0m  * log[IBM SPECTRUM SCALE: Creating GPFS cluster with default profile (SS04).] action write
2016-01-28 09:28:44,583 [ TRACE ] localhost.localdomain
2016-01-28 09:28:52,778 [ TRACE ] localhost.localdomain
#[0m  * execute[create_GPFS_cluster_default_profile] action run
2016-01-28 09:28:52,778 [ TRACE ] localhost.localdomain #[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain ================================================================#[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain #[31mError executing action `run` on resource
                          'execute[create_GPFS_cluster_default_profile]'#[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain ================================================================#[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain #[0m    Mixlib::ShellOut::ShellCommandFailed#[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain     ------------------------------------#[0m
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain     Expected process to exit with [0], but received '1'
2016-01-28 09:28:52,779 [ TRACE ] localhost.localdomain #[0m
---- Begin output of /usr/lpp/mmfs/bin/mmcrcluster -N /tmp/NodesDesc -r /usr/bin/ssh -R /usr/bin/scp \
-C spectrumscale.example.com --profile gpfsprotocoldefaults  ----
2016-01-28 09:28:52,780 [ TRACE ] localhost.localdomain #[0m
STDOUT: mmcrcluster: Performing preliminary node verification ...
2016-01-28 09:28:52,780 [ TRACE ] localhost.localdomain #[0m
mmcrcluster: Processing quorum and other critical nodes ...
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
STDERR: spectrum-scale-102.example.com:
Warning: Permanently added 'spectrum-scale-102.example.com,192.168.100.102' (ECDSA) to the list of known hosts.
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
spectrum-scale-102.example.com:
checkNewClusterNode:success:%%home%%:20_MEMBER_NODE::0:1:localhost:%3A%3A1:
localhost:manager::::::localhost:localhost:1502:4.2.0.1:Linux:Q::::::::
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
spectrum-scale-103.example.com:
Warning: Permanently added 'spectrum-scale-103.example.com,192.168.100.103'
(ECDSA) to the list of known hosts.
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
spectrum-scale-103.example.com:
checkNewClusterNode:success:%%home%%:20_MEMBER_NODE::0:1:localhost:%3A%3A1:
localhost:manager::::::localhost:localhost:1502:4.2.0.1:Linux:Q::::::::
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
mmcrcluster: Removing GPFS cluster files from the nodes in the cluster . . .
2016-01-28 09:28:52,781 [ TRACE ] localhost.localdomain #[0m
mmcrcluster: Command failed. Examine previous error messages to determine cause.
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
---- End output of /usr/lpp/mmfs/bin/mmcrcluster -N /tmp/NodesDesc -r /usr/bin/ssh -R /usr/bin/scp \
-C spectrumscale.example.com --profile gpfsprotocoldefaults  ----
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
Ran /usr/lpp/mmfs/bin/mmcrcluster -N /tmp/NodesDesc -r /usr/bin/ssh -R /usr/bin/scp \
-C spectrumscale.example.com --profile gpfsprotocoldefaults  returned 1#[0m
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m    Resource Declaration:#[0m
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain     ---------------------#[0m
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain     #
In /var/chef/cache/cookbooks/gpfs/recipes/gpfs_cluster_create.rb
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
20: execute 'create_GPFS_cluster_default_profile' do
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
21:   command "#{node['gpfs']['gpfs_path']}/mmcrcluster -N /tmp/NodesDesc -r
#{node['gpfs']['RemoteShellCommand']} -R #{node['gpfs']['RemoteFileCopy']}
```

```
-C #{node['gpfs']['cluster_name']} --profile gpfsprotocoldefaults "
2016-01-28 09:28:52,782 [ TRACE ] localhost.localdomain #[0m
22:   not_if { node['gpfs']['profile'] == 'randomio' }
2016-01-28 09:28:52,783 [ TRACE ] localhost.localdomain #[0m
23:   not_if "#{node['gpfs']['gpfs_path']}/mmlscluster"
2016-01-28 09:28:52,783 [ TRACE ] localhost.localdomain #[0m
24:   action :run
2016-01-28 09:28:52,783 [ TRACE ] localhost.localdomain #[0m
25: end
2016-01-28 09:28:52,783 [ TRACE ] localhost.localdomain #[0m
26:
2016-01-28 09:28:52,783 [ TRACE ] localhost.localdomain #[0m
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m    Compiled Resource:#[0m
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain     ------------------#[0m
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain
# Declared in /var/chef/cache/cookbooks/gpfs/recipes/gpfs_cluster_create.rb:20:in `from_file'
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m
execute("create_GPFS_cluster_default_profile") do
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m
action [:run]
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m
retries 0
2016-01-28 09:28:52,784 [ TRACE ] localhost.localdomain #[0m
retry_delay 2
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
default_guard_interpreter :execute
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
command "/usr/lpp/mmfs/bin/mmcrcluster -N /tmp/NodesDesc -r /usr/bin/ssh -R /usr/bin/scp \
-C spectrumscale.example.com --profile gpfsprotocoldefaults "
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
backup 5
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
returns 0
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
declared_type :execute
2016-01-28 09:28:52,785 [ TRACE ] localhost.localdomain #[0m
cookbook_name "gpfs"
2016-01-28 09:28:52,786 [ TRACE ] localhost.localdomain #[0m
recipe_name "gpfs_cluster_create"
2016-01-28 09:28:52,786 [ TRACE ] localhost.localdomain #[0m
not_if { #code block }
2016-01-28 09:28:52,786 [ TRACE ] localhost.localdomain #[0m
not_if "/usr/lpp/mmfs/bin/mmlscluster"
2016-01-28 09:28:52,786 [ TRACE ] localhost.localdomain #[0m    end
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain #[0m
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain #[0m#[0m
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain Running handlers:#[0m
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain [2016-01-28T09:28:52+00:00]
ERROR: Running exception handlers
2016-01-28 09:28:52,787 [ TRACE ] localhost.localdomain Running handlers complete
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain #[0m[2016-01-28T09:28:52+00:00]
ERROR: Exception handlers complete
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain [2016-01-28T09:28:52+00:00]
FATAL: Stacktrace dumped to /var/chef/cache/chef-stacktrace.out
2016-01-28 09:28:52,788 [ TRACE ] localhost.localdomain Chef Client failed.
3 resources updated in 14.197169001 seconds#[0m
2016-01-28 09:28:52,826 [ TRACE ] localhost.localdomain
[2016-01-28T09:28:52+00:00] ERROR: execute[create_GPFS_cluster_default_profile]
(gpfs::gpfs_cluster_create line 20) had an error:
Mixlib::ShellOut::ShellCommandFailed: Expected process to exit with [0], but received '1'
```

Workaround

In this case, verify that passwordless SSH is set up properly. The installation toolkit performs verification during the precheck phase to ensure that passwordless SSH is set up correctly. you can also verify manually. For more information, see “Passwordless SSH setup” on page 283. Once passwordless SSH is set up properly between all nodes, installation can be initiated again.

The warning that indicates that the host is added to the list of known hosts helped in determining that the passwordless SSH setup is improper. If passwordless SSH were completely set up before this installation, the host would already have existed within the known hosts file.

Note: IBM Spectrum Scale requires all admin nodes to have passwordless SSH to and from all other nodes of the cluster.

Example 2 - Deployment failed and the bottom of the log file contains the following Chef output, which is not indicative of the error

Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-15 15:31:14,912 [ TRACE ] Stopping chef zero
2016-01-15 15:31:14,913 [ ERROR ] The following error was encountered:
Traceback (most recent call last):
  File "/usr/lpp/mmfs/4.2.0.0/installer/espylib/reporting.py", line 222, in log_to_file
    yield handler
  File "/usr/lpp/mmfs/4.2.0.0/installer/espylib/install.py", line 167, in _install
    setup.install(config)
  File "/usr/lpp/mmfs/4.2.0.0/installer/espylib/setup/ces.py", line 325, in install
    self.deploy(config.protocol_nodes, options_fn)
  File "/usr/lpp/mmfs/4.2.0.0/installer/espylib/deploy.py", line 133, in deploy_nodes
    raise DeployError()
DeployError: Installation failed on one or more nodes. Check the log for more details.
2016-01-15 15:31:14,957 [ INFO  ] Detailed error log:
/usr/lpp/mmfs/4.2.0.0/installer/logs/DEPLOY-15-01-2016_15:29:59.log
```

1. To find more details, go to the end of the log file and search upwards for the text FATAL.

   In this example, the first search hit is the last instance of the word FATAL in the log file that is being searched. The output typically shows what was printed to the screen and gives a general indication of where the failure occurred. It is also helpful to search for the terms 'fail' and 'error'. In this case, the failure occurred while installing object packages:

   Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-15 15:31:09,762 [ FATAL ]
objnode4 failure whilst: Installing Object packages  (SS50)
2016-01-15 15:31:09,770 [ WARN  ] SUGGESTED ACTION(S):
2016-01-15 15:31:09,770 [ WARN  ]
Check Object dependencies are available via your package manager or are already met prior to installation.
2016-01-15 15:31:09,770 [ FATAL ]
objnode3 failure whilst: Installing Object packages  (SS50)
```

2. Search upwards further for the text FATAL to find its first occurrence in the log file.

   In this example, the text FATAL is found 8 times, wherein the following is its first occurrence in the log file:

   Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-15 15:31:09,447 [ TRACE ] objnode4 [2016-01-15T15:31:09+05:30] ERROR: Running exception handlers
2016-01-15 15:31:09,447 [ TRACE ] objnode4 Running handlers complete
2016-01-15 15:31:09,447 [ TRACE ] objnode4 [0m[2016-01-15T15:31:09+05:30] ERROR: Exception handlers complete
2016-01-15 15:31:09,448 [ TRACE ] objnode4 [2016-01-15T15:31:09+05:30]
FATAL: Stacktrace dumped to /var/chef/cache/chef-stacktrace.out
2016-01-15 15:31:09,448 [ TRACE ] objnode4 Chef Client failed. 32 resources updated in 46.185382251 seconds[0m
2016-01-15 15:31:09,474 [ TRACE ] objnode4 [2016-01-15T15:31:09+05:30]
ERROR: yum_package[spectrum-scale-object] (swift_on_gpfs::swift_node_install line 14) had an error:
Chef::Exceptions::Exec: yum -d0 -e0 -y install spectrum-scale-object-4.2.0-0 returned 1:
```

   This log snippet mentions the exact Chef recipe (swift_on_gpfs::swift_node_install) that failed during deployment.

3. To find more information, visually search upwards within the log file.

   Root cause output is typically close to the first occurrence of the text FATAL. Following is a snippet of the log text above the first occurrence of FATAL. It shows the start of the **Installing Object packages** portion of the deployment and then shows where the first error occurred: a stanza encapsulated in "=======" symbols. Immediately following this is the command executed by the installation toolkit for deployment:

```
yum -do -e0 -y install spectrum-scale-object-4.2.0.0
```

   Following that is a STDERR: Error showing that a specific package, libcap-ng is already installed on this node with version 0.7.5-4, yet this specific code level requires version 0.7.3-5 of libcap-ng. Because libcap-ng version 0.7.3-5 is a dependency for spectrum-scale-object-4.2.0-0, the deployment has failed.

   Note: The following log text has been adjusted to fit in the PDF margin.

```
2016-01-15 15:30:51,858 [ INFO  ] [objnode3 15-01-2016 15:30:51]
IBM SPECTRUM SCALE: Installing Object packages (SS50)
2016-01-15 15:30:51,858 [ TRACE ] objnode3   * log
[IBM SPECTRUM SCALE: Installing Object packages (SS50).] action write
2016-01-15 15:30:51,859 [ TRACE ] objnode3
2016-01-15 15:31:09,441 [ TRACE ] objnode4 [0m
* yum_package[spectrum-scale-object] action install
2016-01-15 15:31:09,441 [ TRACE ] objnode4     [0m
2016-01-15 15:31:09,441 [ TRACE ] objnode4     ============================================================[0m
2016-01-15 15:31:09,442 [ TRACE ] objnode4     [31mError executing action `install`
                                               on resource 'yum_package[spectrum-scale-object]'[0m
2016-01-15 15:31:09,442 [ TRACE ] objnode4     ============================================================[0m
2016-01-15 15:31:09,442 [ TRACE ] objnode4
2016-01-15 15:31:09,442 [ TRACE ] objnode4 [0m    Chef::Exceptions::Exec[0m
2016-01-15 15:31:09,442 [ TRACE ] objnode4     ----------------------[0m
2016-01-15 15:31:09,442 [ TRACE ] objnode4
yum -d0 -e0 -y install spectrum-scale-object-4.2.0-0 returned 1:
2016-01-15 15:31:09,442 [ TRACE ] objnode4 [0m
STDOUT:  You could try using --skip-broken to work around the problem
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
You could try running: rpm -Va --nofiles --nodigest
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
STDERR: Error: Package: libcap-ng-python-0.7.3-5.el7.x86_64 (ces_object)
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
Requires: libcap-ng = 0.7.3-5.el7
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
Installed: libcap-ng-0.7.5-4.el7.x86_64 (@RHEL7.1)
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
libcap-ng = 0.7.5-4.el7
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,443 [ TRACE ] objnode4 [0m
Resource Declaration:[0m
2016-01-15 15:31:09,444 [ TRACE ] objnode4
---------------------[0m
2016-01-15 15:31:09,444 [ TRACE ] objnode4
# In /var/chef/cache/cookbooks/swift_on_gpfs/recipes/swift_node_install.rb
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
14:   package pkg do
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
15:     retries 3
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
16:     retry_delay 3
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
17:   end
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
18: end
2016-01-15 15:31:09,444 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
Compiled Resource:[0m
2016-01-15 15:31:09,445 [ TRACE ] objnode4     ------------------[0m
2016-01-15 15:31:09,445 [ TRACE ] objnode4
# Declared in /var/chef/cache/cookbooks/swift_on_gpfs/recipes/swift_node_install.rb:14:in `block in from_file'
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
yum_package("spectrum-scale-object") do
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
action :install
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
retries 3
2016-01-15 15:31:09,445 [ TRACE ] objnode4 [0m
retry_delay 3
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
default_guard_interpreter :default
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
package_name "spectrum-scale-object"
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
version "4.2.0-0"
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
timeout 900
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
flush_cache {:before=>false, :after=>false}
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
declared_type :package
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
cookbook_name "swift_on_gpfs"
2016-01-15 15:31:09,446 [ TRACE ] objnode4 [0m
recipe_name "swift_node_install"
2016-01-15 15:31:09,447 [ TRACE ] objnode4 [0m    end
2016-01-15 15:31:09,447 [ TRACE ] objnode4 [0m
2016-01-15 15:31:09,447 [ TRACE ] objnode4 [0m[0m
2016-01-15 15:31:09,447 [ TRACE ] objnode4 Running handlers:[0m
2016-01-15 15:31:09,447 [ TRACE ] objnode4
[2016-01-15T15:31:09+05:30] ERROR: Running exception handlers
2016-01-15 15:31:09,447 [ TRACE ] objnode4
Running handlers complete
2016-01-15 15:31:09,447 [ TRACE ] objnode4
[0m[2016-01-15T15:31:09+05:30] ERROR: Exception handlers complete
2016-01-15 15:31:09,448 [ TRACE ] objnode4
```

```
[2016-01-15T15:31:09+05:30] FATAL: Stacktrace dumped to
/var/chef/cache/chef-stacktrace.out
2016-01-15 15:31:09,448 [ TRACE ] objnode4 Chef Client failed.
32 resources updated in 46.185382251 seconds[0m
2016-01-15 15:31:09,474 [ TRACE ] objnode4
[2016-01-15T15:31:09+05:30] ERROR: yum_package[spectrum-scale-object]
(swift_on_gpfs::swift_node_install line 14) had an error:
Chef::Exceptions::Exec: yum -d0 -e0 -y install spectrum-scale-object-4.2.0-0 returned 1:
```

Workaround

Check the version of `libcap-ng` installed on the node(s) and install the same version of `libcap-ng-python` on the node(s). Once this is done on all nodes, deployment can be initiated again.

Problems due to missing prerequisites

Use this information to ensure that prerequisites are met before using the installation toolkit for installation, deployment, and upgrade.

Passwordless SSH setup

The installation toolkit performs verification during the precheck phase to ensure that passwordless SSH is set up correctly. You can manually verify and set up passwordless SSH as follows.

1. Verify that passwordless SSH is set up using the following commands.

   ```
   ssh <host name of the first node>
   ssh <host name of the second node>
   ```

   Repeat this on all nodes.

   Verify that the user can log into the node successfully without being prompted for any input and that there are no warnings.

   ```
   ssh <FQDN of the first node>
   ssh <FQDN of the second node>
   ```

   Repeat this on all nodes.

   Verify that the user can log into the node successfully without being prompted for any input and that there are no warnings.

   ```
   ssh <IP address of the first node>
   ssh <IP address of the second node>
   ```

   Repeat this on all nodes.

   Verify that the user can log into the node successfully without being prompted for any input and that there are no warnings.

2. If needed, set up passwordless SSH using the following commands.

   Note: This is one of the several possible ways of setting up passwordless SSH.

   ```
   ssh-keygen
   ```

   Repeat this on all cluster nodes.

   ```
   ssh-copy-id <host name of the first node>
   ssh-copy-id <host name of the second node>
   ```

   Repeat this on all nodes.

   ```
   ssh-copy-id <FQDN of the first node>
   ssh-copy-id <FQDN of the second node>
   ```

   Repeat this on all nodes.

Repository setup

- Verify that the repository is set up depending on your operating system. For example, verify that yum repository is set up using the following command on all cluster nodes.

  ```
  yum repolist
  ```

  This command should run clean with no errors if the Yum repository is set up.

Firewall configuration

It is recommended that firewalls are in place to secure all nodes. For more information, see *Securing the IBM Spectrum Scale system using firewall* in *IBM Spectrum Scale: Administration Guide*.

- If you need to open specific ports, use the following steps on Red Hat Enterprise Linux nodes.
  1. Check the firewall status.

     ```
     systemctl status firewalld
     ```

  2. Open ports required by the installation toolkit.

     ```
     firewall-cmd --permanent --add-port 8889/tcp
     firewall-cmd --add-port 8889/tcp
     firewall-cmd --permanent --add-port 10080/tcp
     firewall-cmd --add-port 10080/tcp
     ```

CES IP address allocation

As part of the deployment process, the IBM Spectrum Scale checks routing on the cluster and applies CES IPs as aliases on each protocol node. Furthermore, as service actions or failovers, nodes dynamically lose the alias IPs as they go down and other nodes gain additional aliases to hold all of the IPs passed to them from the down nodes.

Example - Before deployment

The only address here is 192.168.251.161, which is the ssh address for the node. It is held by the `eth0` adapter.

```
# ifconfig -a
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>mtu 1500
        inet 192.168.251.161  netmask 255.255.254.0  broadcast 192.168.251.255
        inet6 2002:90b:e006:84:250:56ff:fea5:1d86  prefixlen 64  scopeid 0x0<global>
        inet6 fe80::250:56ff:fea5:1d86  prefixlen 64  scopeid 0x20<link>
        ether 00:50:56:a5:1d:86  txqueuelen 1000  (Ethernet)
        RX packets 1978638  bytes 157199595 (149.9 MiB)
        RX errors 0  dropped 2291  overruns 0  frame 0
        TX packets 30884  bytes 3918216 (3.7 MiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

# ip addr
2: eth0:<BROADCAST,MULTICAST,UP,LOWER_UP>mtu 1500 qdisc mq state UP qlen 1000
    link/ether 00:50:56:a5:1d:86 brd ff:ff:ff:ff:ff:ff
    inet 192.168.251.161/23 brd 192.168.251.255 scope global eth0
       valid_lft forever preferred_lft forever
    inet6 2002:90b:e006:84:250:56ff:fea5:1d86/64 scope global dynamic
       valid_lft 2591875sec preferred_lft 604675sec
    inet6 fe80::250:56ff:fea5:1d86/64 scope link
       valid_lft forever preferred_lft forever
```

Example - After deployment

Now that the CES IP addresses exist, you can see that aliases called `eth0:0` and `eth0:1` have been created and the CES IP addresses specific to this node have been tagged to it. This allows the ssh IP of the node to exist at the same time as the CES IP address on the same adapter, if necessary. In this example, 192.168.251.161 is the initial ssh IP. The CES IP 192.168.251.165 is aliased onto `eth0:0` and the CES IP

192.168.251.166 is aliased onto eth0:1. This occurs on all protocol nodes that are assigned a CES IP address. NSD server nodes or any client nodes that do not have protocols installed on them do not get a CES IP.

Furthermore, as service actions or failovers, nodes dynamically lose the alias IPs as they go down and other nodes gain additional aliases such as eth0:1 and eth0:2 to hold all of the IPs passed to them from the down nodes.

```
# ifconfig -a
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500
        inet 192.168.251.161  netmask 255.255.254.0  broadcast 192.168.251.255
        inet6 2002:90b:e006:84:250:56ff:fea5:1d86  prefixlen 64  scopeid 0x0<global>
        inet6 fe80::250:56ff:fea5:1d86  prefixlen 64  scopeid 0x20<link>
     ether 00:50:56:a5:1d:86  txqueuelen 1000  (Ethernet)
        RX packets 2909840  bytes 1022774886 (975.3 MiB)
        RX errors 0  dropped 2349  overruns 0  frame 0
        TX packets 712595  bytes 12619844288 (11.7 GiB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
eth0:0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>mtu 1500
        inet 192.168.251.165  netmask 255.255.254.0  broadcast 192.168.251.255
        ether 00:50:56:a5:1d:86  txqueuelen 1000  (Ethernet)
eth0:1: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>mtu 1500
        inet 192.168.251.166  netmask 255.255.254.0  broadcast 192.168.251.255
        ether 00:50:56:a5:1d:86  txqueuelen 1000  (Ethernet)

# ip addr
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP>mtu 1500 qdisc mq state UP qlen 1000
    link/ether 00:50:56:a5:1d:86 brd ff:ff:ff:ff:ff:ff
    inet 192.168.251.161/23 brd 9.11.85.255 scope global eth0
       valid_lft forever preferred_lft forever
    inet 192.168.251.165/23 brd 9.11.85.255 scope global secondary eth0:0
       valid_lft forever preferred_lft forever
    inet 192.168.251.166/23 brd 9.11.85.255 scope global secondary eth0:1
       valid_lft forever preferred_lft forever
    inet6 2002:90b:e006:84:250:56ff:fea5:1d86/64 scope global dynamic
       valid_lft 2591838sec preferred_lft 604638sec
    inet6 fe80::250:56ff:fea5:1d86/64 scope link
       valid_lft forever preferred_lft forever
```

Addition of CES IPs to /etc/hosts

Although it is highly recommended that all CES IPs are maintained in a central DNS and that they are accessible using both forward and reverse DNS lookup, there are times when this might not be possible. IBM Spectrum Scale always verify that forward or reverse DNS lookup is possible. To satisfy this check without a central DNS server containing the CES IPs, you must add the CES IPs to /etc/hosts and create a host name for them within /etc/hosts. The following example shows how a cluster might have multiple networks, nodes, and IPs defined.

For example:

```
# cat /etc/hosts
127.0.0.1   localhost localhost.localdomain localhost4 localhost4.localdomain4
::1         localhost localhost.localdomain localhost6 localhost6.localdomain6

# These are external addresses for GPFS
# Use these for ssh in. You can also use these to form your GPFS cluster if you choose
198.51.100.2   ss-deploy-cluster3-1.example.com   ss-deploy-cluster3-1
198.51.100.4   ss-deploy-cluster3-2.example.com   ss-deploy-cluster3-2
198.51.100.6   ss-deploy-cluster3-3.example.com   ss-deploy-cluster3-3
198.51.100.9   ss-deploy-cluster3-4.example.com   ss-deploy-cluster3-4

# These are addresses for the base adapter used to alias CES-IPs to.
# Do not use these as CES-IPs.
# You could use these for a gpfs cluster if you choose
# Or you could leave these unused as placeholders
```

```
203.0.113.7    ss-deploy-cluster3-1_ces.example.com   ss-deploy-cluster3-1_ces
203.0.113.10   ss-deploy-cluster3-2_ces.example.com   ss-deploy-cluster3-2_ces
203.0.113.12   ss-deploy-cluster3-3_ces.example.com   ss-deploy-cluster3-3_ces
203.0.113.14   ss-deploy-cluster3-4_ces.example.com   ss-deploy-cluster3-4_ces

# These are addresses to use for CES-IPs
203.0.113.17  ss-deploy-cluster3-ces.example.com   ss-deploy-cluster3-ces
203.0.113.20  ss-deploy-cluster3-ces.example.com   ss-deploy-cluster3-ces
203.0.113.21  ss-deploy-cluster3-ces.example.com   ss-deploy-cluster3-ces
203.0.113.23  ss-deploy-cluster3-ces.example.com   ss-deploy-cluster3-ces
```

In this example, the first two sets of addresses have unique host names and the third set of addresses that are associated with CES IPs are not unique. Alternatively, you could give each CES IP a unique host name but this is an arbitrary decision because only the node itself can see its own `/etc/hosts` file. Therefore, these host names are not visible to external clients/nodes unless they too contain a mirror copy of the `/etc/hosts` file. The reason for containing the CES IPs within the `/etc/hosts` file is solely to satisfy the IBM Spectrum Scale CES network verification checks. Without this, in cases with no DNS server, CES IPs cannot be added to a cluster.

Problems due to mixed operating system levels in the cluster

Use the following guidelines to avoid problems due to mixed operating system levels in an IBM Spectrum Scale cluster.

For latest information about supported operating systems, see IBM Spectrum Scale FAQ in IBM Knowledge Center(www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

Verify that the installation toolkit is configured to operate only on supported nodes by using the following command:

```
./spectrumscale node list
```

If any of the listed nodes are of an unsupported OS type, then they need to be removed by using the following command:

```
./spectrumscale node delete node
```

If the node to be removed is an NSD node, then you might have to manually create NSDs and file systems before using the installation toolkit.

The installation toolkit does not need to be made aware of preexisting file systems and NSDs that are present on unsupported node types. Ensure that the file systems are mounted before running the installation toolkit and that they point at the mount points or directory structures.

For information about how the installation toolkit can be used in a cluster that has nodes with mixed operating systems, see **Mixed operating system support with the installation toolkit** in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Upgrades in a mixed OS cluster

Upgrades in a mixed OS cluster need to be performed carefully due to a mix of manual and automated steps. In this case, the installation toolkit can be made aware of a list of nodes that are running on supported OS that are to be upgraded. It can then upgrade these nodes. However, the remaining nodes need to be upgraded manually.

Problems due to using the installation toolkit for functions or configurations not supported

Use this information to determine node types, setups, and functions supported with the installation toolkit, and to understand how to use the toolkit if a setup is not fully supported.

Support for mixed mode of install, deploy, or upgrade

I want to use the installation toolkit but I already have an existing cluster. Can the installation toolkit auto-detect my cluster or do I have to manually configure the toolkit?

The installation toolkit is stateless and it does not import an existing cluster configuration into its cluster definition file. As a workaround to this scenario, use the steps in these topics of *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

- *Deploying protocols on an existing cluster*
- *Deploying protocols authentication on an existing cluster*
- *Adding nodes, NSDs, or file systems to an existing installation*
- *Enabling another protocol on an existing cluster that has protocols enabled*

If NSDs and file systems already exist, you do not need to provide that information to the installation toolkit.

What are valid starting scenarios for which the installation toolkit can be used for an installation or a deployment or an upgrade?

Scenario	Installation toolkit support
No cluster exists and no GPFS RPMs exist on any nodes.	The installation toolkit can be used to install GPFS and create a cluster.
No cluster exists and GPFS RPMs are already installed on nodes.	The installation toolkit can be used to install GPFS and create a cluster.
No cluster exists	The installation toolkit can be used to configure NTP during GPFS installation and cluster configuration.
No cluster exists	The installation GUI can be used to create a cluster.
A cluster exists	The installation toolkit can be used to add NSDs.
A cluster exists	The installation toolkit can be used to add nodes (manager, quorum, admin, nsd, protocol, gui).
A cluster exists and NSDs exist	The installation toolkit can be used to add file systems.
A cluster exists and some NSDs exist	The installation toolkit can be used to add more NSDs.
A cluster exists and some protocols are enabled	The installation toolkit can be used to enable more protocols.
A cluster exists and performance monitoring is enabled	The installation toolkit can be used to reconfigure performance monitoring.
An ESS cluster exists and protocol nodes have been added	The installation toolkit can be used to add protocols to protocol nodes.
SLES 11, Windows, Debian, RHEL 6.8, and AIX nodes exist along with RHEL 7.x (7.1 and later), Ubuntu 16.04, and SLES 12 nodes	The installation toolkit can be used only on RHEL 7.x (7.1 and later), Ubuntu 16.04, and SLES 12 nodes.
A cluster is at mixed levels of 4.2.0.x	The installation toolkit can be used to upgrade all nodes or a subset of nodes to a common code level.

What are invalid starting scenarios for the installation toolkit?

- NSDs were not cleaned up or deleted prior to a cluster deletion.
- Unsupported node types were added to the installation toolkit.
- File systems or NSDs are served by unsupported node types.

The installation toolkit cannot add or change these. It can only use file system paths for protocol configuration.

- An ESS cluster exists and protocol nodes have not yet been added to the cluster.

  Protocol nodes must first be added to the ESS cluster before the installation toolkit can install the protocols.

Does the installation toolkit need to have my entire cluster information?

No, but this depends on the use case. Here are some examples in which the installation toolkit does not need to be made aware of the configuration information of an existing cluster:

- **Deploying protocols on protocol nodes:** The installation toolkit needs only the protocol nodes information and that they are configured to point to `cesSharedRoot`.
- **Upgrading protocol nodes:** The installation toolkit can upgrade a portion of the cluster such as all protocol nodes. In this case, it does not need to be made aware of other NSD or client/server nodes within the cluster.
- **Adding protocols to an ESS cluster:** The installation toolkit does not need to be made aware of the EMS or I/O nodes. The installation toolkit needs only the protocol nodes information and that they are configured to point to `cesSharedRoot`.
- **Adding protocols to a cluster with AIX, SLES, Debian, RHEL6, and Windows nodes:** The installation toolkit does not need to be made aware of any nodes except for the RHEL 7.x, Ubuntu 16.04, and SLES 12 protocol nodes. The installation toolkit needs only the protocol nodes information and that they are configured to point to `cesSharedRoot`.
-

Can the installation toolkit act on some protocol nodes but not all?

Protocol nodes must always be treated as a group of nodes. Therefore, do not use the installation toolkit to run install, deploy, or upgrade commands on a subset of protocol nodes.

Support for DMAPI enabled nodes

On nodes with DMAPI enabled, the installation toolkit does not provide much help to users in case of an error including whether a DMAPI related function is supported or unsupported.

Use the following steps to verify whether DMAPI is enabled on your nodes and to use the installation toolkit on DMAPI enabled nodes.

1. Verify that DMAPI is enabled on a file system using the following command:

```
# mmlsfs all -z
File system attributes for /dev/fs1:
====================================
flag                value                    description
------------------- ------------------------ -----------------------------------
-z                  yes                      Is DMAPI enabled?
```

2. Shut down all functions that are using DMAPI and unmount DMAPI using the following steps:
   a. Shut down all functions that are using DMAPI. This includes HSM policies and IBM Spectrum Archive™.
   b. Unmount the DMAPI file system from all nodes using the following command:

```
# mmunmount fs1 -a
```

   Note: If the DMAPI file system is also the CES shared root file system, then you must first shut down GPFS on all protocol nodes before unmounting the file system.

   1) Check if the DMAPI file system is also the CES shared root file system, use the following command:

```
# mmlsconfig | grep cesSharedRoot
```

   2) Compare the output of this command with that of Step 1 to determine if the CES shared root file system has DMAPI enabled.

3) Shut down GPFS on all protocol nodes using the following command:

```
# mmshutdown -N cesNodes
```

c. Disable DMAPI using the following command:

```
# mmchfs fs1 -z no
```

3. If GPFS was shut down on the protocol nodes in one of the preceding steps, start GPFS on the protocol nodes using the following command:

```
# mmstartup -N cesNodes
```

4. Remount the file system on all nodes using the following command:

```
# mmmount fs1 -a
```

5. Proceed with using the installation toolkit as now it can be used on all file systems.
6. After the task being done using the installation toolkit is completed, enable DMAPI using the following steps:
   a. Unmount the DMAPI file system from all nodes.

      Note: If the DMAPI file system is also the CES shared root file system, shut down GPFS on all protocol nodes before unmounting the file system.

   b. Enable DMAPI using the following command:

      ```
      # mmchfs fs1 -z yes
      ```

   c. Start GPFS on all protocol nodes.
   d. Remount the file system on all nodes.

Support for ESS cluster

For information on using the installation toolkit with a cluster containing ESS, see the following topics in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*:

- *ESS awareness with the installation toolkit*
- *Preparing a cluster that contains ESS for adding protocols*
- *Deploying protocols on an existing cluster*

Understanding supported upgrade functions with installation toolkit

Use this information to understand the setups in which upgrade can be done using the installation toolkit.

- "Scope of the upgrade process"
- "Understanding implications of a failed upgrade" on page 290

Scope of the upgrade process

The upgrade process using the installation toolkit can be summarized as follows:

- The upgrade process acts upon all nodes specified in the cluster definition file (typically using the **./spectrumscale node add** commands).
- All installed/deployed components are upgraded.
- Upgrades are sequential with multiple passes.

The upgrade process using the installation toolkit comprises following passes:

1. Pass 1 of all nodes upgrades GPFS sequentially.
2. Pass 2 of all nodes upgrades Object sequentially.
3. Pass 3 of all nodes upgrades NFS sequentially.
4. Pass 4 of all nodes upgrades SMB sequentially.
5. A post check is done to verify a healthy cluster state after the upgrade.

As an upgrade moves sequentially across nodes, functions such as SMB, NFS, Object, Performance Monitoring, AFM, etc. undergo failovers. This might cause outages on the nodes being upgraded.

Upgrading a subset of nodes is possible because the installation toolkit acts only on the nodes specified in the cluster definition file. If you want to upgrade a subset of cluster nodes, be aware of the node types and the functions being performed on these nodes. For example, all protocol nodes within a cluster must be upgraded by the installation toolkit in one batch.

Understanding implications of a failed upgrade

A failed upgrade might leave a cluster in a state of containing multiple code levels. It is important to analyze console output to determine which nodes or components were upgraded prior to the failure and which node or component was in the process of being upgraded when the failure occurred.

Once the problem has been isolated, a healthy cluster state must be achieved prior to continuing the upgrade. Use the **mmhealth** command in addition to the **mmces state show -a** command to verify that all services are up. It might be necessary to manually start services that were down when the upgrade failed. Starting the services manually helps achieve a state in which all components are healthy prior to continuing the upgrade.

For more information about verifying service status, see **mmhealth command** and **mmces state show command** in *IBM Spectrum Scale: Command and Programming Reference*.

Installation toolkit hangs indefinitely during a GPFS state check

The installation toolkit might hang indefinitely during a GPFS state check operation. This issue occurs either due to multiple versions of Ruby being installed in the environment or if the user pressed Ctrl+C, or because the node crashes or reboots during this timeframe.

If the Chef knife process is hanging, you can use the following command to determine that this issue is occurring due to multiple versions of Ruby.

```
/opt/chef/embedded/bin/chef-zero -H InstallerNodeIP -p 8889
```

If this issue is occurring due to multiple versions of Ruby, this command generates an output similar to the following.

```
/usr/local/share/ruby/site_ruby/rubygems/dependency.rb:311:in `to_specs':
Could not find 'chef-zero' (>= 0) among 8 total         gem(s) (Gem::MissingSpecError)
Checked in 'GEM_PATH=/root/.gem/ruby:/usr/share/gems:/usr/local/share/gems', execute `gem env` for more information
from /usr/local/share/ruby/site_ruby/rubygems/dependency.rb:323:in `to_spec'
from /usr/local/share/ruby/site_ruby/rubygems/core_ext/kernel_gem.rb:65:in `gem'
from /opt/chef/embedded/bin/chef-zero:22:in `<main>'
```

Workaround:

1. Uninstall Ruby packages by issuing the following command from the installer node.
   ```
   yum remove ruby
   ```
2. Set the installer node.
   ```
   ./spectrumscale setup -s InstallerNodeIP
   ```
3. Set the Chef provided Ruby path into `.bash_profile` or export the path during the current session.
   ```
   export PATH="/opt/chef/embedded/bin:${HOME}/.chef/gem/ruby/2.1.0/bin:$PATH"
   ```
4. Retry the installation toolkit operation.

Installation toolkit hangs during a subsequent session after the first session was terminated

An installation toolkit operation that is initiated after the first session was terminated might hang. The first session could have been terminated either by pressing Ctrl+C, by closing the window, or due to a reboot or some other error. This issue occurs because the Chef or Knife processes initiated during the first installation toolkit session did not get terminated.

Workaround:

1. Verify that the subsequent installation toolkit session is hung by viewing the logs and by observing for a few minutes.
2. If the installation toolkit session is hung, press Ctrl+C to terminate the session.
3. Use the following steps on every node on which the installation toolkit is being used to do installation, deployment, or upgrade, including the installer node.
   a. Identify the Chef processes that are active.
      ```
      ps -ef | grep chef
      ```
   b. Kill each Chef process that is still active.
      ```
      kill -9 ProcessID
      ```
   c. Rerun the following command to ensure that none of the Chef processes is active.
      ```
      ps -ef | grep chef
      ```
   d. Identify the Knife processes that are active.
      ```
      ps -ef | grep knife
      ```
   e. Kill each Knife process that is still active.
      ```
      kill -9 ProcessID
      ```
   f. Rerun the following command to ensure that none of the Knife processes is active.
      ```
      ps -ef | grep knife
      ```
4. Retry the installation toolkit operation.

Installation toolkit setup fails with an ssh-agent related error

The installation toolkit setup might fail with an ssh-agent related error.

The error message is similar to the following:

```
ERROR: Net::SSH::Authentication::AgentError: could not get identity count
```

Workaround:

1. Issue the following command on each node added in the installation toolkit cluster definition.
   ```
   eval "$(ssh-agent)";
   ```
2. Retry the installation procedure using the installation toolkit.

Package conflict on SLES 12 SP1 and SP2 nodes while doing installation, deployment, or upgrade using installation toolkit

While doing installation, deployment, or upgrade using the installation toolkit on SLES 12 SP1 and SP2 nodes, you might encounter package conflict issues.

Symptom:

The error message might be similar to the following:

```
[ FATAL ] nodex.example.com File /usr/lib64/libnss_winbind.so.2
[ FATAL ] node2.example.com   from install of
[ FATAL ] node2.example.com   samba-winbind-4.4.2-31.1.x86_64 (FTP3-SUSE-12-2-Updates)
[ FATAL ] node2.example.com   conflicts with file from package
[ FATAL ] node2.example.com   gpfs.smb-1:4.5.5_gpfs_15-1.sles12.x86_64
```

Workaround:

1. Back up the `zypper.rb` file.
   ```
   cp /opt/chef/embedded/apps/chef/lib/chef/provider/package/zypper.rb  /tmp/
   ```
2. Edit the `zypper.rb` file.
   ```
   vim /opt/chef/embedded/apps/chef/lib/chef/provider/package/zypper.rb
   ```
3. Modify the `install_package` function code to add the `--no-recommends` parameter using the following code snippet.
   ```
   def install_package(name, version)
            zypper_package("install --auto-agree-with-licenses --no-recommends", name, version)
       end
   ```
4. Save the changes in the `zypper.rb` file.
5. Copy the changed `zypper.rb` file on every failure node or do the same code changes on every node.
6. Rerun the installation toolkit from the last failure point.

Note: You can also try using this workaround in scenarios with similar package conflict issues.

Related concepts:

"File conflict issue while upgrading SLES 12 on IBM Spectrum Scale nodes" on page 313
While upgrading SLES 12 on IBM Spectrum Scale nodes using the **zypper up** command, you might encounter file conflicts.

---

systemctl commands time out during installation, deployment, or upgrade with the installation toolkit

In some environments, systemctl commands such as **systemctl daemon-reexec** and **systemctl list-unit-files** might time out during installation, deployment, or upgrade using the installation toolkit.

This causes the installation, deployment, or upgrade operation to fail.

When this issue occurs, a message similar to the following might be present in the installation toolkit log:

```
no implicit conversion of false into Array
```

Workaround:

1. List all the scope files without a directory.
   ```
   for j in $(ls /run/systemd/system/session*.scope);
   do if [[ ! -d /run/systemd/system/$j.d ]];
   then echo $j;
   fi;
   done
   ```
2. Remove all the scope files without a directory.
   ```
   for j in $(ls /run/systemd/system/session*.scope);
   do if [[ ! -d /run/systemd/system/$j.d ]];
   then rm -f $j;
   fi;
   done
   ```
3. Rerun installation, deployment, or upgrade using the installation toolkit.

Chef crashes during installation, upgrade, or deployment using the installation toolkit

The installation toolkit uses the Chef configuration management tool. While installing, upgrading or, deploying IBM Spectrum Scale using the installation toolkit, Chef might crash with an error

similar to the following.

```
Error in `chef-client worker: ppid=10676;start=14:58:30;'
: realloc(): invalid next size: 0x0000000003b56620 ***
```

Workaround

1. Kill the chef-client process using its process ID as follows.
   a. Identify the chef-client process by issuing the following command.

      ```
      ps -ef | grep chef
      ```

      This process might be running on multiple nodes. Therefore, you might need to issue this command on each of these nodes. If the installation process failed after the creation of cluster, you can use the **mmdsh** command to identify the chef-client process on each node it is running on.

      ```
      mmdsh ps -ef | grep chef
      ```

   b. Kill the chef-client process on each node it is running on.
2. Delete all the contents of the `/var/chef/cache/cookbooks` directory by issuing the following command.

   ```
   rm -rf /var/chef/cache/cookbooks
   ```

   This command might need to be issued on multiple nodes. Therefore, log in to each of these nodes and issue this command. If the installation process failed after the creation of cluster, you can use the **mmdsh** command as follows to delete the contents of the `/var/chef/cache/cookbooks` directory on each node.

   ```
   mmdsh rm -rf /var/chef/cache/cookbooks
   ```
3. Rerun the installation, upgrade, or deployment using the installation toolkit.

Chef commands require configuration changes to work in an environment that requires proxy servers

Chef commands might not work until Chef is configured correctly, if your environment requires proxy servers to access internet.

You can configure Chef to work in an environment that requires proxy servers by specifying proxy settings with one or more of the following environment variables:

- `http_proxy`
- `https_proxy`
- `ftp_proxy`
- `no_proxy`

Workaround:

1. Issue the following command to determine the current proxy server on Linux platforms by checking the environment variables.

   ```
   env | grep -i proxy
   ```
2. Issue the following command to set up the installer node for the installation toolkit as follows.

   ```
   ./spectrumscale setup -s InstallerNodeIP
   ```

   Note: Make sure that *InstallerNodeIP* has access to the proxy server, if any.

3. On the installer node, make changes to the `knife.rb` file for environments that use an HTTP proxy or an HTTPS proxy as follows.
   a. Open `knife.rb` in a file editor such as vim.
      ```
      vim ~/.chef/knife.rb
      ```
   b. Add `http_proxy` and `https_proxy` at the end of the file.
      ```
      http_proxy '<http proxy hostname with port number>'
      https_proxy '<https proxy hostname with port number>'
      ```
4. Use the installation toolkit to perform installation, deployment, or upgrade.

Installation toolkit setup on Ubuntu fails due to dpkg database lock issue

The installation toolkit setup on Ubuntu nodes might fail because of a Chef installation failure that occurs due to a dpkg database lock issue.

Symptom:

The error message might be similar to one of the following:

```
Could not get lock /var/lib/apt/lists/lock - open (11: Resource temporarily unavailable)
Unable to lock directory /var/lib/apt/lists/
Could not get lock /var/lib/dpkg/lock - open (11: Resource temporarily unavailable)
Unable to lock the administration directory (/var/lib/dpkg/), is another process using it?
```

Workaround:

1. Identify the apt-get process by issuing the following command.
   ```
   ps -ef | grep apt-get
   ```
   This process might be running on multiple nodes. Therefore, you might need to issue this command on each of these nodes. If the installation process failed after the creation of cluster, you can use the **mmdsh** command to identify the apt-get process on each node it is running on.
   ```
   mmdsh ps -ef | grep apt-get
   ```
2. Kill the apt-get process on each node it is running on.
   ```
   sudo kill Process_ID
   ```
3. Retry the installation toolkit setup. If the error persists, issue following commands and then try again.
   ```
   rm /var/lib/apt/lists/lock
   dpkg –configure -a
   ```

Installation toolkit config populate operation fails to detect object endpoint

The installation toolkit deployment precheck might fail in some cases because the config populate operation could not detect the object endpoint.

However, the deployment precheck identifies this issue and suggests the corrective action.

Workaround

1. Issue the following command to the add the object endpoint:
   ```
   ./spectrumscale config object -e EndPoint
   ```
2. Proceed with the installation, deployment, or upgrade with the installation toolkit.

Post installation and configuration problems

This topic describes the issues that you might encounter after installing or configuring IBM Spectrum Scale.

The *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* provides the step-by-step procedure for installing and migrating IBM Spectrum Scale, however, some problems might occur after installation and configuration if the procedures were not properly followed.

Some of those problems might include:

- Not being able to start GPFS after installation of the latest version. Did you reboot your IBM Spectrum Scale nodes before and after the installation/upgrade of IBM Spectrum Scale? If you did, see "GPFS daemon will not come up" on page 301. If not, reboot. For more information, see the *Initialization of the GPFS daemon* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.
- Not being able to access a file system. See "File system fails to mount" on page 317.
- New GPFS functions do not operate. See "GPFS commands are unsuccessful" on page 306.

Cluster is crashed after reinstallation

This topic describes the steps that you need to perform when a cluster crashes after IBM Spectrum Scale reinstallation.

After reinstalling IBM Spectrum Scale code, check whether the **/var/mmfs/gen/mmsdrfs** file was lost. If it was lost, and an up-to-date version of the file is present on the primary GPFS cluster configuration server, restore the file by issuing this command from the node on which it is missing:

```
mmsdrrestore -p primaryServer
```

where *primaryServer* is the name of the primary GPFS cluster configuration server.

If the **/var/mmfs/gen/mmsdrfs** file is not present on the primary GPFS cluster configuration server, but it is present on some other node in the cluster, restore the file by issuing these commands:

```
mmsdrrestore -p remoteNode -F remoteFile
mmchcluster -p LATEST
```

where *remoteNode* is the node that has an up-to-date version of the **/var/mmfs/gen/mmsdrfs** file, and *remoteFile* is the full path name of that file on that node.

One way to ensure that the latest version of the **/var/mmfs/gen/mmsdrfs** file is always available is to use the **mmsdrbackup** user exit.

If you have made modifications to any of the users exist in **/var/mmfs/etc**, you will have to restore them before starting GPFS.

For additional information, see "Recovery from loss of GPFS cluster configuration data file" on page 299.

Node cannot be added to the GPFS cluster

There is an indication leading you to the conclusion that a node cannot be added to a cluster and steps to follow to correct the problem.

That indication is:

- You issue the **mmcrcluster** or **mmaddnode** command and receive the message:

6027-1598
Node *nodeName* was not added to the cluster. The node appears to already belong to a GPFS cluster.

Steps to follow if a node cannot be added to a cluster:

1. Run the **mmlscluster** command to verify that the node is not in the cluster.
2. If the node is not in the cluster, issue this command on the node that could not be added:
   ```
   mmdelnode -f
   ```
3. Reissue the **mmaddnode** command.

Problems with the /etc/hosts file

This topic describes the issues relating to the `/etc/hosts` file that you might come across while installing or configuring IBM Spectrum Scale.

The **/etc/hosts** file must have a unique node name for each node interface to be used by GPFS. Violation of this requirement results in the message:

6027-1941
Cannot handle multiple interfaces for host *hostName*.

If you receive this message, correct the **/etc/hosts** file so that each node interface to be used by GPFS appears only once in the file.

Linux configuration considerations

This topic describes the Linux configuration that you need to consider while installing or configuring IBM Spectrum Scale on your cluster.

Note: This information applies only to Linux nodes.

Depending on your system configuration, you may need to consider:

1. **Why can only one host successfully attach to the Fibre Channel loop and see the Fibre Channel disks?**

   Your host bus adapter may be configured with an enabled *Hard Loop ID* that conflicts with other host bus adapters on the same Fibre Channel loop.

   To see if that is the case, reboot your machine and enter the adapter bios with **<Alt-Q>** when the Fibre Channel adapter bios prompt appears. Under the Configuration Settings menu, select `Host Adapter Settings` and either ensure that the Adapter Hard Loop ID option is disabled or assign a unique Hard Loop ID per machine on the Fibre Channel loop.
2. **Could the GPFS daemon be terminated due to a memory shortage?**

   The Linux virtual memory manager (VMM) exhibits undesirable behavior for low memory situations on nodes, where the processes with the largest memory usage are killed by the kernel (using OOM killer), yet no mechanism is available for prioritizing important processes that should not be initial candidates for the OOM killer. The GPFS **mmfsd** daemon uses a large amount of pinned memory in the page pool for caching data and metadata, and so the **mmfsd** process is a likely candidate for termination if memory must be freed up.
3. **What are the performance tuning suggestions?**

   For an up-to-date list of tuning suggestions, see the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

   For Linux on Z, see also the Device Drivers, Features, and Commands(www.ibm.com/support/knowledgecenter/api/content/linuxonibm/liaaf/lnz_r_dd.html) topic in the Linux on Z library overview.

Python conflicts while deploying object packages using installation toolkit

While deploying object packages using the installation toolkit, you may encounter a dependency conflict between `python-dnspython` and `python-dns`.

Symptom:

The error messages may be similar to the following:

```
[ INFO ] [shepard7lp1.tuc.stglabs.example.com 12-04-2017 16:39:29] IBM SPECTRUM SCALE:
Installing Object packages (SS50)
[ FATAL ] shepard3lp1.tuc.stglabs.example.com failure whilst: Installing Object packages (SS50)
[ WARN ] SUGGESTED ACTION(S):
[ WARN ] Check Object dependencies are available via your package manager or are already met
prior to installation.
```

Workaround

1. Manually remove the conflicting rpm by issuing the following command:

   ```
   yum remove python-dns
   ```

2. Retry deploying the object packages.

Problems with running commands on other nodes

This topic describes the problems that you might encounter relating to running remote commands during installing and configuring IBM Spectrum Scale.

Many of the GPFS administration commands perform operations on nodes other than the node on which the command was issued. This is achieved by utilizing a remote invocation shell and a remote file copy command. By default these items are **/usr/bin/ssh** and **/usr/bin/scp**. You also have the option of specifying your own remote shell and remote file copy commands to be used instead of the default **ssh** and **scp**. The remote shell and copy commands must adhere to the same syntax forms as **ssh** and **scp** but may implement an alternate authentication mechanism. For more information on the **mmcrcluster** and **mmchcluster** commands, see the *mmcrcluster command* and the *mmchcluster command* pages in the *IBM Spectrum Scale: Command and Programming Reference*. These are problems you may encounter with the use of remote commands.

Authorization problems

This topic describes issues with running remote commands due to authorization problems in IBM Spectrum Scale.

The **ssh** and **scp** commands are used by GPFS administration commands to perform operations on other nodes. The **ssh** daemon (**sshd**) on the remote node must recognize the command being run and must obtain authorization to invoke it.

Note: Use the **ssh** and **scp** commands that are shipped with the OpenSSH package supported by GPFS. Refer to the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html) for the latest OpenSSH information.

For more information, see "Problems due to missing prerequisites" on page 283.

For the **ssh** and **scp** commands issued by GPFS administration commands to succeed, each node in the cluster must have an **.rhosts** file in the home directory for the root user, with file permission set to 600. This **.rhosts** file must list each of the nodes and the root user. If such an **.rhosts** file does not exist on each node in the cluster, the **ssh** and **scp** commands issued by GPFS commands will fail with permission errors, causing the GPFS commands to fail in turn.

If you elected to use installation specific remote invocation shell and remote file copy commands, you must ensure:

1. Proper authorization is granted to all nodes in the GPFS cluster.
2. The nodes in the GPFS cluster can communicate without the use of a password, and without any extraneous messages.

Connectivity problems

This topic describes the issues with running GPFS commands on remote nodes due to connectivity problems.

Another reason why **ssh** may fail is that connectivity to a needed node has been lost. Error messages from **mmdsh** may indicate that connectivity to such a node has been lost. Here is an example:

```
mmdelnode -N k145n04
Verifying GPFS is stopped on all affected nodes ...
mmdsh: 6027-1617 There are no available nodes on which to run the command.
mmdelnode: 6027-1271 Unexpected error from verifyDaemonInactive: mmcommon onall.
  Return code: 1
```

If error messages indicate that connectivity to a node has been lost, use the **ping** command to verify whether the node can still be reached:

```
ping k145n04
PING k145n04: (119.114.68.69): 56 data bytes
<Ctrl- C>
----k145n04 PING Statistics----
3 packets transmitted, 0 packets received, 100% packet loss
```

If connectivity has been lost, restore it, then reissue the GPFS command.

GPFS error messages for rsh problems

This topic describes the error messages that are displayed for rsh issues in IBM Spectrum Scale.

When **rsh** problems arise, the system may display information similar to these error messages:

6027-1615
nodeName remote shell process had return code *value*.

6027-1617
There are no available nodes on which to run the command.

Cluster configuration data file issues

This topic describes the issues that you might encounter with respect to the cluster configuration data files while installing or configuring IBM Spectrum Scale.

GPFS cluster configuration data file issues

This topic describes the issues relating to IBM Spectrum Scale cluster configuration data.

GPFS uses a file to serialize access of administration commands to the GPFS cluster configuration data files. This lock file is kept on the primary GPFS cluster configuration server in the **/var/mmfs/gen/mmLockDir** directory. If a system failure occurs before the cleanup of this lock file, the file will remain and subsequent administration commands may report that the GPFS cluster configuration data files are locked. Besides a serialization lock, certain GPFS commands may obtain an additional lock. This lock is designed to prevent GPFS from coming up, or file systems from being mounted, during critical sections of the command processing. If this happens you will see a message that shows the name of the blocking command, similar to message:

6027-1242
GPFS is waiting for *requiredCondition*.

To release the lock:

1. Determine the PID and the system that owns the lock by issuing:
   ```
   mmcommon showLocks
   ```
   The **mmcommon showLocks** command displays information about the lock server, lock name, lock holder, PID, and extended information. If a GPFS administration command is not responding, stopping the command will free the lock. If another process has this PID, another error occurred to the original GPFS command, causing it to die without freeing the lock, and this new process has the same PID. If this is the case, do not kill the process.
2. If any locks are held and you want to release them manually, from any node in the GPFS cluster issue the command:
   ```
   mmcommon freeLocks <lockName>
   ```

GPFS error messages for cluster configuration data file problems

This topic describes the error messages relating to the cluster configuration data file issues in IBM Spectrum Scale.

When GPFS commands are unable to retrieve or update the GPFS cluster configuration data files, the system may display information similar to these error messages:

6027-1628
Cannot determine basic environment information. Not enough nodes are available.

6027-1630
The GPFS cluster data on *nodeName* is back level.

6027-1631
The commit process failed.

6027-1632
The GPFS cluster configuration data on *nodeName* is different than the data on *nodeName*.

6027-1633
Failed to create a backup copy of the GPFS cluster data on *nodeName*.

Recovery from loss of GPFS cluster configuration data file

This topic describes the procedure for recovering the cluster configuration data file in IBM Spectrum Scale.

A copy of the IBM Spectrum Scale cluster configuration data files is stored in the **/var/mmfs/gen/mmsdrfs** file on each node. For proper operation, this file must exist on each node in the IBM Spectrum Scale cluster. The latest level of this file is guaranteed to be on the primary, and secondary if specified, GPFS cluster configuration server nodes that were defined when the IBM Spectrum Scale cluster was first created with the **mmcrcluster** command.

If the **/var/mmfs/gen/mmsdrfs** file is removed by accident from any of the nodes, and an up-to-date version of the file is present on the primary IBM Spectrum Scale cluster configuration server, restore the file by issuing this command from the node on which it is missing:

```
mmsdrrestore -p primaryServer
```

where *primaryServer* is the name of the primary GPFS cluster configuration server.

If the **/var/mmfs/gen/mmsdrfs** file is not present on the primary GPFS cluster configuration server, but is present on some other node in the cluster, restore the file by issuing these commands:

```
mmsdrrestore -p remoteNode -F remoteFile
mmchcluster -p LATEST
```

where *remoteNode* is the node that has an up-to-date version of the **/var/mmfs/gen/mmsdrfs** file and *remoteFile* is the full path name of that file on that node.

One way to ensure that the latest version of the **/var/mmfs/gen/mmsdrfs** file is always available is to use the **mmsdrbackup** user exit.

Automatic backup of the GPFS cluster data

This topic describes the procedure for automatically backing up the cluster data in IBM Spectrum Scale.

The IBM Spectrum Scale provides an exit, **mmsdrbackup**, that can be used to automatically back up the IBM Spectrum Scale configuration data every time it changes. To activate this facility, follow these steps:

1. Modify the IBM Spectrum Scale-provided version of **mmsdrbackup** as described in its prologue, to accomplish the backup of the **mmsdrfs** file however the user desires. This file is **/usr/lpp/mmfs/samples/mmsdrbackup.sample**.
2. Copy this modified **mmsdrbackup.sample** file to **/var/mmfs/etc/mmsdrbackup** on all of the nodes in the cluster. Make sure that the permission bits for **/var/mmfs/etc/mmsdrbackup** are set to permit execution by root.

The IBM Spectrum Scale system invokes the user-modified version of **mmsdrbackup** in **/var/mmfs/etc** every time a change is made to the **mmsdrfs** file. This will perform the backup of the **mmsdrfs** file according to the user's specifications. For more information on GPFS user exits, see the *GPFS user exits* topic in the *IBM Spectrum Scale: Command and Programming Reference*.

GPFS application calls

Error numbers specific to GPFS applications calls

This topic describes the error numbers specific to GPFS application calls.

When experiencing installation and configuration problems, GPFS may report these error numbers in the operating system error log facility, or return them to an application:

ECONFIG = 215, Configuration invalid or inconsistent between different nodes.
: This error is returned when the levels of software on different nodes cannot coexist. For information about which levels may coexist, see the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

ENO_QUOTA_INST = 237, No Quota management enabled.
: To enable quotas for the file system issue the **mmchfs -Q yes** command. To disable quotas for the file system issue the **mmchfs -Q no** command.

EOFFLINE = 208, Operation failed because a disk is offline
: This is most commonly returned when an open of a disk fails. Since GPFS will attempt to continue operation with failed disks, this will be returned when the disk is first needed to complete a command or application request. If this return code occurs, check your disk subsystem for stopped states and check to determine if the network path exists. In rare situations, this will be reported if disk definitions are incorrect.

EALL_UNAVAIL = 218, A replicated read or write failed because none of the replicas were available.
: Multiple disks in multiple failure groups are unavailable. Follow the procedures in Chapter 19, “Disk issues,” on page 349 for unavailable disks.

6027-341 [D]
Node *nodeName* is incompatible because its maximum compatible version (*number*) is less than the version of this node (*number*).

6027-342 [E]
Node *nodeName* is incompatible because its minimum compatible version is greater than the version of this node (*number*).

6027-343 [E]
Node *nodeName* is incompatible because its version (*number*) is less than the minimum compatible version of this node (*number*).

6027-344 [E]
Node *nodeName* is incompatible because its version is greater than the maximum compatible version of this node (*number*).

GPFS modules cannot be loaded on Linux

You must build the GPFS portability layer binaries based on the kernel configuration of your system. For more information, see *The GPFS open source portability layer* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*. During **mmstartup** processing, GPFS loads the **mmfslinux** kernel module.

Some of the more common problems that you may encounter are:

1. If the portability layer is not built, you may see messages similar to:

```
Mon Mar 26 20:56:30 EDT 2012: runmmfs starting
Removing old /var/adm/ras/mmfs.log.* files:
Unloading modules from /lib/modules/2.6.32.12-0.6-ppc64/extra
runmmfs: The /lib/modules/2.6.32.12-0.6-ppc64/extra/mmfslinux.ko kernel extension does not exist.
runmmfs: Unable to verify kernel/module configuration.
Loading modules from /lib/modules/2.6.32.12-0.6-ppc64/extra
runmmfs: The /lib/modules/2.6.32.12-0.6-ppc64/extra/mmfslinux.ko kernel extension does not exist.
runmmfs: Unable to verify kernel/module configuration.
Mon Mar 26 20:56:30 EDT 2012 runmmfs: error in loading or unloading the mmfs kernel extension
Mon Mar 26 20:56:30 EDT 2012 runmmfs: stopping GPFS
```

2. The GPFS kernel modules, **mmfslinux** and **tracedev**, are built with a kernel version that differs from that of the currently running Linux kernel. This situation can occur if the modules are built on another node with a different kernel version and copied to this node, or if the node is rebooted using a kernel with a different version.
3. If the **mmfslinux** module is incompatible with your system, you may experience a kernel panic on GPFS startup. Ensure that the **site.mcr** has been configured properly from the **site.mcr.proto**, and GPFS has been built and installed properly.

For more information about the **mmfslinux** module, see the *Building the GPFS portability layer* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

GPFS daemon issues

This topic describes the GPFS daemon issues that you might encounter while installing or configuring IBM Spectrum Scale.

GPFS daemon will not come up

There are several indications that could lead you to the conclusion that the GPFS daemon (**mmfsd**) will not come up and there are some steps to follow to correct the problem.

Those indications include:

- The file system has been enabled to mount automatically, but the mount has not completed.
- You issue a GPFS command and receive the message:

6027-665
Failed to connect to file system daemon: Connection refused.

- The GPFS log does not contain the message:

  6027-300 [N]
  mmfsd ready

- The GPFS log file contains this error message: 'Error: daemon and kernel extension do not match.' This error indicates that the kernel extension currently loaded in memory and the daemon currently starting have mismatching versions. This situation may arise if a GPFS code update has been applied, and the node has not been rebooted prior to starting GPFS.

  While GPFS scripts attempt to unload the old kernel extension during update and install operations, such attempts may fail if the operating system is still referencing GPFS code and data structures. To recover from this error, ensure that all GPFS file systems are successfully unmounted, and reboot the node. The **mmlsmount** command can be used to ensure that all file systems are unmounted.

Steps to follow if the GPFS daemon does not come up

This topic describes the steps that you need to follow if the GPFS daemon does not come up after installation of IBM Spectrum Scale.

1. See "GPFS modules cannot be loaded on Linux" on page 301 if your node is running Linux, to verify that you have built the portability layer.
2. Verify that the GPFS daemon is active by issuing:

   ```
   ps -e | grep mmfsd
   ```

   The output of this command should list **mmfsd** as operational. For example:

   ```
   12230 pts/8 00:00:00 mmfsd
   ```

   If the output does not show this, the GPFS daemon needs to be started with the **mmstartup** command.
3. If you did not specify the **autoload** option on the **mmcrcluster** or the **mmchconfig** command, you need to manually start the daemon by issuing the **mmstartup** command.

   If you specified the **autoload** option, someone may have issued the **mmshutdown** command. In this case, issue the **mmstartup** command. When using **autoload** for the first time, **mmstartup** must be run manually. The **autoload** takes effect on the next reboot.
4. Verify that the network upon which your GPFS cluster depends is up by issuing:

   ```
   ping nodename
   ```

   to each node in the cluster. A properly working network and node will correctly reply to the ping with no lost packets.

   Query the network interface that GPFS is using with:

   ```
   netstat -i
   ```

   A properly working network will report no transmission errors.
5. Verify that the GPFS cluster configuration data is available by looking in the GPFS log. If you see the message:

   6027-1592
   Unable to retrieve GPFS cluster files from node *nodeName*.

   Determine the problem with accessing node *nodeName* and correct it.
6. Verify that the GPFS environment is properly initialized by issuing these commands and ensuring that the output is as expected.
   - Issue the **mmlscluster** command to list the cluster configuration. This will also update the GPFS configuration data on the node. Correct any reported errors before continuing.

- List all file systems that were created in this cluster. For an AIX node, issue:

  ```
  lsfs -v mmfs
  ```

  For a Linux node, issue:

  ```
  cat /etc/fstab | grep gpfs
  ```

  If any of these commands produce unexpected results, this may be an indication of corrupted GPFS cluster configuration data file information. Follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.

7. GPFS requires a quorum of nodes to be active before any file system operations can be honored. This requirement guarantees that a valid single token management domain exists for each GPFS file system. Prior to the existence of a quorum, most requests are rejected with a message indicating that quorum does not exist.

   To identify which nodes in the cluster have daemons **up** or **down**, issue:

   ```
   mmgetstate -L -a
   ```

   If insufficient nodes are active to achieve quorum, go to any nodes not listed as **active** and perform problem determination steps on these nodes. A quorum node indicates that it is part of a quorum by writing an `mmfsd ready` message to the GPFS log. Remember that your system may have quorum nodes and non-quorum nodes, and only quorum nodes are counted to achieve the quorum.

8. This step applies only to AIX nodes. Verify that GPFS kernel extension is not having problems with its shared segment by invoking:

   ```
   cat /var/adm/ras/mmfs.log.latest
   ```

   Messages such as:

   6027-319
   Could not create shared segment.

   must be corrected by the following procedure:

   a. Issue the **mmshutdown** command.
   b. Remove the shared segment in an AIX environment:
      1) Issue the **mmshutdown** command.
      2) Issue the **mmfsadm cleanup** command.
   c. If you are still unable to resolve the problem, reboot the node.

9. If the previous GPFS daemon was brought down and you are trying to start a new daemon but are unable to, this is an indication that the original daemon did not completely go away. Go to that node and check the state of GPFS. Stopping and restarting GPFS or rebooting this node will often return GPFS to normal operation. If this fails, follow the procedures in "Additional information to collect for GPFS daemon crashes" on page 470, and then contact the IBM Support Center.

Unable to start GPFS after the installation of a new release of GPFS

This topic describes the steps that you need to perform if you are unable to start GPFS after installing a new version of IBM Spectrum Scale.

If one or more nodes in the cluster will not start GPFS, these are the possible causes:

- If message:

  6027-2700 [E]
  A node join was rejected. This could be due to incompatible daemon versions, failure to find the node in the configuration database, or no configuration manager found.

  is written to the GPFS log, incompatible versions of GPFS code exist on nodes within the same cluster.

- If messages stating that functions are not supported are written to the GPFS log, you may not have the correct kernel extensions loaded.

  1. Ensure that the latest GPFS install packages are loaded on your system.

2. If running on Linux, ensure that the latest kernel extensions have been installed and built. See the *Building the GPFS portability layer* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.
3. Reboot the GPFS node after an installation to ensure that the latest kernel extension is loaded.

- The daemon will not start because the configuration data was not migrated. See "Post installation and configuration problems" on page 295.

GPFS error messages for shared segment and network problems

This topic describes the error messages relating to issues in shared segment and network in IBM Spectrum Scale.

For shared segment problems, follow the problem determination and repair actions specified with the following messages:

6027-319
Could not create shared segment.

6027-320
Could not map shared segment.

6027-321
Shared segment mapped at wrong address (is *value*, should be *value*).

6027-322
Could not map shared segment in kernel extension.

For network problems, follow the problem determination and repair actions specified with the following message:

6027-306 [E]
Could not initialize inter-node communication

Error numbers specific to GPFS application calls when the daemon is unable to come up

This topic describes the application call error numbers when the daemon is unable to come up.

When the daemon is unable to come up, GPFS may report these error numbers in the operating system error log, or return them to an application:

ECONFIG = 215, Configuration invalid or inconsistent between different nodes.
This error is returned when the levels of software on different nodes cannot coexist. For information about which levels may coexist, see the IBM Spectrum Scale FAQ in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html).

6027-341 [D]
Node *nodeName* is incompatible because its maximum compatible version (*number*) is less than the version of this node (*number*).

6027-342 [E]
Node *nodeName* is incompatible because its minimum compatible version is greater than the version of this node (*number*).

6027-343 [E]
Node *nodeName* is incompatible because its version (*number*) is less than the minimum compatible version of this node (*number*).

6027-344 [E]
Node *nodeName* is incompatible because its version is greater than the maximum compatible version of this node (*number*).

GPFS daemon went down

There are a number of conditions that can cause the GPFS daemon to exit.

These are all conditions where the GPFS internal checking has determined that continued operation would be dangerous to the consistency of your data. Some of these conditions are errors within GPFS processing but most represent a failure of the surrounding environment.

In most cases, the daemon will exit and restart after recovery. If it is not safe to simply force the unmounted file systems to recover, the GPFS daemon will exit.

Indications leading you to the conclusion that the daemon went down:

- Applications running at the time of the failure will see either `ENODEV` or `ESTALE` errors. The `ENODEV` errors are generated by the operating system until the daemon has restarted. The `ESTALE` error is generated by GPFS as soon as it restarts.

  When quorum is lost, applications with open files receive an `ESTALE` error return code until the files are closed and reopened. New file open operations will fail until quorum is restored and the file system is remounted. Applications accessing these files prior to GPFS return may receive a `ENODEV` return code from the operating system.
- The GPFS log contains the message:

  6027-650 [X]
  The mmfs daemon is shutting down abnormally.

  Most GPFS daemon down error messages are in the **mmfs.log.previous** log for the instance that failed. If the daemon restarted, it generates a new **mmfs.log.latest**. Begin problem determination for these errors by examining the operating system error log.

  If an existing quorum is lost, GPFS stops all processing within the cluster to protect the integrity of your data. GPFS will attempt to rebuild a quorum of nodes and will remount the file system if automatic mounts are specified.
- Open requests are rejected with `no such file` or `no such directory` errors.

  When quorum has been lost, requests are rejected until the node has rejoined a valid quorum and mounted its file systems. If messages indicate lack of quorum, follow the procedures in "GPFS daemon will not come up" on page 301.
- Removing the setuid bit from the permissions of these commands may produce errors for non-root users:
  mmdf
  mmgetacl
  mmlsdisk
  mmlsfs
  mmlsmgr
  mmlspolicy
  mmlsquota
  mmlssnapshot
  mmputacl
  mmsnapdir
  mmsnaplatest

  The GPFS system-level versions of these commands (prefixed by **ts**) may need to be checked for how permissions are set if non-root users see the following message:

  6027-1209
  GPFS is down on this node.

  If the setuid bit is removed from the permissions on the system-level commands, the command cannot be executed and the node is perceived as being down. The system-level versions of the commands are:
  tsdf
  tslsdisk

tslsfs
tslsmgr
tslspolicy
tslsquota
tslssnapshot
tssnapdir
tssnaplatest

These are found in the **/usr/lpp/mmfs/bin** directory.

Note: The mode bits for all listed commands are 4555 or **-r-sr-xr-x**. To restore the default (shipped) permission, enter:

```
chmod 4555 tscommand
```

Attention: Only administration-level versions of GPFS commands (prefixed by **mm**) should be executed. Executing system-level commands (prefixed by **ts**) directly will produce unexpected results.

- For all other errors, follow the procedures in "Additional information to collect for GPFS daemon crashes" on page 470, and then contact the IBM Support Center.

GPFS commands are unsuccessful

GPFS commands can be unsuccessful for various reasons.

Unsuccessful command results will be indicated by:

- Return codes indicating the GPFS daemon is no longer running.
- Command specific problems indicating you are unable to access the disks.
- A nonzero return code from the GPFS command.

Some reasons that GPFS commands can be unsuccessful include:

1. If all commands are generically unsuccessful, this may be due to a daemon failure. Verify that the GPFS daemon is active. Issue:

   ```
   mmgetstate
   ```

   If the daemon is not active, check **/var/adm/ras/mmfs.log.latest** and **/var/adm/ras/mmfs.log.previous** on the local node and on the file system manager node. These files enumerate the failing sequence of the GPFS daemon.

   If there is a communication failure with the file system manager node, you will receive an error and the **errno** global variable may be set to EIO (I/O error).
2. Verify the GPFS cluster configuration data files are not locked and are accessible. To determine if the GPFS cluster configuration data files are locked, see "GPFS cluster configuration data file issues" on page 298.
3. The **ssh** command is not functioning correctly. See "Authorization problems" on page 297.

   If **ssh** is not functioning properly on a node in the GPFS cluster, a GPFS administration command that needs to run work on that node will fail with a 'permission is denied' error. The system displays information similar to:

   ```
   mmlscluster
   sshd: 0826-813 Permission is denied.
   mmdsh: 6027-1615 k145n02 remote shell process had return code 1.
   mmlscluster: 6027-1591 Attention: Unable to retrieve GPFS cluster files from node k145n02
   sshd: 0826-813 Permission is denied.
   mmdsh: 6027-1615 k145n01 remote shell process had return code 1.
   mmlscluster: 6027-1592 Unable to retrieve GPFS cluster files from node k145n01
   ```

   These messages indicate that **ssh** is not working properly on nodes **k145n01** and **k145n02**.

If you encounter this type of failure, determine why **ssh** is not working on the identified node. Then fix the problem.

4. Most problems encountered during file system creation fall into three classes:
   - You did not create network shared disks which are required to build the file system.
   - The creation operation cannot access the disk.

     Follow the procedures for checking access to the disk. This can result from a number of factors including those described in "NSD and underlying disk subsystem failures" on page 349.
   - Unsuccessful attempt to communicate with the file system manager.

     The file system creation runs on the file system manager node. If that node goes down, the **mmcrfs** command may not succeed.
5. If the **mmdelnode** command was unsuccessful and you plan to permanently de-install GPFS from a node, you should first remove the node from the cluster. If this is not done and you run the **mmdelnode** command after the **mmfs** code is removed, the command will fail and display a message similar to this example:

   ```
   Verifying GPFS is stopped on all affected nodes ...
   k145n05: ksh: /usr/lpp/mmfs/bin/mmremote: not found.
   ```

   If this happens, power off the node and run the **mmdelnode** command again.
6. If you have successfully installed and are operating with the latest level of GPFS, but cannot run the new functions available, it is probable that you have not issued the **mmchfs -V full** or **mmchfs -V compat** command to change the version of the file system. This command must be issued for *each* of your file systems.

   In addition to **mmchfs -V**, you may need to run the **mmmigratefs** command. See the *File system format changes between versions of GPFS* topic in the *IBM Spectrum Scale: Administration Guide*.

   Note: Before issuing the **-V** option (with **full** or **compat**), see *Upgrading* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*. You must ensure that all nodes in the cluster have been migrated to the latest level of GPFS code and that you have successfully run the **mmchconfig release=LATEST** command.

   Make sure you have operated with the new level of code for some time and are certain you want to migrate to the latest level of GPFS. Issue the **mmchfs -V full** command only after you have definitely decided to accept the latest level, as this will cause disk changes that are incompatible with previous levels of GPFS.

   For more information about the **mmchfs** command, see the *IBM Spectrum Scale: Command and Programming Reference*.

GPFS error messages for unsuccessful GPFS commands

This topic describes the error messages for unsuccessful GPFS commands.

If message **6027-538** is returned from the **mmcrfs** command, verify that the disk descriptors are specified correctly and that all named disks exist and are online. Issue the **mmlsnsd** command to check the disks.

6027-538
: Error accessing disks.

If the daemon failed while running the command, you will see message **6027-663**. Follow the procedures in "GPFS daemon went down" on page 305.

6027-663
: Lost connection to file system daemon.

If the daemon was not running when you issued the command, you will see message **6027-665**. Follow the procedures in "GPFS daemon will not come up" on page 301.

6027-665
Failed to connect to file system daemon: *errorString*.

When GPFS commands are unsuccessful, the system may display information similar to these error messages:

6027-1627
The following nodes are not aware of the configuration server change: *nodeList*. Do not start GPFS on the preceding nodes until the problem is resolved.

Quorum loss

Each GPFS cluster has a set of quorum nodes explicitly set by the cluster administrator.

These quorum nodes and the selected quorum algorithm determine the availability of file systems owned by the cluster. For more information, see *Quorum* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

When quorum loss or loss of connectivity occurs, any nodes still running GPFS suspend the use of file systems owned by the cluster experiencing the problem. This may result in GPFS access within the suspended file system receiving **ESTALE** errnos. Nodes continuing to function after suspending file system access will start contacting other nodes in the cluster in an attempt to rejoin or reform the quorum. If they succeed in forming a quorum, access to the file system is restarted.

Normally, quorum loss or loss of connectivity occurs if a node goes down or becomes isolated from its peers by a network failure. The expected response is to address the failing condition.

CES configuration issues

The following are the issues that you might encounter while configuring cluster export services in IBM Spectrum Scale.

- Issue: The **mmces** command shows a socket-connection-error.

  Error: Cannot connect to server(localhost), port(`/var/mmfs/mmsysmon/mmsysmonitor.socket`): Connection refused

  Solution: The **mmsysmon-daemon** is not running or is malfunctioning. Submit the **mmsysmoncontrol restart** command to restore the functionality.
- Issue: The **mmlscluster --ces** command does not show any CES IPs, bound to the CES-nodes.

  Solution: Either all CES nodes are unhealthy or no IPs are defined as CES IPs. Try out the following steps to resolve this issue:
  1. Use the **mmces state show -a** to find out the nodes in which the CES service is in the FAILED state. Using the **ssh <nodeName> mmhealth node show** command displays the component that is creating the issue. In some cases, events are created if there are issues with the node health.
  2. Use the **mmces address list** command to list the IPs are defined as CES IPs. You can extend this list by issuing the command **mmces address add --ces-node --ces-ip <ipAddress>**.

Application program errors

When receiving application program errors, there are various courses of action to take.

Follow these steps to help resolve application program errors:

1. Loss of file system access usually appears first as an error received by an application. Such errors are normally encountered when the application tries to access an unmounted file system.

The most common reason for losing access to a single file system is a failure somewhere in the path to a large enough number of disks to jeopardize your data if operation continues. These errors may be reported in the operating system error log on any node because they are logged in the first node to detect the error. Check all error logs for errors.

The **mmlsmount all -L** command can be used to determine the nodes that have successfully mounted a file system.

2. There are several cases where the state of a given disk subsystem will prevent access by GPFS. This will be seen by the application as I/O errors of various types and will be reported in the error logs as **MMFS_SYSTEM_UNMOUNT** or **MMFS_DISKFAIL** records. This state can be found by issuing the **mmlsdisk** command.
3. If allocation of data blocks or files (which quota limits should allow) fails, issue the **mmlsquota** command for the user, group or fileset.

   If filesets are involved, use these steps to determine which fileset was being accessed at the time of the failure:

   a. From the error messages generated, obtain the path name of the file being accessed.
   b. Go to the directory just obtained, and use this **mmlsattr -L** command to obtain the fileset name:

      ```
      mmlsattr -L . | grep "fileset name:"
      ```

      The system produces output similar to:

      ```
      fileset name: myFileset
      ```

   c. Use the **mmlsquota -j** command to check the quota limit of the fileset. For example, using the fileset name found in the previous step, issue this command:

      ```
      mmlsquota -j myFileset -e
      ```

      The system produces output similar to:

      ```
                       Block Limits                        |     File Limits
      Filesystem type      KB  quota limit in_doubt grace | files quota limit in_doubt grace  Remarks
      fs1        FILESET 2152      0     0        0  none |   250     0   250        0  none
      ```

   The **mmlsquota** output is similar when checking the user and group quota. If usage is equal to or approaching the hard limit, or if the grace period has expired, make sure that no quotas are lost by checking *in doubt* values.

   If quotas are exceeded in the *in doubt* category, run the **mmcheckquota** command. For more information, see "The mmcheckquota command" on page 262.

   Note: There is no way to force GPFS nodes to relinquish all their local shares in order to check for lost quotas. This can only be determined by running the **mmcheckquota** command immediately after mounting the file system, and before any allocations are made. In this case, the value *in doubt* is the amount lost.

   To display the latest quota usage information, use the **-e** option on either the **mmlsquota** or the **mmrepquota** commands. Remember that the **mmquotaon** and **mmquotaoff** commands do not enable and disable quota management. These commands merely control enforcement of quota limits. Usage continues to be counted and recorded in the quota files regardless of enforcement.

   Reduce quota usage by deleting or compressing files or moving them out of the file system. Consider increasing quota limit.

GPFS error messages for application program errors

This topic describes the error messages that IBM Spectrum Scale displays for application program errors.

Application program errors can be associated with these GPFS message numbers:

6027-506
program: *loadFile* is already loaded at *address*.

6027-695 [E]
File system is read-only.

Windows issues

The topics that follow apply to Windows Server 2008.

Home and .ssh directory ownership and permissions

This topic describes the issues related to .ssh directory ownership and permissions.

Make sure users own their home directories, which is not normally the case on Windows. They should also own **~/.ssh** and the files it contains. Here is an example of file attributes that work:

```
bash-3.00$ ls -l -d ~
drwx------  1 demyn  Domain Users  0 Dec  5 11:53 /dev/fs/D/Users/demyn
bash-3.00$ ls -l -d ~/.ssh
drwx------  1 demyn  Domain Users  0 Oct 26 13:37 /dev/fs/D/Users/demyn/.ssh
bash-3.00$ ls -l ~/.ssh
total 11
drwx------  1 demyn  Domain Users     0 Oct 26 13:37 .
drwx------  1 demyn  Domain Users     0 Dec  5 11:53 ..
-rw-r--r--  1 demyn  Domain Users   603 Oct 26 13:37 authorized_keys2
-rw-------  1 demyn  Domain Users   672 Oct 26 13:33 id_dsa
-rw-r--r--  1 demyn  Domain Users   603 Oct 26 13:33 id_dsa.pub
-rw-r--r--  1 demyn  Domain Users  2230 Nov 11 07:57 known_hosts
bash-3.00$
```

Problems running as Administrator

You might have problems using SSH when running as the domain **Administrator** user. These issues do not apply to other accounts, even if they are members of the **Administrators** group.

GPFS Windows and SMB2 protocol (CIFS serving)

SMB2 is a version of the Server Message Block (SMB) protocol that was introduced with Windows Vista and Windows Server 2008.

Various enhancements include the following (among others):

- reduced "chattiness" of the protocol
- larger buffer sizes
- faster file transfers
- caching of metadata such as directory content and file properties
- better scalability by increasing the support for number of users, shares, and open files per server

The SMB2 protocol is negotiated between a client and the server during the establishment of the SMB connection, and it becomes active only if both the client and the server are SMB2 capable. If either side is not SMB2 capable, the default SMB (version 1) protocol gets used.

The SMB2 protocol does active metadata caching on the client redirector side, and it relies on Directory Change Notification on the server to invalidate and refresh the client cache. However, GPFS on Windows currently does not support Directory Change Notification. As a result, if SMB2 is used for serving out a IBM Spectrum Scale file system, the SMB2 redirector cache on the client will not see any cache-invalidate operations if the actual metadata is changed, either directly on the server or via another CIFS client. In such a case, the SMB2 client will continue to see its cached version of the directory contents until the redirector cache expires. Therefore, the use of SMB2 protocol for CIFS sharing of GPFS file systems can result in the CIFS clients seeing an inconsistent view of the actual GPFS namespace.

A workaround is to disable the SMB2 protocol on the CIFS server (that is, the GPFS compute node). This will ensure that the SMB2 never gets negotiated for file transfer even if any CIFS client is SMB2 capable.

To disable SMB2 on the GPFS compute node, follow the instructions under the "MORE INFORMATION" section at the Microsoft Support website (support.microsoft.com/kb/974103).

Chapter 16. Upgrade issues

This topic describes the issues that you might encounter while upgrading IBM Spectrum Scale from one version to another.

Upgrading Ubuntu 16.04.x causes Chef client to be upgraded to an unsupported version for the installation toolkit

While upgrading Ubuntu 16.04.x to apply latest updates, the Chef client might get upgraded to an unsupported version. This causes the installation toolkit operation to fail.

After upgrading Ubuntu 16.04.x, use the following workaround on all nodes in the cluster before using the installation toolkit to ensure that an unsupported version of the Chef client is not installed.

Workaround:

1. Check the version of the installed Chef client.

   ```
   dpkg -l chef
   ```

   A sample output is as follows.

   ```
   Desired=Unknown/Install/Remove/Purge/Hold
   | Status=Not/Inst/Conf-files/Unpacked/halF-conf/Half-inst/trig-aWait/Trig-pend
   / Err?=(none)/Reinst-required (Status,Err: uppercase=bad)
   ||/ Name           Version      Architecture Description
   +++-==============-============-============-=================================
   ii  chef           12.3.0-3ubun all          systems integration framework - c
   ```

2. If the Chef client version is not 12.0.3, remove the Chef client and the associated configuration files.

   ```
   apt-get purge chef
   ```

3. Proceed with the installation toolkit operation that you want to perform. The required Chef client is bundled in the installation toolkit and will be automatically be installed as part of the toolkit operation.

File conflict issue while upgrading SLES 12 on IBM Spectrum Scale nodes

While upgrading SLES 12 on IBM Spectrum Scale nodes using the **zypper up** command, you might encounter file conflicts.

This occurs because of the installation of unnecessary, conflicting packages.

Workaround:

Do the SLES 12 upgrade on IBM Spectrum Scale nodes using the **zypper up --no-recommends** command to avoid the installation of conflicting packages.

Related concepts:

"Package conflict on SLES 12 SP1 and SP2 nodes while doing installation, deployment, or upgrade using installation toolkit" on page 291
While doing installation, deployment, or upgrade using the installation toolkit on SLES 12 SP1 and SP2 nodes, you might encounter package conflict issues.

NSD nodes cannot connect to storage after upgrading from SLES 12 SP1 to SP2

After upgrading from SLES 12 SP1 to SP2, NSD nodes might be unable to connect to the storage.

This occurs because of a change in the way regular expressions are evaluated in SLES 12. After this change, glibc-provided regular expressions are used in SLES 12. Therefore, to match an arbitrary string, you must now use ".*" instead of "*".

Workaround:

1. In the blacklist section of the `/etc/multipath.conf` file, replace "*" with ".*".
2. Restart `multipathd.service` by issuing the **`systemctl restart multipathd.service`** command.
3. Verify that LUNs from storage can be detected by issuing the **`multipath -ll`** command.

Chapter 17. Network issues

This topic describes network issues that you might encounter while using IBM Spectrum Scale.

For firewall settings, see the topic *Securing the IBM Spectrum Scale system using firewall* in the *IBM Spectrum Scale: Administration Guide*.

IBM Spectrum Scale failures due to a network failure

For proper functioning, GPFS depends both directly and indirectly on correct network operation.

This dependency is direct because various IBM Spectrum Scale internal messages flow on the network, and may be indirect if the underlying disk technology is dependent on the network. Symptoms included in an indirect failure would be inability to complete I/O or GPFS moving disks to the **down** state.

The problem can also be first detected by the GPFS network communication layer. If network connectivity is lost between nodes or GPFS heart beating services cannot sustain communication to a node, GPFS will declare the node dead and perform recovery procedures. This problem will manifest itself by messages appearing in the GPFS log such as:

```
Mon Jun 25 22:23:36.298 2007: Close connection to 192.168.10.109 c5n109. Attempting reconnect.
Mon Jun 25 22:23:37.300 2007: Connecting to 192.168.10.109 c5n109
Mon Jun 25 22:23:37.398 2007: Close connection to 192.168.10.109 c5n109
Mon Jun 25 22:23:38.338 2007: Recovering nodes: 9.114.132.109
Mon Jun 25 22:23:38.722 2007: Recovered 1 nodes.
```

Nodes mounting file systems owned and served by other clusters may receive error messages similar to this:

```
Mon Jun 25 16:11:16 2007: Close connection to 89.116.94.81 k155n01
Mon Jun 25 16:11:21 2007: Lost membership in cluster remote.cluster. Unmounting file systems.
```

If a sufficient number of nodes fail, GPFS will lose the quorum of nodes, which exhibits itself by messages appearing in the GPFS log, similar to this:

```
Mon Jun 25 11:08:10 2007: Close connection to 179.32.65.4 gpfs2
Mon Jun 25 11:08:10 2007: Lost membership in cluster gpfsxx.kgn.ibm.com. Unmounting file system.
```

When either of these cases occur, perform problem determination on your network connectivity. Failing components could be network hardware such as switches or host bus adapters.

OpenSSH connection delays

OpenSSH can be sensitive to network configuration issues that often do not affect other system components. One common symptom is a substantial delay (20 seconds or more) to establish a connection. When the environment is configured correctly, a command such as **ssh gandalf date** should only take one or two seconds to complete.

If you are using OpenSSH and experiencing an SSH connection delay (and if IPv6 is not supported in your environment), try disabling IPv6 on your Windows nodes and remove or comment out any IPv6 addresses from the **/etc/resolv.conf** file.

Analyze network problems with the mmnetverify command

You can use the **mmnetverify** command to detect network problems and to identify nodes where a network problem exists.

The **mmnetverify** command is useful for detecting network problems and for identifying the type and node location of a network problem. The command can run 16 types of network checks in the areas of connectivity, ports, data, bandwidth, and flooding.

The following examples illustrate some of the uses of this command:

- Before you create a cluster, to verify that all your nodes are ready to be included in a cluster together, you can run the following command:

  ```
  mmnetverify --configuration-file File connectivity -N all
  ```

  This command runs several types of connectivity checks between each node and all the other nodes in the group and reports the results on the console. Because a cluster does not exist yet, you must include a configuration file *File* in which you list all the nodes that you want to test.
- To check for network outages in a cluster, you can run the following command:

  ```
  mmnetverify ping -N all
  ```

  This command runs several types of ping checks between each node and all the other nodes in the cluster and reports the results on the console.
- Before you make a node a quorum node, you can run the following check to verify that other nodes can communicate with the daemon:

  ```
  mmnetverify connectivity port
  ```
- To investigate a possible lag in large-data transfers between two nodes, you can run the following command:

  ```
  mmnetverify data-large -N node2 --target-nodes node3 --verbose
  min-bandwidth Bandwidth
  ```

  This command establishes a TCP connection from `node2` to `node3` and causes the two nodes to exchange a series of large-sized data messages. If the bandwidth falls below the level that is specified, the command generates an error. The output of the command to the console indicates the results of the test.
- To analyze a problem with connectivity between nodes, you can run the following command:

  ```
  mmnetverify connectivity -N all --target-nodes all --verbose
  --log-file File
  ```

  This command runs connectivity checks between each node and all the other nodes in the cluster, one pair at a time, and writes the results of each test to the console and to the specified log file.

Chapter 18. File system issues

Suspect a GPFS file system problem when a file system will not mount or unmount.

You can also suspect a file system problem if a file system unmounts unexpectedly, or you receive an error message indicating that file system activity can no longer continue due to an error, and the file system is being unmounted to preserve its integrity. Record all error messages and log entries that you receive relative to the problem, making sure that you look on all affected nodes for this data.

These are some of the errors encountered with GPFS file systems:

File system fails to mount

There are indications leading you to the conclusion that your file system will not mount and courses of action you can take to correct the problem.

Some of those indications include:

- On performing a manual mount of the file system, you get errors from either the operating system or GPFS.
- If the file system was created with the option of an automatic mount, you will have failure return codes in the GPFS log.
- Your application cannot access the data it needs. Check the GPFS log for messages.
- Return codes or error messages from the **mmmount** command.
- The **mmlsmount** command indicates that the file system is not mounted on certain nodes.

If your file system will not mount, follow these steps:

1. On a quorum node in the cluster that owns the file system, verify that quorum has been achieved. Check the GPFS log to see if an `mmfsd ready` message has been logged, and that no errors were reported on this or other nodes.
2. Verify that a conflicting command is not running. This applies only to the cluster that owns the file system. However, other clusters would be prevented from mounting the file system if a conflicting command is running in the cluster that owns the file system.

For example, a **mount** command may not be issued while the **mmfsck** command is running. The **mount** command may not be issued until the conflicting command completes. Note that interrupting the **mmfsck** command is not a solution because the file system will not be mountable until the command completes. Try again after the conflicting command has completed.

3. Verify that sufficient disks are available to access the file system by issuing the **mmlsdisk** command. GPFS requires a minimum number of disks to find a current copy of the core metadata. If sufficient disks cannot be accessed, the mount will fail. The corrective action is to fix the path to the disk. See "NSD and underlying disk subsystem failures" on page 349.

   Missing disks can also cause GPFS to be unable to find critical metadata structures. The output of the **mmlsdisk** command will show any unavailable disks. If you have not specified metadata replication, the failure of one disk may result in your file system being unable to mount. If you have specified metadata replication, it will require two disks in different failure groups to disable the entire file system. If there are down disks, issue the **mmchdisk start** command to restart them and retry the mount.

   For a remote file system, **mmlsdisk** provides information about the disks of the file system. However **mmchdisk** must be run from the cluster that owns the file system.

   If there are no disks down, you can also look locally for error log reports, and follow the problem determination and repair actions specified in your storage system vendor problem determination guide. If the disk has failed, follow the procedures in "NSD and underlying disk subsystem failures" on page 349.
4. Verify that communication paths to the other nodes are available. The lack of communication paths between all nodes in the cluster may impede contact with the file system manager.
5. Verify that the file system is not already mounted. Issue the **mount** command.
6. Verify that the GPFS daemon on the file system manager is available. Run the **mmlsmgr** command to determine which node is currently assigned as the file system manager. Run a trivial data access command such as an **ls** on the mount point directory. If the command fails, see "GPFS daemon went down" on page 305.
7. Check to see if the mount point directory exists and that there is an entry for the file system in the **/etc/fstab** file (for Linux) or **/etc/filesystems** file (for AIX). The device name for a file system mount point will be listed in column one of the **/etc/fstab** entry or as a **dev=** attribute in the **/etc/filesystems** stanza entry. A corresponding device name must also appear in the **/dev** file system.

   If any of these elements are missing, an update to the configuration information may not have been propagated to this node. Issue the **mmrefresh** command to rebuild the configuration information on the node and reissue the **mmmount** command.

   Do not add GPFS file system information to **/etc/filesystems** (for AIX) or **/etc/fstab** (for Linux) directly. If after running **mmrefresh -f** the file system information is still missing from **/etc/filesystems** (for AIX) or **/etc/fstab** (for Linux), follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.
8. Check the number of file systems that are already mounted. There is a maximum number of 256 mounted file systems for a GPFS cluster. Remote file systems are included in this number.
9. If you issue **mmchfs -V compat**, it enables backwardly-compatible format changes only. Nodes in remote clusters that were able to mount the file system before will still be able to do so.

   If you issue **mmchfs -V full**, it enables all new functions that require different on-disk data structures. Nodes in remote clusters running an older GPFS version will no longer be able to mount the file system. If there are any nodes running an older GPFS version that have the file system mounted at the time this command is issued, the **mmchfs** command will fail. For more information, see the *Completing the upgrade to a new level of IBM Spectrum Scale* section in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

   All nodes that access the file system must be upgraded to the same level of GPFS. Check for the possibility that one or more of the nodes was accidently left out of an effort to upgrade a multi-node

system to a new GPFS release. If you need to return to the earlier level of GPFS, you must re-create the file system from the backup medium and restore the content in order to access it.

10. If DMAPI is enabled for the file system, ensure that a data management application is started and has set a disposition for the mount event. Refer to the *IBM Spectrum Scale: Command and Programming Reference* and the user's guide from your data management vendor. The data management application must be started in the cluster that owns the file system. If the application is not started, other clusters will not be able to mount the file system. Remote mounts of DMAPI managed file systems may take much longer to complete than those not managed by DMAPI.
11. Issue the **mmlsfs -A** command to check whether the automatic mount option has been specified. If automatic mount option is expected, check the GPFS log in the cluster that owns and serves the file system, for progress reports indicating:

    ```
    starting ...
    mounting ...
    mounted ....
    ```
12. If quotas are enabled, check if there was an error while reading quota files. See "MMFS_QUOTA" on page 216.
13. Verify the **maxblocksize** configuration parameter on all clusters involved. If **maxblocksize** is less than the block size of the local or remote file system you are attempting to mount, you will not be able to mount it.
14. If the file system has encryption rules, see "Mount failure for a file system with encryption rules" on page 367.
15. To mount a file system on a remote cluster, ensure that the cluster that owns and serves the file system and the remote cluster have proper authorization in place. The authorization between clusters is set up with the **mmauth** command.

    Authorization errors on AIX are similar to the following:

    ```
    c13c1apv6.gpfs.net:  Failed to open remotefs.
    c13c1apv6.gpfs.net:  Permission denied
    c13c1apv6.gpfs.net:  Cannot mount /dev/remotefs on /gpfs/remotefs: Permission denied
    ```

    Authorization errors on Linux are similar to the following:

    ```
    mount: /dev/remotefs is write-protected, mounting read-only
    mount: cannot mount /dev/remotefs read-only
    mmmount: 6027-1639 Command failed. Examine previous error messages to determine cause.
    ```

    For more information about mounting a file system that is owned and served by another GPFS cluster, see the *Mounting a remote GPFS file system* topic in the *IBM Spectrum Scale: Administration Guide*.

GPFS error messages for file system mount problems

There are error messages specific to file system reading, failure, mounting, and remounting.

6027-419
: Failed to read a file system descriptor.

6027-482 [E]
: Remount failed for device *name*: *errnoDescription*

6027-549
: Failed to open *name*.

6027-580
: Unable to access vital system metadata. Too many disks are unavailable.

6027-645
: Attention: mmcommon getEFOptions *fileSystem* failed. Checking *fileName*.

Error numbers specific to GPFS application calls when a file system mount is not successful

There are specific error numbers for unsuccessful file system mounting.

When a mount of a file system is not successful, GPFS may report these error numbers in the operating system error log or return them to an application:

ENO_QUOTA_INST = 237, No Quota management enabled.
: To enable quotas for the file system, issue the **mmchfs -Q yes** command. To disable quotas for the file system issue the **mmchfs -Q no** command.

Mount failure due to client nodes joining before NSD servers are online

While mounting a file system, specially during automounting, if a client node joins the GPFS cluster and attempts file system access prior to the file system's NSD servers being active, the mount fails. Use **mmchconfig** command to specify the amount of time for GPFS mount requests to wait for an NSD server to join the cluster.

If a client node joins the GPFS cluster and attempts file system access prior to the file system's NSD servers being active, the mount fails. This is especially true when automount is used. This situation can occur during cluster startup, or any time that an NSD server is brought online with client nodes already active and attempting to mount a file system served by the NSD server.

The file system mount failure produces a message similar to this:

```
Mon Jun 25 11:23:34 EST 2007: mmmount: Mounting file systems ...
No such device
Some file system data are inaccessible at this time.
Check error log for additional information.
After correcting the problem, the file system must be unmounted and then
mounted again to restore normal data access.
Failed to open fs1.
No such device
Some file system data are inaccessible at this time.
Cannot mount /dev/fs1 on /fs1: Missing file or filesystem
```

The GPFS log contains information similar to this:

```
Mon Jun 25 11:23:54 2007: Command: mount fs1 32414
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sdcnsd.
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sddnsd.
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sdensd.
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sdgnsd.
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sdhnsd.
Mon Jun 25 11:23:58 2007: Disk failure. Volume fs1. rc = 19. Physical volume sdinsd.
Mon Jun 25 11:23:58 2007: File System fs1 unmounted by the system with return code 19
reason code 0
Mon Jun 25 11:23:58 2007: No such device
Mon Jun 25 11:23:58 2007: File system manager takeover failed.
Mon Jun 25 11:23:58 2007: No such device
Mon Jun 25 11:23:58 2007: Command: err 52: mount fs1 32414
Mon Jun 25 11:23:58 2007: Missing file or filesystem
```

Two **mmchconfig** command options are used to specify the amount of time for GPFS mount requests to wait for an NSD server to join the cluster:

nsdServerWaitTimeForMount
: Specifies the number of seconds to wait for an NSD server to come up at GPFS cluster startup time, after a quorum loss, or after an NSD server failure.

Valid values are between 0 and 1200 seconds. The default is 300. The interval for checking is 10 seconds. If **nsdServerWaitTimeForMount** is 0, **nsdServerWaitTimeWindowOnMount** has no effect.

nsdServerWaitTimeWindowOnMount
Specifies a time window to determine if quorum is to be considered *recently formed*.

Valid values are between 1 and 1200 seconds. The default is 600. If **nsdServerWaitTimeForMount** is 0, **nsdServerWaitTimeWindowOnMount** has no effect.

The GPFS daemon need not be restarted in order to change these values. The scope of these two operands is the GPFS cluster. The **-N** flag can be used to set different values on different nodes. In this case, the settings on the file system manager node take precedence over the settings of nodes trying to access the file system.

When a node rejoins the cluster (after it was expelled, experienced a communications problem, lost quorum, or other reason for which it dropped connection and rejoined), that node resets all the failure times that it knows about. Therefore, when a node rejoins it sees the NSD servers as never having failed. From the node's point of view, it has rejoined the cluster and old failure information is no longer relevant.

GPFS checks the cluster formation criteria first. If that check falls outside the window, GPFS then checks for NSD server fail times being within the window.

File system fails to unmount

There are indications leading you to the conclusion that your file system will not unmount and a course of action to correct the problem.

Those indications include:

- Return codes or error messages indicate the file system will not unmount.
- The **mmlsmount** command indicates that the file system is still mounted on one or more nodes.
- Return codes or error messages from the **mmumount** command.

If your file system will not unmount, follow these steps:

1. If you get an error message similar to:

   ```
   umount: /gpfs1: device is busy
   ```

   the file system will not unmount until all processes are finished accessing it. If **mmfsd** is up, the processes accessing the file system can be determined. See "The lsof command" on page 255. These processes can be killed with the command:

   ```
   lsof filesystem | grep -v COMMAND | awk '{print $2}' | xargs kill -9
   ```

   If **mmfsd** is not operational, the **lsof** command will not be able to determine which processes are still accessing the file system.

   For Linux nodes it is possible to use the **/proc** pseudo file system to determine current file access. For each process currently running on the system, there is a subdirectory **/proc/***pid***/fd**, where *pid* is the numeric process ID number. This subdirectory is populated with symbolic links pointing to the files that this process has open. You can examine the contents of the **fd** subdirectory for all running processes, manually or with the help of a simple script, to identify the processes that have open files in GPFS file systems. Terminating all of these processes may allow the file system to unmount successfully.

| To unmount a CES protocol node, suspend the CES function using the following command:

| ```
| mmces node suspend
| ```

| • Stop the NFS service using the following command:

| `mmces service stop NFS`
| • Stop the SMB service using the following command:
| `mmces service stop SMB`
| • Stop the Object service using the following command:
| `mmces service stop OBJ`

2. Verify that there are no disk media failures.

   Look on the NSD server node for error log entries. Identify any NSD server node that has generated an error log entry. See "Disk media failure" on page 357 for problem determination and repair actions to follow.

3. If the file system *must* be unmounted, you can force the unmount by issuing the **mmumount -f** command:

   Note:

   a. See "File system forced unmount" for the consequences of doing this.
   b. Before forcing the unmount of the file system, issue the **lsof** command and close any files that are open.
   c. On Linux, you might encounter a situation where a GPFS file system cannot be unmounted, even if you issue the **mmumount -f** command. In this case, you must reboot the node to clear the condition. You can also try the system **umount** command before you reboot. For example:

      `umount -f /fileSystem`

4. If a file system that is mounted by a remote cluster needs to be unmounted, you can force the unmount by issuing the command:

   `mmumount fileSystem -f -C RemoteClusterName`

Remote node expelled after remote file system successfully mounted

This problem produces 'node expelled from cluster' messages.

One cause of this condition is when the **subnets** attribute of the **mmchconfig** command has been used to specify subnets to GPFS, and there is an incorrect netmask specification on one or more nodes of the clusters involved in the remote mount. Check to be sure that all netmasks are correct for the network interfaces used for GPFS communication.

File system forced unmount

There are indications that lead you to the conclusion that your file system has been forced to unmount and various courses of action that you can take to correct the problem.

Those indications are:

- Forced unmount messages in the GPFS log.
- Your application no longer has access to data.
- Your application is getting `ESTALE` or `ENOENT` return codes.
- Multiple unsuccessful attempts to appoint a file system manager may cause the cluster manager to unmount the file system everywhere.

  Such situations involve the failure of paths to disk resources from many, if not all, nodes. The underlying problem may be at the disk subsystem level, or lower. The error logs for each node that unsuccessfully attempted to appoint a file system manager will contain records of a file system unmount with an error that are either coded **212**, or that occurred when attempting to assume management of the file system. Note that these errors apply to a specific file system although it is possible that shared disk communication paths will cause the unmount of multiple file systems.
- File system unmounts with an error indicating too many disks are unavailable.

The **mmlsmount -L** command can be used to determine which nodes currently have a given file system mounted.

If your file system has been forced to unmount, follow these steps:

1. With the failure of a single disk, if you have not specified multiple failure groups and replication of metadata, GPFS will not be able to continue because it cannot write logs or other critical metadata. If you have specified multiple failure groups and replication of metadata, the failure of multiple disks in different failure groups will put you in the same position. In either of these situations, GPFS will forcibly unmount the file system. This will be indicated in the error log by records indicating exactly which access failed, with an **MMFS_SYSTEM_UNMOUNT** record indicating the forced unmount. The user response to this is to take the needed actions to restore the disk access and issue the **mmchdisk** command to disks that are shown as down in the information displayed by the **mmlsdisk** command.
2. Internal errors in processing data on a single file system may cause loss of file system access. These errors may clear with the invocation of the **umount** command, followed by a remount of the file system, but they should be reported as problems to the IBM Support Center.
3. If an **MMFS_QUOTA** error log entry containing `Error writing quota file...` is generated, the quota manager continues operation if the next write for the user, group, or fileset is successful. If not, further allocations to the file system will fail. Check the error code in the log and make sure that the disks containing the quota file are accessible. Run the **mmcheckquota** command. For more information, see "The mmcheckquota command" on page 262.

   If the file system must be repaired without quotas:

   a. Disable quota management by issuing the command:

      ```
      mmchfs Device -Q no
      ```

   b. Issue the **mmmount** command for the file system.

   c. Make any necessary repairs and install the backup quota files.

   d. Issue the **mmumount -a** command for the file system.

   e. Restore quota management by issuing the **mmchfs** *Device* **-Q yes** command.

   f. Run the **mmcheckquota** command with the **-u**, **-g**, and **-j** options. For more information, see "The mmcheckquota command" on page 262.

   g. Issue the **mmmount** command for the file system.
4. If errors indicate that too many disks are unavailable, see "Additional failure group considerations."

Additional failure group considerations

GPFS uses *file system descriptor* to be replicated on a subset of the disks as changes to the file system occur, such as adding or deleting disks. To reduce the risk of multiple failure GPFS picks disks to hold the replicas in different failure group.

There is a structure in GPFS called the *file system descriptor* that is initially written to every disk in the file system, but is replicated on a subset of the disks as changes to the file system occur, such as adding or deleting disks. Based on the number of failure groups and disks, GPFS creates between one and five replicas of the descriptor:

- If there are at least five different failure groups, five replicas are created.
- If there are at least three different disks, three replicas are created.
- If there are only one or two disks, a replica is created on each disk.

Once it is decided how many replicas to create, GPFS picks disks to hold the replicas, so that all replicas will be in different failure groups, if possible, to reduce the risk of multiple failures. In picking replica locations, the current state of the disks is taken into account. Stopped or suspended disks are avoided. Similarly, when a failed disk is brought back online, GPFS may modify the subset to rebalance the file system descriptors across the failure groups. The subset can be found by issuing the **mmlsdisk -L** command.

GPFS requires a majority of the replicas on the subset of disks to remain available to sustain file system operations:

- If there are at least five different failure groups, GPFS will be able to tolerate a loss of two of the five groups. If disks out of three different failure groups are lost, the file system descriptor may become inaccessible due to the loss of the majority of the replicas.
- If there are at least three different failure groups, GPFS will be able to tolerate a loss of one of the three groups. If disks out of two different failure groups are lost, the file system descriptor may become inaccessible due to the loss of the majority of the replicas.
- If there are fewer than three failure groups, a loss of one failure group may make the descriptor inaccessible.

  If the subset consists of three disks and there are only two failure groups, one failure group must have two disks and the other failure group has one. In a scenario that causes one entire failure group to disappear all at once, if the half of the disks that are unavailable contain the single disk that is part of the subset, everything stays up. The file system descriptor is moved to a new subset by updating the remaining two copies and writing the update to a new disk added to the subset. But if the downed failure group contains a majority of the subset, the file system descriptor cannot be updated and the file system has to be force unmounted.

  Introducing a third failure group consisting of a single disk that is used solely for the purpose of maintaining a copy of the file system descriptor can help prevent such a scenario. You can designate this disk by using the **descOnly** designation for disk usage on the disk descriptor. For more information on disk replication, see the *NSD creation considerations* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* and the *Data mirroring and replication* topic in the *IBM Spectrum Scale: Administration Guide*.

GPFS error messages for file system forced unmount problems

There are many error messages for file system forced unmount problems due to unavailable disk space.

Indications there are not enough disks available:

6027-418
Inconsistent file system quorum. readQuorum=*value* writeQuorum=*value* quorumSize=*value*.

6027-419
Failed to read a file system descriptor.

Indications the file system has been forced to unmount:

6027-473 [X]
File System *fileSystem* unmounted by the system with return code *value* reason code *value*

6027-474 [X]
Recovery Log I/O failed, unmounting file system *fileSystem*

Error numbers specific to GPFS application calls when a file system has been forced to unmount

There are error numbers to indicate that a file system is forced to unmount for GPFS application calls.

When a file system has been forced to unmount, GPFS may report these error numbers in the operating system error log or return them to an application:

EPANIC = 666, A file system has been forcibly unmounted because of an error. Most likely due to the failure of one or more disks containing the last copy of metadata.
See "Operating system error logs" on page 214 for details.

EALL_UNAVAIL = 218, A replicated read or write failed because none of the replicas were available.
Multiple disks in multiple failure groups are unavailable. Follow the procedures in Chapter 19, "Disk issues," on page 349 for unavailable disks.

Automount file system will not mount

The automount fails to mount the file system and the courses of action that you can take to correct the problem.

If an automount fails when you **cd** into the mount point directory, first check that the file system in question is of automount type. Use the **mmlsfs -A** command for local file systems. Use the **mmremotefs show** command for remote file systems.

Steps to follow if automount fails to mount on Linux

There are course of actions that you can take if the automount fails to mount on Linux system.

On Linux, perform these steps:

1. Verify that the GPFS file system mount point is actually a symbolic link to a directory in the **automountdir** directory. If **automountdir=/gpfs/automountdir** then the mount point **/gpfs/gpfs66** would be a symbolic link to **/gpfs/automountdir/gpfs66**.
   a. First, verify that GPFS is up and running.
   b. Use the **mmlsconfig** command to verify the **automountdir** directory. The default **automountdir** is named **/gpfs/automountdir**. If the GPFS file system mount point is not a symbolic link to the GPFS **automountdir** directory, then accessing the mount point will not cause the automounter to mount the file system.
   c. If the command **/bin/ls -ld** of the mount point shows a directory, then run the command **mmrefresh -f**. If the directory is empty, the command **mmrefresh -f** will remove the directory and create a symbolic link. If the directory is not empty, you need to move or remove the files contained in that directory, or change the mount point of the file system. For a local file system, use the **mmchfs** command. For a remote file system, use the **mmremotefs** command.
   d. Once the mount point directory is empty, run the **mmrefresh -f** command.
2. Verify that the **autofs** mount has been established. Issue this command:

   ```
   mount | grep automount
   ```

   The output must be similar to this:

   ```
   automount(pid20331) on /gpfs/automountdir type autofs (rw,fd=5,pgrp=20331,minproto=2,maxproto=3)
   ```

   For Red Hat Enterprise Linux 5, verify the following line is in the default master map file (/etc/auto.master):

   ```
   /gpfs/automountdir program:/usr/lpp/mmfs/bin/mmdynamicmap
   ```

   For example, issue:

   ```
   grep mmdynamicmap /etc/auto.master
   ```

   Output should be similar to this:

   ```
   /gpfs/automountdir program:/usr/lpp/mmfs/bin/mmdynamicmap
   ```

   This is an **autofs** program map, and there will be a single mount entry for all GPFS automounted file systems. The symbolic link points to this directory, and access through the symbolic link triggers the mounting of the target GPFS file system. To create this GPFS **autofs** mount, issue the **mmcommon startAutomounter** command, or stop and restart GPFS using the **mmshutdown** and **mmstartup** commands.
3. Verify that the automount daemon is running. Issue this command:

```
ps -ef | grep automount
```

The output must be similar to this:

```
root 5116 1 0 Jun25 pts/0 00:00:00 /usr/sbin/automount /gpfs/automountdir program
    /usr/lpp/mmfs/bin/mmdynamicmap
```

For Red Hat Enterprise Linux 5, verify that the **autofs** daemon is running. Issue this command:

```
ps -ef | grep automount
```

The output must be similar to this:

```
root 22646 1 0 01:21 ? 00:00:02 automount
```

To start the automount daemon, issue the **mmcommon startAutomounter** command, or stop and restart GPFS using the **mmshutdown** and **mmstartup** commands.

Note: If **automountdir** is mounted (as in step 2) and the **mmcommon startAutomounter** command is not able to bring up the automount daemon, manually **umount** the **automountdir** before issuing the **mmcommon startAutomounter** again.

4. Verify that the mount command was issued to GPFS by examining the GPFS log. You should see something like this:

   ```
   Mon Jun 25 11:33:03 2004: Command: mount gpfsx2.kgn.ibm.com:gpfs55 5182
   ```

5. Examine **/var/log/messages** for **autofs** error messages. The following is an example of what you might see if the remote file system name does not exist.

   ```
   Jun 25 11:33:03 linux automount[20331]: attempting to mount entry /gpfs/automountdir/gpfs55
   Jun 25 11:33:04 linux automount[28911]: >> Failed to open gpfs55.
   Jun 25 11:33:04 linux automount[28911]: >> No such device
   Jun 25 11:33:04 linux automount[28911]: >> mount: fs type gpfs not supported by kernel
   Jun 25 11:33:04 linux automount[28911]: mount(generic): failed to mount /dev/gpfs55 (type gpfs)
       on /gpfs/automountdir/gpfs55
   ```

6. After you have established that GPFS has received a mount request from **autofs** (Step 4) and that mount request failed (Step 5), issue a mount command for the GPFS file system and follow the directions in "File system fails to mount" on page 317.

Steps to follow if automount fails to mount on AIX

There are course of actions that you can take if the automount fails to mount on AIX server.

On AIX, perform these steps:

1. First, verify that GPFS is up and running.
2. Verify that GPFS has established **autofs** mounts for each automount file system. Issue the following command:

   ```
   mount | grep autofs
   ```

   The output is similar to this:

   ```
   /var/mmfs/gen/mmDirectMap /gpfs/gpfs55 autofs Jun 25 15:03 ignore
   /var/mmfs/gen/mmDirectMap /gpfs/gpfs88 autofs Jun 25 15:03 ignore
   ```

   These are direct mount **autofs** mount entries. Each GPFS automount file system will have an **autofs** mount entry. These **autofs** direct mounts allow GPFS to mount on the GPFS mount point. To create any missing GPFS **autofs** mounts, issue the **mmcommon startAutomounter** command, or stop and restart GPFS using the **mmshutdown** and **mmstartup** commands.
3. Verify that the **autofs** daemon is running. Issue this command:

   ```
   ps -ef | grep automount
   ```

   Output is similar to this:

   ```
   root 9820 4240 0 15:02:50 - 0:00 /usr/sbin/automountd
   ```

To start the automount daemon, issue the **mmcommon startAutomounter** command, or stop and restart GPFS using the **mmshutdown** and **mmstartup** commands.

4. Verify that the mount command was issued to GPFS by examining the GPFS log. You should see something like this:

   ```
   Mon Jun 25 11:33:03 2007: Command: mount gpfsx2.kgn.ibm.com:gpfs55 5182
   ```

5. Since the **autofs** daemon logs status using **syslogd**, examine the **syslogd** log file for status information from **automountd**. Here is an example of a failed automount request:

   ```
   Jun 25 15:55:25 gpfsa1 automountd [9820 ] :mount of /gpfs/gpfs55:status 13
   ```

6. After you have established that GPFS has received a mount request from **autofs** (Step 4) and that mount request failed (Step 5), issue a mount command for the GPFS file system and follow the directions in "File system fails to mount" on page 317.
7. If automount fails for a non-GPFS file system and you are using file **/etc/auto.master**, use file **/etc/auto_master** instead. Add the entries from **/etc/auto.master** to **/etc/auto_master** and restart the automount daemon.

Remote file system will not mount

The remote file system mounting failure reasons and the course of action that you can take to resolve the issue.

When a remote file system does not mount, the problem might be with how the file system was defined to both the local and remote nodes, or the communication paths between them. Review the *Mounting a file system owned and served by another GPFS cluster* topic in the *IBM Spectrum Scale: Administration Guide* to ensure that your setup is correct.

These are some of the errors encountered when mounting remote file systems:

- "Remote file system I/O fails with the "Function not implemented" error message when UID mapping is enabled"
- "Remote file system will not mount due to differing GPFS cluster security configurations" on page 328
- "Cannot resolve contact node address" on page 328
- "The remote cluster name does not match the cluster name supplied by the mmremotecluster command" on page 329
- "Contact nodes down or GPFS down on contact nodes" on page 329
- "GPFS is not running on the local node" on page 330
- "The NSD disk does not have an NSD server specified and the mounting cluster does not have direct access to the disks" on page 330
- "The cipherList option has not been set properly" on page 330
- "Remote mounts fail with the "permission denied" error message" on page 331

Remote file system I/O fails with the "Function not implemented" error message when UID mapping is enabled

There are error messages when remote file system has an I/O failure and the course of action that you can take to correct this issue.

When user ID (UID) mapping in a multi-cluster environment is enabled, certain kinds of mapping infrastructure configuration problems might result in I/O requests on a remote file system failing:

```
ls -l /fs1/testfile
ls: /fs1/testfile: Function not implemented
```

To troubleshoot this error, verify the following configuration details:

1. That **/var/mmfs/etc/mmuid2name** and **/var/mmfs/etc/mmname2uid** helper scripts are present and executable on all nodes in the local cluster and on all quorum nodes in the file system home cluster, along with any data files needed by the helper scripts.
2. That UID mapping is enabled in both local cluster and remote file system home cluster configuration by issuing the **mmlsconfig enableUIDremap** command.
3. That UID mapping helper scripts are working correctly.

For more information about configuring UID mapping, see the IBM white paper entitled *UID Mapping for GPFS in a Multi-cluster Environment* in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/SSFKCN/com.ibm.cluster.gpfs.doc/gpfs_uid/uid_gpfs.html).

Remote file system will not mount due to differing GPFS cluster security configurations

There are indications leading you to the conclusion that the remote file system will not mount and courses of action you can take to correct the problem.

A mount command fails with a message similar to this:

```
Cannot mount gpfsxx2.ibm.com:gpfs66: Host is down.
```

The GPFS log on the cluster issuing the mount command should have entries similar to these:

```
There is more information in the log file /var/adm/ras/mmfs.log.latest
Mon Jun 25 16:39:27 2007: Waiting to join remote cluster gpfsxx2.ibm.com
Mon Jun 25 16:39:27 2007: Command: mount gpfsxx2.ibm.com:gpfs66 30291
Mon Jun 25 16:39:27 2007: The administrator of 199.13.68.12 gpfslx2 requires
secure connections.  Contact the administrator to obtain the target clusters
key and register the key using "mmremotecluster update".
Mon Jun 25 16:39:27 2007: A node join was rejected.  This could be due to
incompatible daemon versions, failure to find the node
in the configuration database, or no configuration manager found.
Mon Jun 25 16:39:27 2007: Failed to join remote cluster gpfsxx2.ibm.com
Mon Jun 25 16:39:27 2007: Command err 693: mount gpfsxx2.ibm.com:gpfs66 30291
```

The GPFS log file on the cluster that owns and serves the file system will have an entry indicating the problem as well, similar to this:

```
Mon Jun 25 16:32:21 2007: Kill accepted connection from 199.13.68.12 because security is required, err 74
```

To resolve this problem, contact the administrator of the cluster that owns and serves the file system to obtain the key and register the key using **mmremotecluster** command.

The SHA digest field of the **mmauth show** and **mmremotecluster** commands may be used to determine if there is a key mismatch, and on which cluster the key should be updated. For more information on the SHA digest, see "The SHA digest" on page 267.

Cannot resolve contact node address

There are error messages which are displayed if the contact node address does not get resolved and the courses of action you can take to correct the problem.

The following error may occur if the contact nodes for **gpfsyy2.ibm.com** could not be resolved. You would expect to see this if your DNS server was down, or the contact address has been deleted.

```
Mon Jun 25 15:24:14 2007: Command: mount gpfsyy2.ibm.com:gpfs14 20124
Mon Jun 25 15:24:14 2007: Host 'gpfs123.ibm.com' in gpfsyy2.ibm.com is not valid.
Mon Jun 25 15:24:14 2007: Command err 2: mount gpfsyy2.ibm.com:gpfs14 20124
```

To resolve the problem, correct the contact list and try the mount again.

The remote cluster name does not match the cluster name supplied by the mmremotecluster command

There are error messages that gets displayed if the remote cluster name does not match with the cluster name , provided by mmremotecluster command, and the courses of action you can take to correct the problem.

A mount command fails with a message similar to this:

```
Cannot mount gpfslx2:gpfs66: Network is unreachable
```

and the GPFS log contains message similar to this:

```
Mon Jun 25 12:47:18 2007: Waiting to join remote cluster gpfslx2
Mon Jun 25 12:47:18 2007: Command: mount gpfslx2:gpfs66 27226
Mon Jun 25 12:47:18 2007: Failed to join remote cluster gpfslx2
Mon Jun 25 12:47:18 2007: Command err 719: mount gpfslx2:gpfs66 27226
```

Perform these steps:

1. Verify that the remote cluster name reported by the **mmremotefs show** command is the same name as reported by the **mmlscluster** command from one of the contact nodes.
2. Verify the list of contact nodes against the list of nodes as shown by the **mmlscluster** command from the remote cluster.

In this example, the correct cluster name is **gpfslx2.ibm.com** and not **gpfslx2**

```
mmlscluster
```

Output is similar to this:

```
GPFS cluster information
========================
  GPFS cluster name:         gpfslx2.ibm.com
  GPFS cluster id:           649437685184692490
  GPFS UID domain:           gpfslx2.ibm.com
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           server-based

GPFS cluster configuration servers:
-----------------------------------
  Primary server:    gpfslx2.ibm.com
  Secondary server:  (none)

 Node  Daemon node name   IP address      Admin node name       Designation
--------------------------------------------------------------------------
   1   gpfslx2            198.117.68.68   gpfslx2.ibm.com       quorum
```

Contact nodes down or GPFS down on contact nodes

There are error messages that gets displayed if the contact nodes are down or the GPFS is down on the contact nodes, and the courses of action you can take to correct the problem.

A mount command fails with a message similar to this:

```
GPFS: 6027-510 Cannot mount /dev/gpfs22 on /gpfs22: A remote host did not respond
    within the timeout period.
```

The GPFS log will have entries similar to this:

```
Mon Jun 25 13:11:14 2007: Command: mount gpfslx22:gpfs22 19004
Mon Jun 25 13:11:14 2007: Waiting to join remote cluster gpfslx22
Mon Jun 25 13:11:15 2007: Connecting to 199.13.68.4 gpfslx22
Mon Jun 25 13:16:36 2007: Failed to join remote cluster gpfslx22
Mon Jun 25 13:16:36 2007: Command err 78: mount gpfslx22:gpfs22 19004
```

To resolve the problem, use the **mmremotecluster show** command and verify that the cluster name matches the remote cluster and the contact nodes are valid nodes in the remote cluster. Verify that GPFS is active on the contact nodes in the remote cluster. Another way to resolve this problem is to change the contact nodes using the **mmremotecluster update** command.

GPFS is not running on the local node

There are error messages that gets displayed if the GPFS does not run on the local nodes, and the courses of action that you can take to correct the problem.

A mount command fails with a message similar to this:

```
mount: fs type gpfs not supported by kernel
```

Follow your procedures for starting GPFS on the local node.

The NSD disk does not have an NSD server specified and the mounting cluster does not have direct access to the disks

There are error messages that gets displayed if the file system mounting gets failed, and the courses of action that you can take to correct the problem.

A file system mount fails with a message similar to this:

```
Failed to open gpfs66.
No such device
mount: Stale NFS file handle
Some file system data are inaccessible at this time.
Check error log for additional information.
Cannot mount gpfslx2.ibm.com:gpfs66: Stale NFS file handle
```

The GPFS log will contain information similar to this:

```
Mon Jun 25 14:10:46 2007: Command: mount gpfslx2.ibm.com:gpfs66 28147
Mon Jun 25 14:10:47 2007: Waiting to join remote cluster gpfslx2.ibm.com
Mon Jun 25 14:10:47 2007: Connecting to 199.13.68.4 gpfslx2
Mon Jun 25 14:10:47 2007: Connected to 199.13.68.4 gpfslx2
Mon Jun 25 14:10:47 2007: Joined remote cluster gpfslx2.ibm.com
Mon Jun 25 14:10:48 2007: Global NSD disk, gpfs1nsd, not found.
Mon Jun 25 14:10:48 2007: Disk failure.  Volume gpfs66. rc = 19. Physical volume gpfs1nsd.
Mon Jun 25 14:10:48 2007: File System gpfs66 unmounted by the system with return code 19 reason code 0
Mon Jun 25 14:10:48 2007: No such device
Mon Jun 25 14:10:48 2007: Command err 666: mount gpfslx2.ibm.com:gpfs66 28147
```

To resolve the problem, the cluster that owns and serves the file system must define one or more NSD servers.

The cipherList option has not been set properly

There remote mount failure,due to invalid value of cipherList, leads the error messages and the course of actions that you can take to resolve the issue.

Another reason for remote mount to fail is if **cipherList** is not set to a valid value. A mount command would fail with messages similar to this:

```
6027-510 Cannot mount /dev/dqfs1 on /dqfs1: A remote host is not available.
```

The GPFS log would contain messages similar to this:

```
Wed Jul 18 16:11:20.496 2007: Command: mount remote.cluster:fs3 655494
Wed Jul 18 16:11:20.497 2007: Waiting to join remote cluster remote.cluster
Wed Jul 18 16:11:20.997 2007: Remote mounts are not enabled within this cluster. \
See the Advanced Administration Guide for instructions. In particular ensure keys have been \
generated and a cipherlist has been set.
```

```
Wed Jul 18 16:11:20.998 2007: A node join was rejected.  This could be due to
incompatible daemon versions, failure to find the node
in the configuration database, or no configuration manager found.
Wed Jul 18 16:11:20.999 2007: Failed to join remote cluster remote.cluster
Wed Jul 18 16:11:20.998 2007: Command: err 693: mount remote.cluster:fs3 655494
Wed Jul 18 16:11:20.999 2007: Message failed because the destination node refused the connection.
```

The **mmchconfig cipherlist=AUTHONLY** command must be run on both the cluster that owns and controls the file system, and the cluster that is attempting to mount the file system.

Remote mounts fail with the "permission denied" error message

There are many reasons why remote mounts can fail with a "permission denied" error message.

Follow these steps to resolve permission denied problems:

1. Check with the remote cluster's administrator to make sure that the proper keys are in place. The **mmauth show** command on both clusters will help with this.
2. Check that the grant access for the remote mounts has been given on the remote cluster with the **mmauth grant** command. Use the **mmauth show** command from the remote cluster to verify this.
3. Check that the file system access permission is the same on both clusters using the **mmauth show** command and the **mmremotefs show** command. If a remote cluster is only allowed to do a read-only mount (see the **mmauth show** command), the remote nodes must specify **-o ro** on their mount requests (see the **mmremotefs show** command). If you try to do remote mounts with read/write (**rw**) access for remote mounts that have read-only (**ro**) access, you will get a "permission denied" error.

For detailed information about the **mmauth** command and the **mmremotefs** command, see the *mmauth command* and the *mmremotefs command* pages in the *IBM Spectrum Scale: Command and Programming Reference*.

Unable to determine whether a file system is mounted

Certain GPFS file system commands cannot be performed when the file system in question is mounted.

In certain failure situations, GPFS cannot determine whether the file system in question is mounted or not, and so cannot perform the requested command. In such cases, message **6027-1996** (Command was unable to determine whether file system *fileSystem* is mounted) is issued.

If you encounter this message, perform problem determination, resolve the problem, and reissue the command. If you cannot determine or resolve the problem, you may be able to successfully run the command by first shutting down the GPFS daemon on all nodes of the cluster (using **mmshutdown -a**), thus ensuring that the file system is not mounted.

GPFS error messages for file system mount status

The GPFS file system commands displays error message when they are unable to determine if the file system in question is mounted.

6027-1996
Command was unable to determine whether file system *fileSystem* is mounted.

Multiple file system manager failures

The correct operation of GPFS requires that one node per file system function as the file system manager at all times. This instance of GPFS has additional responsibilities for coordinating usage of the file system.

When the file system manager node fails, another file system manager is appointed in a manner that is not visible to applications except for the time required to switch over.

There are situations where it may be impossible to appoint a file system manager. Such situations involve the failure of paths to disk resources from many, if not all, nodes. In this event, the cluster manager nominates several host names to successively try to become the file system manager. If none succeed, the cluster manager unmounts the file system everywhere. See "NSD and underlying disk subsystem failures" on page 349.

The required action here is to address the underlying condition that caused the forced unmounts and then remount the file system. In most cases, this means correcting the path to the disks required by GPFS. If NSD disk servers are being used, the most common failure is the loss of access through the communications network. If SAN access is being used to all disks, the most common failure is the loss of connectivity through the SAN.

GPFS error messages for multiple file system manager failures

Certain GPFS error messages are displayed for multiple file system manager failures.

The inability to successfully appoint a file system manager after multiple attempts can be associated with both the error messages listed in "File system forced unmount" on page 322, as well as these additional messages:

- When a forced unmount occurred on all nodes:

  6027-635 [E]
  The current file system manager failed and no new manager will be appointed.
- If message **6027-636** is displayed, it means that there may be a disk failure. See "NSD and underlying disk subsystem failures" on page 349 for NSD problem determination and repair procedures.

  6027-636 [E]
  Disk marked as stopped or offline.
- Message **6027-632** is the last message in this series of messages. See the accompanying messages:

  6027-632
  Failed to appoint new manager for *fileSystem*.
- Message **6027-631** occurs on each attempt to appoint a new manager (see the messages on the referenced node for the specific reason as to why it failed):

  6027-631
  Failed to appoint node *nodeName* as manager for *fileSystem*.
- Message **6027-638** indicates which node had the original error (probably the original file system manager node):

  6027-638 [E]
  File system *fileSystem* unmounted by node *nodeName*

Error numbers specific to GPFS application calls when file system manager appointment fails

Certain error numbers and messages are displayed when the file system manager appointment fails .

When the appointment of a file system manager is unsuccessful after multiple attempts, GPFS may report these error numbers in error logs, or return them to an application:

ENO_MGR = 212, The current file system manager failed and no new manager could be appointed.
This usually occurs when a large number of disks are unavailable or when there has been a major network failure. Run **mmlsdisk** to determine whether disks have failed and take corrective action if they have by issuing the **mmchdisk** command.

Discrepancy between GPFS configuration data and the on-disk data for a file system

There is an indication leading you to the conclusion that there may be a discrepancy between the GPFS configuration data and the on-disk data for a file system.

You issue a disk command (for example, **mmadddisk**, **mmdeldisk**, or **mmrpldisk**) and receive the message:

6027-1290
GPFS configuration data for file system *fileSystem* may not be in agreement with the on-disk data for the file system. Issue the command:

```
mmcommon recoverfs fileSystem
```

Before a disk is added to or removed from a file system, a check is made that the GPFS configuration data for the file system is in agreement with the on-disk data for the file system. The preceding message is issued if this check was not successful. This may occur if an earlier GPFS disk command was unable to complete successfully for some reason. Issue the **mmcommon recoverfs** command to bring the GPFS configuration data into agreement with the on-disk data for the file system.

If running **mmcommon recoverfs** does not resolve the problem, follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.

Errors associated with storage pools, filesets and policies

There are certain error messages associated with the storage pools, filesets and policies.

When an error is suspected while working with storage pools, policies and filesets, check the relevant section in the *IBM Spectrum Scale: Administration Guide* to ensure that your setup is correct.

When you are sure that your setup is correct, see if your problem falls into one of these categories:

A NO_SPACE error occurs when a file system is known to have adequate free space

The GPFS commands display a NO_SPACE error even if a file system has free space and the course of actions that you can take to correct this issue.

A **ENOSPC** (**NO_SPACE**) message can be returned even if a file system has remaining space. The **NO_SPACE** error might occur even if the **df** command shows that the file system is not full.

The user might have a policy that writes data into a specific storage pool. When the user tries to create a file in that storage pool, it returns the **ENOSPC** error if the storage pool is full. The user next issues the

df command, which indicates that the file system is not full, because the problem is limited to the one storage pool in the user's policy. In order to see if a particular storage pool is full, the user must issue the **mmdf** command.

The following is a sample scenario:

1. The user has a policy rule that says files whose name contains the word 'tmp' should be put into storage pool **sp1** in the file system **fs1**. This command displays the rule:

   ```
   mmlspolicy fs1 -L
   ```

   The system displays an output similar to this:

   ```
   /* This is a policy for GPFS file system fs1 */

   /* File Placement Rules */
   RULE SET POOL 'sp1' WHERE name like '%tmp%'
   RULE 'default' SET POOL 'system'
   /* End of Policy */
   ```

2. The user moves a file from the **/tmp** directory to **fs1** that has the word 'tmp' in the file name, meaning data of **tmpfile** should be placed in storage pool **sp1**:

   ```
   mv /tmp/tmpfile /fs1/
   ```

   The system produces output similar to this:

   ```
   mv: writing `/fs1/tmpfile': No space left on device
   ```

   This is an out-of-space error.

3. This command shows storage information for the file system:

   ```
    df |grep fs1
   ```

   The system produces output similar to this:

   ```
    /dev/fs1             280190976 140350976 139840000  51% /fs1
   ```

   This output indicates that the file system is only 51% full.

4. To query the storage usage for an individual storage pool, the user must issue the **mmdf** command.

   ```
   mmdf fs1
   ```

   The system produces output similar to this:

   ```
   disk                disk size  failure holds    holds              free KB             free KB
   name                    in KB    group metadata data        in full blocks        in fragments
   --------------- ------------- -------- -------- ----- -------------------- -------------------
   Disks in storage pool: system
   gpfs1nsd            140095488     4001 yes      yes       139840000 (100%)         19936 ( 0%)
                   -------------                         -------------------- -------------------
   (pool total)        140095488                               139840000 (100%)         19936 ( 0%)

   Disks in storage pool: sp1
   gpfs2nsd            140095488     4001 no       yes               0s (  0%)           248 ( 0%)
                   -------------                         -------------------- -------------------
   (pool total)        140095488                                     0 (  0%)           248 ( 0%)

                   =============                         ==================== ===================
   (data)              280190976                               139840000 ( 50%)         20184 ( 0%)
   (metadata)          140095488                               139840000 (100%)         19936 ( 0%)
                   =============                         ==================== ===================
   (total)             280190976                               139840000 ( 50%)         20184 ( 0%)

   Inode Information
   -----------------
   ```

```
Number of used inodes: 74
Number of free inodes: 137142
Number of allocated inodes: 137216
Maximum number of inodes: 150016
```

In this case, the user sees that storage pool **sp1** has 0% free space left and that is the reason for the **NO_SPACE** error message.

5. To resolve the problem, the user must change the placement policy file to avoid putting data in a full storage pool, delete some files in storage pool **sp1**, or add more space to the storage pool.

Negative values occur in the 'predicted pool utilizations', when some files are 'ill-placed'

A scenario where an ill-placed files may cause GPFS to produce a 'Predicted Pool Utilization' of a negative value and the course of action that you can take to resolve this issue.

This is a hypothetical situation where ill-placed files can cause GPFS to produce a 'Predicted Pool Utilization' of a negative value.

Suppose that 2 GB of data from a 5 GB file named **abc**, that is supposed to be in the **system** storage pool, are actually located in another pool. This 2 GB of data is said to be 'ill-placed'. Also, suppose that 3 GB of this file are in the **system** storage pool, and no other file is assigned to the **system** storage pool.

If you run the **mmapplypolicy** command to schedule file **abc** to be moved from the **system** storage pool to a storage pool named **YYY**, the **mmapplypolicy** command does the following:

1. Starts with the 'Current pool utilization' for the **system** storage pool, which is 3 GB.
2. Subtracts 5 GB, the size of file **abc**.
3. Arrives at a 'Predicted Pool Utilization' of negative 2 GB.

The **mmapplypolicy** command does not know how much of an 'ill-placed' file is currently in the wrong storage pool and how much is in the correct storage pool.

When there are ill-placed files in the **system** storage pool, the 'Predicted Pool Utilization' can be any positive or negative value. The positive value can be capped by the **LIMIT** clause of the **MIGRATE** rule. The 'Current Pool Utilizations' should always be between 0% and 100%.

Policies - usage errors

Certain misunderstandings that may be encountered while using policies and the suggestions to overcome such mistakes.

The following are common mistakes and misunderstandings encountered when dealing with policies:

- You are advised to test your policy rules using the **mmapplypolicy** command with the **-I test** option.

  Also consider specifying a test-subdirectory within your file system. Do not apply a policy to an entire file system of vital files until you are confident that the rules correctly express your intentions. Even then, you are advised to do a sample run with the **mmapplypolicy -I test** command using the option **-L 3** or higher, to better understand which files are selected as candidates, and which candidates are chosen. The **-L** flag of the **mmapplypolicy** command can be used to check a policy before it is applied. For examples and more information on this flag, see "The mmapplypolicy -L command" on page 256.
- There is a 1 MB limit on the total size of the policy file installed in GPFS.
- Ensure that all clocks on all nodes of the GPFS cluster are synchronized. Depending on the policies in effect, variations in the clock times can cause unexpected behavior.

  The **mmapplypolicy** command uses the time on the node on which it is run as the current time. Policy rules may refer to a file's last access time or modification time, which is set by the node which last accessed or modified the file. If the clocks are not synchronized, files may be treated as older or

younger than their actual age, and this could cause files to be migrated or deleted prematurely, or not at all. A suggested solution is to use NTP to keep the clocks synchronized on all nodes in the cluster.

- The rules of a policy file are evaluated in order. A new file is assigned to the storage pool of the first rule that it matches. If the file fails to match any rule, the file creation fails with an **EINVAL** error code. A suggested solution is to put a **DEFAULT** clause as the last entry of the policy file.
- When a policy file is installed, GPFS verifies that the named storage pools exist. However, GPFS allows an administrator to delete pools that are mentioned in the policy file. This allows more freedom for recovery from hardware errors. Consequently, the administrator must be careful when deleting storage pools referenced in the policy.

Errors encountered with policies

The analysis of those errors which may be encountered while dealing with the policies.

These are errors encountered with policies and how to analyze them:

- Policy file never finishes, appears to be looping.

  The **mmapplypolicy** command runs by making two passes over the file system - one over the inodes and one over the directory structure. The policy rules are applied to each file to determine a list of candidate files. The list is sorted by the weighting specified in the rules, then applied to the file system. No file is ever moved more than once. However, due to the quantity of data involved, this operation may take a long time and appear to be hung or looping.

  The time required to run **mmapplypolicy** is a function of the number of files in the file system, the current load on the file system, and on the node in which **mmapplypolicy** is run. If this function appears to not finish, you may need to reduce the load on the file system or run **mmapplypolicy** on a less loaded node in the cluster.

- Initial file placement is not correct.

  The placement rules specify a single pool for initial placement. The first rule that matches the file's attributes selects the initial pool. If that pool is incorrect, then the placement rules must be updated to select a different pool. You may see current placement rules by running **mmlspolicy -L**. For existing files, the file can be moved to its desired pool using the **mmrestripefile** or **mmchattr** commands.

  For examples and more information on **mmlspolicy -L**, see "The mmapplypolicy -L command" on page 256.

- Data migration, deletion or exclusion not working properly.

  The **mmapplypolicy** command selects a list of candidate files to be migrated or deleted. The list is sorted by the weighting factor specified in the rules, then applied to a sufficient number of files on the candidate list to achieve the utilization thresholds specified by the pools. The actual migration and deletion are done in parallel. The following are the possibilities for an incorrect operation:

  – The file was not selected as a candidate for the expected rule. Each file is selected as a candidate for only the first rule that matched its attributes. If the matched rule specifies an invalid storage pool, the file is not moved. The **-L 4** option on **mmapplypolicy** displays the details for candidate selection and file exclusion.

  – The file was a candidate, but was not operated on. Only the candidates necessary to achieve the desired pool utilizations are migrated. Using the **-L 3** option displays more information on candidate selection and files chosen for migration.

    For more information on **mmlspolicy -L**, see "The mmapplypolicy -L command" on page 256.

  – The file was scheduled for migration but was not moved. In this case, the file will be shown as 'ill-placed' by the **mmlsattr -L** command, indicating that the migration did not succeed. This occurs if the new storage pool assigned to the file did not have sufficient free space for the file when the actual migration was attempted. Since migrations are done in parallel, it is possible that the target pool had files which were also migrating, but had not yet been moved. If the target pool now has sufficient free space, the files can be moved using the commands: **mmrestripefs**, **mmrestripefile**, **mmchattr**.

- Asserts or error messages indicating a problem.

The policy rule language can only check for some errors at runtime. For example, a rule that causes a divide by zero cannot be checked when the policy file is installed. Errors of this type generate an error message and stop the policy evaluation for that file.

Note: I/O errors while migrating files indicate failing storage devices and must be addressed like any other I/O error. The same is true for any file system error or panic encountered while migrating files.

Filesets - usage errors

The misunderstandings while dealing with the filesets and the course of actions to correct them.

These are common mistakes and misunderstandings encountered when dealing with filesets:

1. Fileset junctions look very much like ordinary directories, but they cannot be deleted by the usual commands such as **rm -r** or **rmdir**. Using these commands on a fileset junction could result in a `Not owner` message on an AIX system, or an `Operation not permitted` message on a Linux system.

   As a consequence these commands may fail when applied to a directory that is a fileset junction. Similarly, when **rm -r** is applied to a directory that contains a fileset junction, it will fail as well.

   On the other hand, **rm -r** will delete all the files contained in the filesets linked under the specified directory. Use the **mmunlinkfileset** command to remove fileset junctions.
2. Files and directories may not be moved from one fileset to another, nor may a hard link cross fileset boundaries.

   If the user is unaware of the locations of fileset junctions, **mv** and **ln** commands may fail unexpectedly. In most cases, the **mv** command will automatically compensate for this failure and use a combination of **cp** and **rm** to accomplish the desired result. Use the **mmlsfileset** command to view the locations of fileset junctions. Use the **mmlsattr -L** command to determine the fileset for any given file.
3. Because a snapshot saves the contents of a fileset, deleting a fileset included in a snapshot cannot completely remove the fileset.

   The fileset is put into a 'deleted' state and continues to appear in **mmlsfileset** output. Once the last snapshot containing the fileset is deleted, the fileset will be completely removed automatically. The **mmlsfileset --deleted** command indicates deleted filesets and shows their names in parentheses.
4. Deleting a large fileset may take some time and may be interrupted by other failures, such as disk errors or system crashes.

   When this occurs, the recovery action leaves the fileset in a 'being deleted' state. Such a fileset may not be linked into the namespace. The corrective action it to finish the deletion by reissuing the fileset delete command:

   ```
   mmdelfileset fs1 fsname1 -f
   ```

   The **mmlsfileset** command identifies filesets in this state by displaying a status of 'Deleting'.
5. If you unlink a fileset that has other filesets linked below it, any filesets linked to it (that is, child filesets) become inaccessible. The child filesets remain linked to the parent and will become accessible again when the parent is re-linked.
6. By default, the **mmdelfileset** command will not delete a fileset that is not empty.

   To empty a fileset, first unlink all its immediate child filesets, to remove their junctions from the fileset to be deleted. Then, while the fileset itself is still linked, use **rm -rf** or a similar command, to remove the rest of the contents of the fileset. Now the fileset may be unlinked and deleted. Alternatively, the fileset to be deleted can be unlinked first and then **mmdelfileset** can be used with the **-f** (force) option. This will unlink its child filesets, then destroy the files and directories contained in the fileset.
7. When deleting a small dependent fileset, it may be faster to use the **rm -rf** command instead of the **mmdelfileset** command with the **-f** option.

Errors encountered with filesets

The analysis of those errors which may be encountered while dealing with the filesets.

These are errors encountered with filesets and how to analyze them:

1. Problems can arise when running backup and archive utilities against a file system with unlinked filesets. See the *Filesets and backup* topic in the *IBM Spectrum Scale: Administration Guide* for details.
2. In the rare case that the **mmfsck** command encounters a serious error checking the file system's fileset metadata, it may not be possible to reconstruct the fileset name and comment. These cannot be inferred from information elsewhere in the file system. If this happens, **mmfsck** will create a dummy name for the fileset, such as 'Fileset911' and the comment will be set to the empty string.
3. Sometimes **mmfsck** encounters orphaned files or directories (those without a parent directory), and traditionally these are reattached in a special directory called 'lost+found' in the file system root. When a file system contains multiple filesets, however, orphaned files and directories are reattached in the 'lost+found' directory in the root of the fileset to which they belong. For the root fileset, this directory appears in the usual place, but other filesets may each have their own 'lost+found' directory.

Active file management fileset errors

When the **mmafmctl** *Device* **getstate** command displays a `NeedsResync` target/fileset state, inconsistencies exist between the home and cache. To ensure that the cached data is synchronized with the home and the fileset is returned to `Active` state, either the file system must be unmounted and mounted or the fileset must be unlinked and linked. Once this is done, the next update to fileset data will trigger an automatic synchronization of data from the cache to the home.

Storage pools - usage errors

The misunderstandings while dealing with the storage pools and the course of actions to correct them.

These are common mistakes and misunderstandings encountered when dealing with storage pools:

1. Only the **system** storage pool is allowed to store metadata. All other pools must have the **dataOnly** attribute.
2. Take care to create your storage pools with sufficient numbers of failure groups to enable the desired level of replication.

   When the file system is created, GPFS requires all of the initial pools to have at least as many failure groups as defined by the default replication (**-m** and **-r** flags on the **mmcrfs** command). However, once the file system has been created, the user can create a storage pool with fewer failure groups than the default replication.

   The **mmadddisk** command issues a warning, but it allows the disks to be added and the storage pool defined. To use the new pool, the user must define a policy rule to create or migrate files into the new pool. This rule should be defined to set an appropriate replication level for each file assigned to the pool. If the replication level exceeds the number of failure groups in the storage pool, all files assigned to the pool incur added overhead on each write to the file, in order to mark the file as ill-replicated.

   To correct the problem, add additional disks to the storage pool, defining a different failure group, or insure that all policy rules that assign files to the pool also set the replication appropriately.
3. GPFS does not permit the **mmchdisk** or **mmrpldisk** command to change a disk's storage pool assignment. Changing the pool assignment requires all data residing on the disk to be moved to another disk before the disk can be reassigned. Moving the data is a costly and time-consuming operation; therefore GPFS requires an explicit **mmdeldisk** command to move it, rather than moving it as a side effect of another command.
4. Some storage pools allow larger disks to be added than do other storage pools.

When the file system is created, GPFS defines the maximum size disk that can be supported using the on-disk data structures to represent it. Likewise, when defining a new storage pool, the newly created on-disk structures establish a limit on the maximum size disk that can be added to that pool.

To add disks that exceed the maximum size allowed by a storage pool, simply create a new pool using the larger disks.

The **mmdf** command can be used to find the maximum disk size allowed for a storage pool.

5. If you try to delete a storage pool when there are files still assigned to the pool, consider this:

   A storage pool is deleted when all disks assigned to the pool are deleted. To delete the last disk, all data residing in the pool must be moved to another pool. Likewise, any files assigned to the pool, whether or not they contain data, must be reassigned to another pool. The easiest method for reassigning all files and migrating all data is to use the **mmapplypolicy** command with a single rule to move all data from one pool to another. You should also install a new placement policy that does not assign new files to the old pool. Once all files have been migrated, reissue the **mmdeldisk** command to delete the disk and the storage pool.

   If all else fails, and you have a disk that has failed and cannot be recovered, follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center for commands to allow the disk to be deleted without migrating all data from it. Files with data left on the failed device will lose data. If the entire pool is deleted, any existing files assigned to that pool are reassigned to a "broken" pool, which prevents writes to the file until the file is reassigned to a valid pool.

6. Ill-placed files - understanding and correcting them.

   The **mmapplypolicy** command migrates a file between pools by first assigning it to a new pool, then moving the file's data. Until the existing data is moved, the file is marked as 'ill-placed' to indicate that some of its data resides in its previous pool. In practice, **mmapplypolicy** assigns all files to be migrated to their new pools, then it migrates all of the data in parallel. Ill-placed files indicate that the **mmapplypolicy** or **mmchattr** command did not complete its last migration or that **-I defer** was used.

   To correct the placement of the ill-placed files, the file data needs to be migrated to the assigned pools. You can use the **mmrestripefs**, or **mmrestripefile** commands to move the data.

7. Using the **-P** *PoolName* option on the **mmrestripefs**, command:

   This option restricts the restripe operation to a single storage pool. For example, after adding a disk to a pool, only the data in that pool needs to be restriped. In practice, **-P** *PoolName* simply restricts the operation to the files assigned to the specified pool. Files assigned to other pools are not included in the operation, even if the file is ill-placed and has data in the specified pool.

Errors encountered with storage pools

The analysis of those errors which may be encountered while dealing with the storage pools.

These are error encountered with policies and how to analyze them:

1. Access time to one pool appears slower than the others.

   A consequence of striping data across the disks is that the I/O throughput is limited by the slowest device. A device encountering hardware errors or recovering from hardware errors may effectively limit the throughput to all devices. However using storage pools, striping is done only across the disks assigned to the pool. Thus a slow disk impacts only its own pool; all other pools are not impeded.

   To correct the problem, check the connectivity and error logs for all disks in the slow pool.

2. Other storage pool problems might really be disk problems and should be pursued from the standpoint of making sure that your disks are properly configured and operational. See Chapter 19, "Disk issues," on page 349.

Snapshot problems

Use the **mmlssnapshot** command as a general hint for snapshot-related problems, to find out what snapshots exist, and what state they are in. Use the **mmsnapdir** command to find the snapshot directory name used to permit access.

The **mmlssnapshot** command displays the list of **all** snapshots of a file system. This command lists the snapshot name, some attributes of the snapshot, as well as the snapshot's status. The **mmlssnapshot** command does not require the file system to be mounted.

Problems with locating a snapshot

Use the **mmlssnapshot** command and **mmsnapdir** command to find the snapshot detail and locate them.

The **mmlssnapshot** and **mmsnapdir** commands are provided to assist in locating the snapshots in the file system directory structure. Only valid snapshots are visible in the file system directory structure. They appear in a hidden subdirectory of the file system's root directory. By default the subdirectory is named **.snapshots**. The valid snapshots appear as entries in the snapshot directory and may be traversed like any other directory. The **mmsnapdir** command can be used to display the assigned snapshot directory name.

Problems not directly related to snapshots

There are errors which are returned from the snapshot commands but are not linked with the snapshots directly.

Many errors returned from the snapshot commands are not specifically related to the snapshot. For example, disk failures or node failures could cause a snapshot command to fail. The response to these types of errors is to fix the underlying problem and try the snapshot command again.

GPFS error messages for indirect snapshot errors

There are GPFS error messages which may be associated with snapshots directly but does not show a clear relation to snapshot issues.

The error messages for this type of problem do not have message numbers, but can be recognized by their message text:

- 'Unable to sync all nodes, rc=*errorCode*.'
- 'Unable to get permission to create snapshot, rc=*errorCode*.'
- 'Unable to quiesce all nodes, rc=*errorCode*.'
- 'Unable to resume all nodes, rc=*errorCode*.'
- 'Unable to delete snapshot *filesystemName* from file system *snapshotName*, rc=*errorCode*.'
- 'Error restoring inode *number*, error *errorCode*.'
- 'Error deleting snapshot *snapshotName* in file system *filesystemName*, error *errorCode*.'
- '*commandString* failed, error *errorCode*.'
- 'None of the nodes in the cluster is reachable, or GPFS is down on all of the nodes.'
- 'File system *filesystemName* is not known to the GPFS cluster.'

Snapshot usage errors

Certain error in the GPFS error messages are related to the snapshot usage restrictions or incorrect snapshot names .

Many errors returned from the snapshot commands are related to usage restrictions or incorrect snapshot names.

An example of a snapshot restriction error is exceeding the maximum number of snapshots allowed at one time. For simple errors of these types, you can determine the source of the error by reading the error message or by reading the description of the command. You can also run the **mmlssnapshot** command to see the complete list of existing snapshots.

Examples of incorrect snapshot name errors are trying to delete a snapshot that does not exist or trying to create a snapshot using the same name as an existing snapshot. The rules for naming global and fileset snapshots are designed to minimize conflicts between the file system administrator and the fileset owners. These rules can result in errors when fileset snapshot names are duplicated across different filesets or when the snapshot command `-j` option (specifying a qualifying fileset name) is provided or omitted incorrectly. To resolve name problems review the **mmlssnapshot** output with careful attention to the `Fileset` column. You can also specify the `-s` or `-j` options of the **mmlssnapshot** command to limit the output. For snapshot deletion, the `-j` option must exactly match the `Fileset` column.

For more information about snapshot naming conventions, see the **mmcrsnapshot** command in the *IBM Spectrum Scale: Command and Programming Reference*.

GPFS error messages for snapshot usage errors

Certain error messages for snapshot usage errors have no error message numbers but may be recognized using the message texts.

The error messages for this type of problem do not have message numbers, but can be recognized by their message text:

- 'File system *filesystemName* does not contain a snapshot *snapshotName*, rc=*errorCode*.'
- 'Cannot create a new snapshot until an existing one is deleted. File system *filesystemName* has a limit of *number* online snapshots.'
- 'Cannot restore snapshot. *snapshotName* is mounted on *number* nodes and in use on *number* nodes.'
- 'Cannot create a snapshot in a DM enabled file system, rc=*errorCode*.'

Snapshot status errors

There are certain snapshot commands like **mmdelsnapshot** and **mmrestorefs**, which lets snapshot go invalid if they got interrupted while running.

Some snapshot commands like **mmdelsnapshot** and **mmrestorefs** may require a substantial amount of time to complete. If the command is interrupted, say by the user or due to a failure, the snapshot may be left in an invalid state. In many cases, the command must be completed before other snapshot commands are allowed to run. The source of the error may be determined from the error message, the command description, or the snapshot status available from **mmlssnapshot**.

GPFS error messages for snapshot status errors

Certain error messages for snapshot status error have no error message numbers and may be recognized by the message texts only.

The error messages for this type of problem do not have message numbers, but can be recognized by their message text:

- 'Cannot delete snapshot *snapshotName* which is *snapshotState*, error = *errorCode*.'
- 'Cannot restore snapshot *snapshotName* which is *snapshotState*, error = *errorCode*.'
- 'Previous snapshot *snapshotName* is invalid and must be deleted before a new snapshot may be created.'
- 'Previous snapshot *snapshotName* must be restored before a new snapshot may be created.'
- 'Previous snapshot *snapshotName* is invalid and must be deleted before another snapshot may be deleted.'
- 'Previous snapshot *snapshotName* is invalid and must be deleted before another snapshot may be restored.'

- 'More than one snapshot is marked for restore.'
- 'Offline snapshot being restored.'

Snapshot directory name conflicts

The snapshot generated by **mmcrsnapshot** command may not be accessed due to directory conflicts and the course of action to correct the snapshot directory name conflict.

By default, all snapshots appear in a directory named **.snapshots** in the root directory of the file system. This directory is dynamically generated when the first snapshot is created and continues to exist even after the last snapshot is deleted. If the user tries to create the first snapshot, and a normal file or directory named **.snapshots** already exists, the **mmcrsnapshot** command will be successful but the snapshot may not be accessed.

There are two ways to fix this problem:

1. Delete or rename the existing file or directory
2. Tell GPFS to use a different name for the dynamically-generated directory of snapshots by running the **mmsnapdir** command.

It is also possible to get a name conflict as a result of issuing the **mmrestorefs** command. Since **mmsnapdir** allows changing the name of the dynamically-generated snapshot directory, it is possible that an older snapshot contains a normal file or directory that conflicts with the current name of the snapshot directory. When this older snapshot is restored, the **mmrestorefs** command will recreate the old, normal file or directory in the file system root directory. The **mmrestorefs** command will not fail in this case, but the restored file or directory will hide the existing snapshots. After invoking **mmrestorefs** it may therefore appear as if the existing snapshots have disappeared. However, **mmlssnapshot** should still show all existing snapshots.

The fix is the similar to the one mentioned before. Perform one of these two steps:

1. After the **mmrestorefs** command completes, rename the conflicting file or directory that was restored in the root directory.
2. Run the **mmsnapdir** command to select a different name for the dynamically-generated snapshot directory.

Finally, the **mmsnapdir -a** option enables a dynamically-generated snapshot directory in every directory, not just the file system root. This allows each user quick access to snapshots of their own files by going into **.snapshots** in their home directory or any other of their directories.

Unlike **.snapshots** in the file system root, **.snapshots** in other directories is invisible, that is, an **ls -a** command will not list **.snapshots**. This is intentional because recursive file system utilities such as **find**, **du** or **ls -R** would otherwise either fail or produce incorrect or undesirable results. To access snapshots, the user must explicitly specify the name of the snapshot directory, for example: **ls ~/.snapshots**. If there is a name conflict (that is, a normal file or directory named **.snapshots** already exists in the user's home directory), the user must rename the existing file or directory.

The inode numbers that are used for and within these special **.snapshots** directories are constructed dynamically and do not follow the standard rules. These inode numbers are visible to applications through standard commands, such as **stat**, **readdir**, or **ls**. The inode numbers reported for these directories can also be reported differently on different operating systems. Applications should not expect consistent numbering for such inodes.

Errors encountered when restoring a snapshot

There are errors which may be displayed while restoring a snapshot.

The following errors might be encountered when restoring from a snapshot:

- The **mmrestorefs** command fails with an **ENOSPC** message. In this case, there are not enough free blocks in the file system to restore the selected snapshot. You can add space to the file system by adding a new disk. As an alternative, you can delete a different snapshot from the file system to free some existing space. You cannot delete the snapshot that is being restored. After there is additional free space, issue the **mmrestorefs** command again.
- The **mmrestorefs** command fails with quota exceeded errors. Try adjusting the quota configuration or disabling quota, and then issue the command again.
- The **mmrestorefs** command is interrupted and some user data is not be restored completely. Try repeating the **mmrestorefs** command in this instance.
- The **mmrestorefs** command fails because of an incorrect file system, fileset, or snapshot name. To fix this error, issue the command again with the correct name.
- The **mmrestorefs -j** command fails with the following error:

  6027-953
  Failed to get a handle for fileset *filesetName*, snapshot *snapshotName* in file system *fileSystem*. *errorMessage*.

  In this case, the file system that contains the snapshot to restore should be mounted, and then the fileset of the snapshot should be linked.

If you encounter additional errors that cannot be resolved, contact the IBM Support Center.

Errors encountered when restoring a snapshot

There are errors which may be displayed while restoring a snapshot.

The following errors might be encountered when restoring from a snapshot:

- The **mmrestorefs** command fails with an **ENOSPC** message. In this case, there are not enough free blocks in the file system to restore the selected snapshot. You can add space to the file system by adding a new disk. As an alternative, you can delete a different snapshot from the file system to free some existing space. You cannot delete the snapshot that is being restored. After there is additional free space, issue the **mmrestorefs** command again.
- The **mmrestorefs** command fails with quota exceeded errors. Try adjusting the quota configuration or disabling quota, and then issue the command again.
- The **mmrestorefs** command is interrupted and some user data is not be restored completely. Try repeating the **mmrestorefs** command in this instance.
- The **mmrestorefs** command fails because of an incorrect file system, fileset, or snapshot name. To fix this error, issue the command again with the correct name.
- The **mmrestorefs -j** command fails with the following error:

  6027-953
  Failed to get a handle for fileset *filesetName*, snapshot *snapshotName* in file system *fileSystem*. *errorMessage*.

  In this case, the file system that contains the snapshot to restore should be mounted, and then the fileset of the snapshot should be linked.

If you encounter additional errors that cannot be resolved, contact the IBM Support Center.

Failures using the mmpmon command

The **mmpmon** command manages performance monitoring and displays performance information.

The **mmpmon** command is thoroughly documented in "Monitoring GPFS I/O performance with the mmpmon command" on page 4 and the *mmpmon command* page in the *IBM Spectrum Scale: Command and Programming Reference*. Before proceeding with **mmpmon** problem determination, review all of this material to ensure that you are using the **mmpmon** command correctly.

Setup problems using mmpmon

The issues associated with the set up of **mmpmon** command and limitations of this command.

Remember these points when using the **mmpmon** command:

- You must have root authority.
- The GPFS daemon must be active.
- The input file must contain valid input requests, one per line. When an incorrect request is detected by **mmpmon**, it issues an error message and terminates.

  Input requests that appear in the input file before the first incorrect request are processed by **mmpmon**.
- Do not alter the input file while **mmpmon** is running.
- Output from **mmpmon** is sent to standard output (STDOUT) and errors are sent to standard (STDERR).
- Up to five instances of **mmpmon** may run on a given node concurrently. See "Monitoring GPFS I/O performance with the mmpmon command" on page 4. For the limitations regarding concurrent usage of **mmpmon**, see "Running mmpmon concurrently from multiple users on the same node" on page 6.
- The **mmpmon** command does **not** support:
  - Monitoring read requests without monitoring writes, or the other way around.
  - Choosing which file systems to monitor.
  - Monitoring on a per-disk basis.
  - Specifying different size or latency ranges for reads and writes.
  - Specifying different latency values for a given size range.

Incorrect output from mmpmon

The analysis of incorrect output of **mmpmon** command.

If the output from **mmpmon** is incorrect, such as zero counters when you know that I/O activity is taking place, consider these points:

1. Someone may have issued the `reset` or `rhist reset` requests.
2. Counters may have wrapped due to a large amount of I/O activity, or running **mmpmon** for an extended period of time. For a discussion of counter sizes and counter wrapping, see *Counter sizes and counter wrapping* section in "Monitoring GPFS I/O performance with the mmpmon command" on page 4.
3. See the *Other information about mmpmon output* section in "Monitoring GPFS I/O performance with the mmpmon command" on page 4. This section gives specific instances where **mmpmon** output may be different than what was expected.

Abnormal termination or hang in mmpmon

The course of action which must be followed if **mmpmon** command hangs or terminates.

If **mmpmon** hangs, perform these steps:

1. Ensure that sufficient time has elapsed to cover the **mmpmon** timeout value. It is controlled using the **-t** flag on the **mmpmon** command.
2. Issue the **ps** command to find the PID for **mmpmon**.
3. Issue the **kill** command to terminate this PID.
4. Try the function again.
5. If the problem persists, issue this command:

   ```
   mmfsadm dump eventsExporter
   ```
6. Copy the output of **mmfsadm** to a safe location.
7. Follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.

If **mmpmon** terminates abnormally, perform these steps:

1. Determine if the GPFS daemon has failed, and if so restart it.
2. Review your invocation of **mmpmon**, and verify the input.
3. Try the function again.
4. If the problem persists, follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.

Tracing the mmpmon command

The course of action to be followed if the **mmpmon** command does not perform as expected.

When the **mmpmon** command does not work properly, there are two trace classes used to determine the cause of the problem. Use these only when requested by the IBM Support Center.

eventsExporter
: Reports attempts to connect and whether or not they were successful.

mmpmon
: Shows the command string that came in to the **mmpmon** command, and whether it was successful or not.

Note: Do not use the **perfmon** trace class of the GPFS trace to diagnose **mmpmon** problems. This trace event does not provide the necessary data.

Failures using the mmbackup command

Use the **mmbackup** command to back up the files in a GPFS file system to storage on a IBM Spectrum Protect server. A number of factors can cause **mmbackup** to fail.

The most common of these are:

- The file system is not mounted on the node issuing the **mmbackup** command.
- The file system is not mounted on the IBM Spectrum Protect client nodes.
- The **mmbackup** command was issued to back up a file system owned by a remote cluster.
- The IBM Spectrum Protect clients are not able to communicate with the IBM Spectrum Protect server due to authorization problems.
- The IBM Spectrum Protect server is down or out of storage space.
- When the target of the backup is tape, the IBM Spectrum Protect server may be unable to handle all of the backup client processes because the value of the IBM Spectrum Protect server's MAXNUMMP parameter is set lower than the number of client processes. This failure is indicated by message ANS1312E from IBM Spectrum Protect.

The errors from **mmbackup** normally indicate the underlying problem.

GPFS error messages for mmbackup errors

Error messages that are displayed for mmbackup errors

6027-1995
: Device *deviceName* is not mounted on node *nodeName*.

IBM Spectrum Protect error messages

Error message displayed for server media mount.

ANS1312E
: Server media mount not possible.

Data integrity

GPFS takes extraordinary care to maintain the integrity of customer data. However, certain hardware failures, or in extremely unusual circumstances, the occurrence of a programming error can cause the loss of data in a file system.

GPFS performs extensive checking to validate metadata and ceases using the file system if metadata becomes inconsistent. This can appear in two ways:

1. The file system will be unmounted and applications will begin seeing `ESTALE` return codes to file operations.
2. Error log entries indicating an **MMFS_SYSTEM_UNMOUNT** and a corruption error are generated.

If actual disk data corruption occurs, this error will appear on each node in succession. Before proceeding with the following steps, follow the procedures in "Information to be collected before contacting the IBM Support Center" on page 469, and then contact the IBM Support Center.

1. Examine the error logs on the NSD servers for any indication of a disk error that has been reported.
2. Take appropriate disk problem determination and repair actions prior to continuing.
3. After completing any required disk repair actions, run the offline version of the **mmfsck** command on the file system.
4. If your error log or disk analysis tool indicates that specific disk blocks are in error, use the **mmfileid** command to determine which files are located on damaged areas of the disk, and then restore these files. See "The mmfileid command" on page 264 for more information.
5. If data corruption errors occur in only one node, it is probable that memory structures within the node have been corrupted. In this case, the file system is probably good but a program error exists in GPFS or another authorized program with access to GPFS data structures.

   Follow the directions in "Data integrity" and then reboot the node. This should clear the problem. If the problem repeats on one node without affecting other nodes check the programming specifications code levels to determine that they are current and compatible and that no hardware errors were reported. Refer to the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* for correct software levels.

Error numbers specific to GPFS application calls when data integrity may be corrupted

If there is a possibility of the corruption of data integrity, GPFS displays specific error messages or returns them to the application.

When there is the possibility of data corruption, GPFS may report these error numbers in the operating system error log, or return them to an application:

EVALIDATE=214, Invalid checksum or other consistency check failure on disk data structure.
: This indicates that internal checking has found an error in a metadata structure. The severity of the error depends on which data structure is involved. The cause of this is usually GPFS software, disk hardware or other software between GPFS and the disk. Running **mmfsck** should repair the error. The urgency of this depends on whether the error prevents access to some file or whether basic metadata structures are involved.

Messages requeuing in AFM

The course of actions to be followed for resolving the requeued messages in the Gateway node

Sometimes requests in the AFM messages queue on the gateway node get requested because of errors at home. For example, if there is no space at home to perform a new write, a write message that is queued is not successful and gets requeued. The administrator would see the failed message getting requeued in

the queue on the gateway node. The administrator has to resolve the issue by adding more space at home and running the **mmafmctl resumeRequeued** command, so that the requeued messages are executed at home again. If **mmafmctl resumeRequeued** is not run by an administrator, AFM would still execute the message in the regular order of message executions from cache to home.

Running the **mmfsadm dump afm all** command on the gateway node shows the queued messages. Requeued messages show in the dumps similar to the following example:

```
c12c4apv13.gpfs.net:    Normal Queue:  (listed by execution order) (state: Active)
c12c4apv13.gpfs.net:      Write [612457.552962] requeued file3 (43 @ 293) chunks 0 bytes 0 0
```

Chapter 19. Disk issues

GPFS uses only disk devices prepared as Network Shared Disks (NSDs). However NSDs might exist on top of a number of underlying disk technologies.

NSDs, for example, might be defined on top of Fibre Channel SAN connected disks. This information provides detail on the creation, use, and failure of NSDs and their underlying disk technologies.

These are some of the errors encountered with GPFS disks and NSDs:

- "NSD and underlying disk subsystem failures"
- "GPFS has declared NSDs built on top of AIX logical volumes as down" on page 356
- "Disk accessing commands fail to complete due to problems with some non-IBM disks" on page 357
- "Persistent Reserve errors" on page 361
- "GPFS is not using the underlying multipath device" on page 363

NSD and underlying disk subsystem failures

There are indications that will lead you to the conclusion that your file system has disk failures.

Some of those indications include:

- Your file system has been forced to unmount. For more information about forced file system unmount, see "File system forced unmount" on page 322.
- The **mmlsmount** command indicates that the file system is not mounted on certain nodes.
- Your application is getting EIO errors.
- Operating system error logs indicate you have stopped using a disk in a replicated system, but your replication continues to operate.
- The **mmlsdisk** command shows that disks are down.

Note: If you are reinstalling the operating system on one node and erasing all partitions from the system, GPFS descriptors will be removed from any NSD this node can access locally. The results of this action might require recreating the file system and restoring from backup. If you experience this problem, do not unmount the file system on any node that is currently mounting the file system. Contact the IBM Support Center immediately to see if the problem can be corrected.

Error encountered while creating and using NSD disks

Use **mmcrnsd** command to prepare NSD disks. While preparing the NSD disks, there are several errors conditions encountered.

GPFS requires that disk devices be prepared as NSDs. This is done using the **mmcrnsd** command. The input to the **mmcrnsd** command is given in the form of disk stanzas. For a complete explanation of disk stanzas, see the *Stanza files* section in the *IBM Spectrum Scale: Administration Guide*, and the following topics from the *IBM Spectrum Scale: Command and Programming Reference*:

- **mmcdisk** command
- **mmchnsd** command
- **mmcrfs** command
- **mmcrnsd** command

For disks that are SAN-attached to all nodes in the cluster, **device**=*DiskName* should refer to the disk device name in **/dev** on the node where the **mmcrnsd** command is issued. If a server list is specified, **device**=*DiskName* must refer to the name of the disk on the first server node. The same disk can have different local names on different nodes.

When you specify an NSD server node, that node performs all disk I/O operations on behalf of nodes in the cluster that do not have connectivity to the disk. You can also specify up to eight additional NSD server nodes. These additional NSD servers will become active if the first NSD server node fails or is unavailable.

When the **mmcrnsd** command encounters an error condition, one of these messages is displayed:

6027-2108
Error found while processing stanza

or

6027-1636
Error found while checking disk descriptor *descriptor*

Usually, this message is preceded by one or more messages describing the error more specifically.

Another possible error from **mmcrnsd** is:

6027-2109
Failed while processing disk stanza on node *nodeName*.

or

6027-1661
Failed while processing disk descriptor *descriptor* on node *nodeName*.

One of these errors can occur if an NSD server node does not have read and write access to the disk. The NSD server node needs to write an NSD volume ID to the raw disk. If an additional NSD server node is specified, that NSD server node will scan its disks to find this NSD volume ID string. If the disk is SAN-attached to all nodes in the cluster, the NSD volume ID is written to the disk by the node on which the **mmcrnsd** command is running.

Displaying NSD information

Use **mmlsnsd** command to display the NSD information and analyze the cluster details pertaining to NSDs.

Use the **mmlsnsd** command to display information about the currently defined NSDs in the cluster. For example, if you issue **mmlsnsd**, your output may be similar to this:

```
 File system   Disk name    NSD servers
---------------------------------------------------------------------------
 fs1           t65nsd4b     (directly attached)
 fs5           t65nsd12b    c26f4gp01.ppd.pok.ibm.com,c26f4gp02.ppd.pok.ibm.com
 fs6           t65nsd13b    c26f4gp01.ppd.pok.ibm.com,c26f4gp02.ppd.pok.ibm.com,c26f4gp03.ppd.pok.ibm.com
```

This output shows that:

- There are three NSDs in this cluster: **t65nsd4b**, **t65nsd12b**, and **t65nsd13b**.
- NSD disk **t65nsd4b** of file system **fs1** is SAN-attached to all nodes in the cluster.
- NSD disk **t65nsd12b** of file system **fs5** has 2 NSD server nodes.
- NSD disk **t65nsd13b** of file system **fs6** has 3 NSD server nodes.

If you need to find out the local device names for these disks, you could use the **-m** option on the **mmlsnsd** command. For example, issuing:

```
mmlsnsd -m
```

produces output similar to this example:

```
 Disk name    NSD volume ID      Device         Node name                Remarks
---------------------------------------------------------------------------------------
 t65nsd12b    0972364D45EF7B78   /dev/hdisk34   c26f4gp01.ppd.pok.ibm.com   server node
 t65nsd12b    0972364D45EF7B78   /dev/hdisk34   c26f4gp02.ppd.pok.ibm.com   server node
 t65nsd12b    0972364D45EF7B78   /dev/hdisk34   c26f4gp04.ppd.pok.ibm.com
 t65nsd13b    0972364D00000001   /dev/hdisk35   c26f4gp01.ppd.pok.ibm.com   server node
 t65nsd13b    0972364D00000001   /dev/hdisk35   c26f4gp02.ppd.pok.ibm.com   server node
 t65nsd13b    0972364D00000001   -              c26f4gp03.ppd.pok.ibm.com   (not found) server node
 t65nsd4b     0972364D45EF7614   /dev/hdisk26   c26f4gp04.ppd.pok.ibm.com
```

From this output we can tell that:

- The local disk name for **t65nsd12b** on NSD server **c26f4gp01** is **hdisk34**.
- NSD disk **t65nsd13b** is not attached to node on which the **mmlsnsd** command was issued, node**c26f4gp04**.
- The **mmlsnsd** command was not able to determine the local device for NSD disk **t65nsd13b** on **c26f4gp03** server.

To find the nodes to which disk **t65nsd4b** is attached and the corresponding local devices for that disk, issue:

```
mmlsnsd -d t65nsd4b -M
```

Output is similar to this example:

```
 Disk name    NSD volume ID      Device         Node name                Remarks
---------------------------------------------------------------------------------------
 t65nsd4b     0972364D45EF7614   /dev/hdisk92   c26f4gp01.ppd.pok.ibm.com
 t65nsd4b     0972364D45EF7614   /dev/hdisk92   c26f4gp02.ppd.pok.ibm.com
 t65nsd4b     0972364D45EF7614   -              c26f4gp03.ppd.pok.ibm.com   (not found) directly attached
 t65nsd4b     0972364D45EF7614   /dev/hdisk26   c26f4gp04.ppd.pok.ibm.com
```

From this output we can tell that NSD **t65nsd4b** is:

- Known as **hdisk92** on node **c26f4gp01** and **c26f4gp02**.
- Known as **hdisk26** on node **c26f4gp04**
- Is not attached to node **c26f4gp03**

To display extended information about a node's view of its NSDs, the **mmlsnsd -X** command can be used:

```
mmlsnsd -X -d "hd3n97;sdfnsd;hd5n98"
```

The system displays information similar to:

```
 Disk name    NSD volume ID      Device         Devtype  Node name                Remarks
---------------------------------------------------------------------------------------------
 hd3n97       0972846145C8E927   /dev/hdisk3    hdisk    c5n97g.ppd.pok.ibm.com   server node,pr=no
 hd3n97       0972846145C8E927   /dev/hdisk3    hdisk    c5n98g.ppd.pok.ibm.com   server node,pr=no
 hd5n98       0972846245EB501C   /dev/hdisk5    hdisk    c5n97g.ppd.pok.ibm.com   server node,pr=no
 hd5n98       0972846245EB501C   /dev/hdisk5    hdisk    c5n98g.ppd.pok.ibm.com   server node,pr=no
 sdfnsd       0972845E45F02E81   /dev/sdf       generic  c5n94g.ppd.pok.ibm.com   server node
 sdfnsd       0972845E45F02E81   /dev/sdm       generic  c5n96g.ppd.pok.ibm.com   server node
```

From this output we can tell that:

- Disk **hd3n97** is an hdisk known as **/dev/hdisk3** on NSD server node **c5n97** and **c5n98**.

- Disk **sdfnsd** is a generic disk known as **/dev/sdf** and **/dev/sdm** on NSD server node **c5n94g** and **c5n96g**, respectively.
- In addition to the preceding information, the NSD volume ID is displayed for each disk.

Note: The **-m**, **-M** and **-X** options of the **mmlsnsd** command can be very time consuming, especially on large clusters. Use these options judiciously.

Disk device name is an existing NSD name

Learn how to respond to an NSD creation error message in which the device name is an existing NSD name.

When you run the **mmcrnsd** command to create an NSD, the command might display an error message saying that a *DiskName* value that you specified refers to an existing NSD name.

This type of error message indicates one of the following situations:
- The disk is an existing NSD.
- The disk is a previous NSD that was removed from the cluster with the **mmdelnsd** command but is not yet marked as available.

In second situation, you can override the check by running the **mmcrnsd** command again with the `-v no` option. Do not take this step unless you are sure that another cluster is not using this disk. Enter the following command:

```
mmcrnsd -F StanzaFile -v no
```

A possible cause for the NSD creation error message is that a previous **mmdelnsd** command failed to zero internal data structures on the disk, even though the disk is functioning correctly. To complete the deletion, run the **mmdelnsd** command with the `-p NSDId` option. Do not take this step unless you are sure that another cluster is not using this disk. The following command is an example:

```
mmdelsd -p NSDId -N Node
```

GPFS has declared NSDs as down

GPFS reactions to NSD failures and the recovery procedure.

There are several situations in which disks can appear to fail to GPFS. Almost all of these situations involve a failure of the underlying disk subsystem. The following information describes how GPFS reacts to these failures and how to find the cause.

GPFS will stop using a disk that is determined to have failed. This event is marked as **MMFS_DISKFAIL** in an error log entry (see "Operating system error logs" on page 214). The state of a disk can be checked by issuing the **mmlsdisk** command.

The consequences of stopping disk usage depend on what is stored on the disk:
- Certain data blocks may be unavailable because the data residing on a stopped disk is not replicated.
- Certain data blocks may be unavailable because the controlling metadata resides on a stopped disk.
- In conjunction with other disks that have failed, all copies of critical data structures may be unavailable resulting in the unavailability of the entire file system.

The disk will remain unavailable until its status is explicitly changed through the **mmchdisk** command. After that command is issued, any replicas that exist on the failed disk are updated before the disk is used.

GPFS can declare disks **down** for a number of reasons:

- If the first NSD server goes down and additional NSD servers were not assigned, or all of the additional NSD servers are also down and no local device access is available on the node, the disks are marked as stopped.
- A failure of an underlying disk subsystem may result in a similar marking of disks as stopped.
  1. Issue the **mmlsdisk** command to verify the status of the disks in the file system.
  2. Issue the **mmchdisk** command with the **-a** option to start all stopped disks.
- Disk failures should be accompanied by error log entries (see The operating system error log facility) for the failing disk. GPFS error log entries labelled **MMFS_DISKFAIL** will occur on the node detecting the error. This error log entry will contain the identifier of the failed disk. Follow the problem determination and repair actions specified in your disk vendor problem determination guide. After performing problem determination and repair issue the **mmchdisk** command to bring the disk back up.

Unable to access disks

Access to the disk might be restricted due to incorrect disk specification or configuration failure during disk subsystem initialization.

If you cannot open a disk, the specification of the disk may be incorrect. It is also possible that a configuration failure may have occurred during disk subsystem initialization. For example, on Linux you should consult **/var/log/messages** to determine if disk device configuration errors have occurred.

```
Feb 16 13:11:18 host123 kernel: SCSI device sdu: 35466240 512-byte hdwr sectors (18159 MB)
Feb 16 13:11:18 host123 kernel:  sdu: I/O error: dev 41:40, sector 0
Feb 16 13:11:18 host123 kernel:  unable to read partition table
```

On AIX, consult "Operating system error logs" on page 214 for hardware configuration error log entries.

Accessible disk devices will generate error log entries similar to this example for a SSA device:

```
---------------------------------------------------------------------------
LABEL:          SSA_DEVICE_ERROR
IDENTIFIER:     FE9E9357

Date/Time:       Wed Sep  8 10:28:13 edt
Sequence Number: 54638
Machine Id:      000203334C00
Node Id:         c154n09
Class:           H
Type:            PERM
Resource Name:   pdisk23
Resource Class:  pdisk
Resource Type:   scsd
Location:        USSA4B33-D3
VPD:
        Manufacturer................IBM
        Machine Type and Model......DRVC18B
        Part Number.................09L1813
        ROS Level and ID............0022
        Serial Number...............6800D2A6HK
        EC Level....................E32032
        Device Specific.(Z2)........CUSHA022
        Device Specific.(Z3)........09L1813
        Device Specific.(Z4)........99168

Description
DISK OPERATION ERROR

Probable Causes
DASD DEVICE

Failure Causes
DISK DRIVE
```

```
        Recommended Actions
        PERFORM PROBLEM DETERMINATION PROCEDURES

Detail Data
ERROR CODE
2310 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000 0000
---------------------------------------------------------------------------
```

or this one from GPFS:

```
---------------------------------------------------------------------------
LABEL:          MMFS_DISKFAIL
IDENTIFIER:     9C6C05FA

Date/Time:       Tue Aug  3 11:26:34 edt
Sequence Number: 55062
Machine Id:      000196364C00
Node Id:         c154n01
Class:           H
Type:            PERM
Resource Name:   mmfs
Resource Class:  NONE
Resource Type:   NONE
Location:

Description
DISK FAILURE

Probable Causes
STORAGE SUBSYSTEM
DISK

Failure Causes
STORAGE SUBSYSTEM
DISK

        Recommended Actions
        CHECK POWER
        RUN DIAGNOSTICS AGAINST THE FAILING DEVICE

Detail Data
EVENT CODE
     1027755
VOLUME
fs3
RETURN CODE
          19
PHYSICAL VOLUME
vp31n05
---------------------------------------------------------------
```

Guarding against disk failures

Protection methods to guard against data loss due to disk media failure.

There are various ways to guard against the loss of data due to disk media failures. For example, the use of a RAID controller, which masks disk failures with parity disks, or a twin-tailed disk, could prevent the need for using these recovery steps.

GPFS offers a method of protection that is called *replication*, which overcomes disk failure at the expense of extra disk space. GPFS allows replication of data and metadata. This means that up to three instances of data, metadata, or both can be automatically created and maintained for any file in a GPFS file system. If one instance becomes unavailable due to disk failure, another instance is used instead. You can set

different replication specifications for each file, or apply default settings that are specified at file system creation. Refer to the *File system replication parameters* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Disk connectivity failure and recovery

The GPFS has certain error messages defined for local connection failure from NSD servers.

If a disk is defined to have a local connection and to be connected to defined NSD servers, and the local connection fails, GPFS bypasses the broken local connection and uses the NSD servers to maintain disk access. The following error message appears in the GPFS log:

6027-361 [E]
Local access to *disk* failed with EIO, switching to access the disk remotely.

This is the default behavior, and can be changed with the **useNSDserver** file system mount option. See the *NSD server considerations* topic in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

For a file system using the default mount option **useNSDserver=asneeded**, disk access fails over from local access to remote NSD access. Once local access is restored, GPFS detects this fact and switches back to local access. The detection and switch over are not instantaneous, but occur at approximately five minute intervals.

Note: In general, after fixing the path to a disk, you must run the **mmnsddiscover** command on the server that lost the path to the NSD. (Until the **mmnsddiscover** command is run, the reconnected node will see its local disks and start using them by itself, but it will not act as the NSD server.)

After that, you must run the command on all client nodes that need to access the NSD on that server; or you can achieve the same effect with a single **mmnsddiscover** invocation if you utilize the **-N** option to specify a node list that contains all the NSD servers and clients that need to rediscover paths.

Partial disk failure

Partial disk failures when you have chosen not to implement hardware protection against media failures and the course of action to correct this problem.

If the disk has only partially failed and you have chosen not to implement hardware protection against media failures, the steps to restore your data depends on whether you have used replication. If you have replicated neither your data nor metadata, you will need to issue the offline version of the **mmfsck** command, and then restore the lost information from the backup media. If it is just the data which was not replicated, you will need to restore the data from the backup media. There is no need to run the **mmfsck** command if the metadata is intact.

If both your data and metadata have been replicated, implement these recovery actions:

1. Unmount the file system:
   ```
   mmumount fs1 -a
   ```
2. Delete the disk from the file system:
   ```
   mmdeldisk fs1 gpfs10nsd -c
   ```
3. If you are replacing the disk, add the new disk to the file system:
   ```
   mmadddisk fs1 gpfs11nsd
   ```
4. Then restripe the file system:
   ```
   mmrestripefs fs1 -b
   ```

   Note: Ensure there is sufficient space elsewhere in your file system for the data to be stored by using the **mmdf** command.

GPFS has declared NSDs built on top of AIX logical volumes as down

Earlier releases of GPFS allowed AIX logical volumes to be used in GPFS file systems. Using AIX logical volumes in GPFS file systems is now discouraged as they are limited with regard to their clustering ability and cross platform support.

Existing file systems using AIX logical volumes are however still supported, and this information might be of use when working with those configurations.

Verify whether the logical volumes are properly defined

Logical volumes must be properly configured to map between the NSD and the underlying disks.

To verify the logical volume configuration, issue the following command:

```
mmlsnsd -m
```

The system displays any underlying physical device present on this node, which is backing the NSD. If the underlying device is a logical volume, issue the following command to map from the logical volume to the volume group.

```
lsvg -o | lsvg -i -l
```

The system displays a list of logical volumes and corresponding volume groups. Now, issue the **lsvg** command for the volume group that contains the logical volume. For example:

```
lsvg gpfs1vg
```

The system displays information similar to the following example:

```
VOLUME GROUP:   gpfs1vg                 VG IDENTIFIER:  000195600004c00000000ee60c66352
VG STATE:       active                  PP SIZE:        16 megabyte(s)
VG PERMISSION:  read/write              TOTAL PPs:      542 (8672 megabytes)
MAX LVs:        256                     FREE PPs:       0 (0 megabytes)
LVs:            1                       USED PPs:       542 (8672 megabytes)
OPEN LVs:       1                       QUORUM:         2
TOTAL PVs:      1                       VG DESCRIPTORS: 2
STALE PVs:      0                       STALE PPs:      0
ACTIVE PVs:     1                       AUTO ON:        no
MAX PPs per PV: 1016                    MAX PVs:        32
LTG size:       128 kilobyte(s)         AUTO SYNC:      no
HOT SPARE:      no
```

Check the volume group on each node

All the disks in GPFS cluster has to be properly defined to all the nodes.

Make sure that all disks are properly defined to all nodes in the GPFS cluster:

1. Issue the AIX **lspv** command on all nodes in the GPFS cluster and save the output.
2. Compare the **pvid** and volume group fields for all GPFS volume groups.

   Each volume group must have the same **pvid** and volume group name on each node. The **hdisk** name for these disks may vary.

For example, to verify the volume group **gpfs1vg** on the five nodes in the GPFS cluster, for each node in the cluster issue:

```
lspv | grep gpfs1vg
```

The system displays information similar to:

```
k145n01:  hdisk3         00001351566acb07    gpfs1vg    active
k145n02:  hdisk3         00001351566acb07    gpfs1vg    active
k145n03:  hdisk5         00001351566acb07    gpfs1vg    active
k145n04:  hdisk5         00001351566acb07    gpfs1vg    active
k145n05:  hdisk7         00001351566acb07    gpfs1vg    active
```

Here the output shows that on each of the five nodes the volume group **gpfs1vg** is the same physical disk (has the same **pvid**). The **hdisk** numbers vary, but the fact that they may be called different **hdisk** names on different nodes has been accounted for in the GPFS product. This is an example of a properly defined volume group.

If any of the **pvid**s were different for the same volume group, this would indicate that the same volume group name has been used when creating volume groups on different physical volumes. This will not work for GPFS. A volume group name can be used only for the same physical volume shared among nodes in a cluster. For more information, refer to AIX in IBM Knowledge Center (www.ibm.com/support/knowledgecenter/ssw_aix/welcome) and search for *operating system and device management*.

Volume group varyon problems

Use **varyoffvg** command for the volume group at all nodes to correct **varyonvg** issues at the volume group layer.

If an NSD backed by an underlying logical volume will not come online to a node, it may be due to **varyonvg** problems at the volume group layer. Issue the **varyoffvg** command for the volume group at all nodes and restart GPFS. On startup, GPFS will **varyon** any underlying volume groups in proper sequence.

Disk accessing commands fail to complete due to problems with some non-IBM disks

Certain disk commands, such as **mmcrfs**, **mmadddisk**, **mmrpldisk**, **mmmount** and the operating system's **mount**, might issue the **varyonvg -u** command if the NSD is backed by an AIX logical volume.

For some non-IBM disks, when many **varyonvg -u** commands are issued in parallel, some of the AIX **varyonvg -u** invocations do not complete, causing the disk command to hang.

This situation is recognized by the GPFS disk command not completing after a long period of time, and the persistence of the **varyonvg** processes as shown by the output of the **ps -ef** command on some of the nodes of the cluster. In these cases, **kill** the **varyonvg** processes that were issued by the GPFS disk command on the nodes of the cluster. This allows the GPFS disk command to complete. Before mounting the affected file system on any node where a **varyonvg** process was killed, issue the **varyonvg -u** command (**varyonvg -u** *vgname*) on the node to make the disk available to GPFS. Do this on each of the nodes in question, one by one, until all of the GPFS volume groups are varied online.

Disk media failure

Recovery procedures to recover lost data in case of disk media failure.

Regardless of whether you have chosen additional hardware or replication to protect your data against media failures, you first need to determine that the disk has completely failed. If the disk has completely failed and it is not the path to the disk which has failed, follow the procedures defined by your disk vendor. Otherwise:

1. Check on the states of the disks for the file system:

   ```
   mmlsdisk fs1 -e
   ```

   GPFS will mark disks **down** if there have been problems accessing the disk.

2. To prevent any I/O from going to the down disk, issue these commands *immediately*:

   ```
   mmchdisk fs1 suspend -d gpfs1nsd
   mmchdisk fs1 stop -d gpfs1nsd
   ```

   Note: If there are any GPFS file systems with pending I/O to the down disk, the I/O will timeout if the system administrator does not stop it.

   To see if there are any threads that have been waiting a long time for I/O to complete, on all nodes issue:

   ```
   mmfsadm dump waiters 10 | grep "I/O completion"
   ```

3. The next step is *irreversible*! Do not run this command unless data and metadata have been replicated. This command scans file system metadata for disk addresses belonging to the disk in question, then replaces them with a special "broken disk address" value, which may take a while.

   CAUTION:
   Be extremely careful with using the -p option of mmdeldisk, because by design it destroys references to data blocks, making affected blocks unavailable. This is a last-resort tool, to be used when data loss may have already occurred, to salvage the remaining data–which means it cannot take any precautions. If you are not absolutely certain about the state of the file system and the impact of running this command, do not attempt to run it without first contacting the IBM Support Center.

   ```
   mmdeldisk fs1 gpfs1n12 -p
   ```

4. Invoke the **mmfileid** command with the operand **:BROKEN**:

   ```
   mmfileid fs1 -d :BROKEN
   ```

   For more information, see "The mmfileid command" on page 264.

5. After the disk is properly repaired and available for use, you can add it back to the file system.

Replicated metadata and data

The course of actions to be followed to recover the lost files if you have replicated metadata and data and only disks in a single failure group has failed.

If you have replicated metadata and data and only disks in a single failure group have failed, everything should still be running normally but with slightly degraded performance. You can determine the replication values set for the file system by issuing the **mmlsfs** command. Proceed with the appropriate course of action:

1. After the failed disk has been repaired, issue an **mmadddisk** command to add the disk to the file system:

   ```
   mmadddisk fs1 gpfs12nsd
   ```

   You can rebalance the file system at the same time by issuing:

   ```
   mmadddisk fs1 gpfs12nsd -r
   ```

   Note: Rebalancing of files is an I/O intensive and time consuming operation, and is important only for file systems with large files that are mostly invariant. In many cases, normal file update and creation will rebalance your file system over time, without the cost of the rebalancing.

2. To re-replicate data that only has single copy, issue:

   ```
   mmrestripefs fs1 -r
   ```

   Optionally, use the **-b** flag instead of the **-r** flag to rebalance across all disks.

   Note: Rebalancing of files is an I/O intensive and time consuming operation, and is important only for file systems with large files that are mostly invariant. In many cases, normal file update and creation will rebalance your file system over time, without the cost of the rebalancing.

3. Optionally, check the file system for metadata inconsistencies by issuing the offline version of **mmfsck**:

   ```
   mmfsck fs1
   ```

   If **mmfsck** succeeds, you may still have errors that occurred. Check to verify no files were lost. If files containing user data were lost, you will have to restore the files from the backup media.

   If **mmfsck** fails, sufficient metadata was lost and you need to recreate your file system and restore the data from backup media.

Replicated metadata only

Using replicated metadata for lost data recovery.

If you have only replicated metadata, you should be able to recover some, but not all, of the user data. Recover any data to be kept using normal file operations or erase the file. If you read a file in block-size chunks and get a failure return code and an **EIO** errno, that block of the file has been lost. The rest of the file may have useful data to recover, or it can be erased.

Strict replication

Use **mmchfs -K no** command to perform disk action for strict replication.

If data or metadata replication is enabled, and the status of an existing disk changes so that the disk is no longer available for block allocation (if strict replication is enforced), you may receive an **errno** of ENOSPC when you create or append data to an existing file. A disk becomes unavailable for new block allocation if it is being deleted, replaced, or it has been suspended. If you need to delete, replace, or suspend a disk, and you need to write new data while the disk is offline, you can disable strict replication by issuing the **mmchfs -K no** command before you perform the disk action. However, data written while replication is disabled will not be replicated properly. Therefore, after you perform the disk action, you must re-enable strict replication by issuing the **mmchfs -K** command with the original value of the **-K** option (**always** or **whenpossible**) and then run the **mmrestripefs -r** command. To determine if a disk has strict replication enforced, issue the **mmlsfs -K** command.

Note: A disk in a **down** state that has not been explicitly suspended is still available for block allocation, and thus a spontaneous disk failure will not result in application I/O requests failing with ENOSPC. While new blocks will be allocated on such a disk, nothing will actually be written to the disk until its availability changes to **up** following an **mmchdisk start** command. Missing replica updates that took place while the disk was down will be performed when **mmchdisk start** runs.

No replication

Perform unmounting yourself if no replication has been done and the system metadata has been lost. You can follow the course of actions for manual unmounting.

When there is no replication, the system metadata has been lost and the file system is basically irrecoverable. You may be able to salvage some of the user data, but it will take work and time. A forced unmount of the file system will probably already have occurred. If not, it probably will very soon if you try to do any recovery work. You can manually force the unmount yourself:

1. Mount the file system in **read-only** mode (see "Read-only mode mount" on page 255). This will bypass recovery errors and let you read whatever you can find. Directories may be lost and give errors, and parts of files will be missing. Get what you can now, for all will soon be gone. On a single node, issue:

   ```
   mount -o ro /dev/fs1
   ```

2. If you read a file in block-size chunks and get an **EIO** return code that block of the file has been lost. The rest of the file may have useful data to recover or it can be erased. To save the file system parameters for recreation of the file system, issue:

   ```
   mmlsfs fs1   > fs1.saveparms
   ```

Note: This next step is *irreversible*!

To delete the file system, issue:

```
mmdelfs fs1
```

3. To repair the disks, see your disk vendor problem determination guide. Follow the problem determination and repair actions specified.
4. Delete the affected NSDs. Issue:

   ```
   mmdelnsd nsdname
   ```

   The system displays output similar to this:

   ```
   mmdelnsd: Processing disk nsdname
   mmdelnsd: 6027-1371 Propagating the cluster configuration data to all
     affected nodes.  This is an asynchronous process.
   ```
5. Create a disk descriptor file for the disks to be used. This will include recreating NSDs for the new file system.
6. Recreate the file system with either different parameters or the same as you used before. Use the disk descriptor file.
7. Restore lost data from backups.

GPFS error messages for disk media failures

There are some GPFS error messages associated with disk media failures.

Disk media failures can be associated with these GPFS message numbers:

6027-418
Inconsistent file system quorum. readQuorum=*value* writeQuorum=*value* quorumSize=*value*

6027-482 [E]
Remount failed for device *name*: *errnoDescription*

6027-485
Perform **mmchdisk** for any disk failures and re-mount.

6027-636 [E]
Disk marked as stopped or offline.

Error numbers specific to GPFS application calls when disk failure occurs

There are certain error numbers associated with GPFS application calls when disk failure occurs.

When a disk failure has occurred, GPFS may report these error numbers in the operating system error log, or return them to an application:

EOFFLINE = 208, Operation failed because a disk is offline
This error is most commonly returned when an attempt to open a disk fails. Since GPFS will attempt to continue operation with failed disks, this will be returned when the disk is first needed to complete a command or application request. If this return code occurs, check your disk for stopped states, and check to determine if the network path exists.

To repair the disks, see your disk vendor problem determination guide. Follow the problem determination and repair actions specified.

ENO_MGR = 212, The current file system manager failed and no new manager could be appointed.
This error usually occurs when a large number of disks are unavailable or when there has been a major network failure. Run the **mmlsdisk** command to determine whether disks have failed. If disks have failed, check the operating system error log on all nodes for indications of errors. Take corrective action by issuing the **mmchdisk** command.

To repair the disks, see your disk vendor problem determination guide. Follow the problem determination and repair actions specified.

Persistent Reserve errors

You can use Persistent Reserve (PR) to provide faster failover times between disks that support this feature. PR allows the stripe group manager to "fence" disks during node failover by removing the reservation keys for that node. In contrast, non-PR disk failovers cause the system to wait until the disk lease expires.

GPFS allows file systems to have a mix of PR and non-PR disks. In this configuration, GPFS will fence PR disks for node failures and recovery and non-PR disk will use disk leasing. If all of the disks are PR disks, disk leasing is not used, so recovery times improve.

GPFS uses the **mmchconfig** command to enable PR. Issuing this command with the appropriate **usePersistentReserve** option configures disks automatically. If this command fails, the most likely cause is either a hardware or device driver problem. Other PR-related errors will probably be seen as file system unmounts that are related to disk reservation problems. This type of problem should be debugged with existing trace tools.

Understanding Persistent Reserve

The AIX server displays the value of *reserve_policy* and *PR_key_value* for Persistent Reserve. Use **chdev** command to set the values for *reserve_policy* and *PR_key_value*.

Note: While Persistent Reserve (PR) is supported on both AIX and Linux, *reserve_policy* is applicable only to AIX.

Persistent Reserve refers to a set of Small Computer Systems Interface-3 (SCSI-3) standard commands and command options. These PR commands and command options give SCSI initiators the ability to establish, preempt, query, and reset a reservation policy with a specified target disk. The functions provided by PR commands are a superset of current reserve and release mechanisms. These functions are not compatible with legacy reserve and release mechanisms. Target disks can only support reservations from either the legacy mechanisms or the current mechanisms.

Note: Attempting to mix Persistent Reserve commands with legacy reserve and release commands will result in the target disk returning a reservation conflict error.

Persistent Reserve establishes an interface through a *reserve_policy* attribute for SCSI disks. You can optionally use this attribute to specify the type of reservation that the device driver will establish before accessing data on the disk. For devices that do not support the *reserve_policy* attribute, the drivers will use the value of the *reserve_lock* attribute to determine the type of reservation to use for the disk. GPFS supports four values for the *reserve_policy* attribute:

no_reserve::
Specifies that no reservations are used on the disk.

single_path::
Specifies that legacy reserve/release commands are used on the disk.

PR_exclusive::
Specifies that Persistent Reserve is used to establish exclusive host access to the disk.

PR_shared::
Specifies that Persistent Reserve is used to establish shared host access to the disk.

Persistent Reserve support affects both the parallel (scdisk) and SCSI-3 (scsidisk) disk device drivers and configuration methods. When a device is opened (for example, when the **varyonvg** command opens the

underlying **hdisks**), the device driver checks the ODM for *reserve_policy* and *PR_key_value* and then opens the device appropriately. For PR, each host attached to the shared disk must use unique registration key values for *reserve_policy* and *PR_key_value*. On AIX, you can display the values assigned to *reserve_policy* and *PR_key_value* by issuing:

```
lsattr -El hdiskx -a reserve_policy,PR_key_value
```

If needed, use the AIX **chdev** command to set *reserve_policy* and *PR_key_value*.

Note: GPFS manages *reserve_policy* and *PR_key_value* using `reserve_policy=PR_shared` when Persistent Reserve support is enabled and `reserve_policy=no_reserve` when Persistent Reserve is disabled.

Checking Persistent Reserve

For Persistent Reserve to function properly, follow the course of actions to determine the PR status.

For Persistent Reserve to function properly, you must have PR enabled on all of the disks that are PR-capable. To determine the PR status in the cluster:

1. Determine if PR is enabled on the cluster
   a. Issue **mmlsconfig**
   b. Check for `usePersistentReserve=yes`
2. Determine if PR is enabled for all disks on all nodes
   a. Make sure that GPFS has been started and mounted on all of the nodes
   b. Enable PR by issuing **mmchconfig**
   c. Issue the command **mmlsnsd -X** and look for `pr=yes` on all the hdisk lines

Notes:

1. To view the keys that are currently registered on a disk, issue the following command from a node that has access to the disk:

   ```
   /usr/lpp/mmfs/bin/tsprreadkeys hdiskx
   ```

2. To check the AIX ODM status of a single disk on a node, issue the following command from a node that has access to the disk:

   ```
   lsattr -El hdiskx -a reserve_policy,PR_key_value
   ```

Clearing a leftover Persistent Reserve reservation

You can clear leftover Persistent Reserve reservation.

Message number **6027-2202** indicates that a specified disk has a SCSI-3 PR reservation, which prevents the **mmcrnsd** command from formatting it. The following example is specific to a Linux environment. Output on AIX is similar but not identical.

Before trying to clear the PR reservation, use the following instructions to verify that the disk is really intended for GPFS use. Note that in this example, the device name is specified without a prefix (**/dev/sdp** is specified as **sdp**).

1. Display all the registration key values on the disk:

   ```
   /usr/lpp/mmfs/bin/tsprreadkeys sdp
   ```

   The system displays information similar to:

   ```
   Registration keys for sdp
   1. 00006d0000000001
   ```

   If the registered key values all start with `0x00006d`, which indicates that the PR registration was issued by GPFS, proceed to the next step to verify the SCSI-3 PR reservation type. Otherwise, contact your system administrator for information about clearing the disk state.

2. Display the reservation type on the disk:
   ```
   /usr/lpp/mmfs/bin/tsprreadres sdp
   ```

   The system displays information similar to:
   ```
   yes:LU_SCOPE:WriteExclusive-AllRegistrants:0000000000000000
   ```

If the output indicates a PR reservation with type **WriteExclusive-AllRegistrants**, proceed to the following instructions for clearing the SCSI-3 PR reservation on the disk.

If the output does not indicate a PR reservation with this type, contact your system administrator for information about clearing the disk state.

To clear the SCSI-3 PR reservation on the disk, follow these steps:

1. Choose a hex value (*HexValue*); for example, **0x111abc** that is not in the output of the **tsprreadkeys** command run previously. Register the local node to the disk by entering the following command with the chosen *HexValue*:
   ```
   /usr/lpp/mmfs/bin/tsprregister sdp 0x111abc
   ```
2. Verify that the specified *HexValue* has been registered to the disk:
   ```
   /usr/lpp/mmfs/bin/tsprreadkeys sdp
   ```

   The system displays information similar to:
   ```
   Registration keys for sdp
   1. 00006d0000000001
   2. 0000000000111abc
   ```
3. Clear the SCSI-3 PR reservation on the disk:
   ```
   /usr/lpp/mmfs/bin/tsprclear sdp 0x111abc
   ```
4. Verify that the PR registration has been cleared:
   ```
   /usr/lpp/mmfs/bin/tsprreadkeys sdp
   ```

   The system displays information similar to:
   ```
   Registration keys for sdp
   ```
5. Verify that the reservation has been cleared:
   ```
   /usr/lpp/mmfs/bin/tsprreadres sdp
   ```

   The system displays information similar to:
   ```
   no:::
   ```

The disk is now ready to use for creating an NSD.

Manually enabling or disabling Persistent Reserve

The PR status can be set manually with the help of IBM Support Center.

Attention: Manually enabling or disabling Persistent Reserve should only be done under the supervision of the IBM Support Center with GPFS stopped on the node.

The IBM Support Center will help you determine if the PR state is incorrect for a disk. If the PR state is incorrect, you may be directed to correct the situation by manually enabling or disabling PR on that disk.

GPFS is not using the underlying multipath device

You can view the underlying disk device where I/O is performed on an NSD disk by using the **mmlsdisk** command with the **-M** option.

The **mmlsdisk** command output might show unexpected results for multipath I/O devices. For example if you issue this command:

```
mmlsdisk dmfs2 -M
```

The system displays information similar to:

```
Disk name     IO performed on node     Device             Availability
------------  -----------------------  -----------------  ------------
m0001         localhost                /dev/sdb           up
```

The following command is available on Linux only.

```
# multipath -ll
mpathae (36005076304ffc0e5000000000000001) dm-30 IBM,2107900
[size=10G][features=1 queue_if_no_path][hwhandler=0]
\_ round-robin 0 [prio=8][active]
 \_ 1:0:5:1  sdhr 134:16  [active][ready]
 \_ 1:0:4:1  sdgl 132:16  [active][ready]
 \_ 1:0:1:1  sdff 130:16  [active][ready]
 \_ 1:0:0:1  sddz 128:16  [active][ready]
 \_ 0:0:7:1  sdct 70:16   [active][ready]
 \_ 0:0:6:1  sdbn 68:16   [active][ready]
 \_ 0:0:5:1  sdah 66:16   [active][ready]
 \_ 0:0:4:1  sdb  8:16    [active][ready]
```

The **mmlsdisk** output shows that I/O for NSD **m0001** is being performed on disk **/dev/sdb**, but it should show that I/O is being performed on the device-mapper multipath (DMM) **/dev/dm-30**. Disk **/dev/sdb** is one of eight paths of the DMM **/dev/dm-30** as shown from the **multipath** command.

This problem could occur for the following reasons:

- The previously installed user exit **/var/mmfs/etc/nsddevices** is missing. To correct this, restore user exit **/var/mmfs/etc/nsddevices** and restart GPFS.
- The multipath device type does not match the GPFS known device type. For a list of known device types, see **/usr/lpp/mmfs/bin/mmdevdiscover**. After you have determined the device type for your multipath device, use the **mmchconfig** command to change the NSD disk to a known device type and then restart GPFS.

The following output shows that device type **dm-30** is **dmm**:

```
/usr/lpp/mmfs/bin/mmdevdiscover | grep dm-30
dm-30 dmm
```

To change the NSD device type to a known device type, create a file that contains the NSD name and device type pair (one per line) and issue this command:

```
mmchconfig updateNsdType=/tmp/filename
```

where the contents of **/tmp/filename** are:

```
m0001 dmm
```

The system displays information similar to:

```
mmchconfig: Command successfully completed
mmchconfig: Propagating the cluster configuration data to all
affected nodes. This is an asynchronous process.
```

Kernel panics with a 'GPFS dead man switch timer has expired, and there are still outstanding I/O requests' message

This problem can be detected by an error log with a label of KERNEL_PANIC, and the PANIC MESSAGES or a PANIC STRING.

For example:

```
GPFS Deadman Switch timer has expired, and there are still outstanding I/O requests
```

GPFS is designed to tolerate node failures through per-node metadata logging (journaling). The log file is called the *recovery log*. In the event of a node failure, GPFS performs recovery by replaying the recovery log for the failed node, thus restoring the file system to a consistent state and allowing other nodes to continue working. Prior to replaying the recovery log, it is critical to ensure that the failed node has indeed failed, as opposed to being active but unable to communicate with the rest of the cluster.

In the latter case, if the failed node has direct access (as opposed to accessing the disk with an NSD server) to any disks that are a part of the GPFS file system, it is necessary to ensure that no I/O requests submitted from this node complete once the recovery log replay has started. To accomplish this, GPFS uses the disk lease mechanism. The disk leasing mechanism guarantees that a node does not submit any more I/O requests once its disk lease has expired, and the surviving nodes use disk lease time out as a guideline for starting recovery.

This situation is complicated by the possibility of 'hung I/O'. If an I/O request is submitted prior to the disk lease expiration, but for some reason (for example, device driver malfunction) the I/O takes a long time to complete, it is possible that it may complete after the start of the recovery log replay during recovery. This situation would present a risk of file system corruption. In order to guard against such a contingency, when I/O requests are being issued directly to the underlying disk device, GPFS initiates a kernel timer, referred to as **dead man switch**. The **dead man switch** timer goes off in the event of disk lease expiration, and checks whether there is any outstanding I/O requests. If there is any I/O pending, a kernel panic is initiated to prevent possible file system corruption.

Such a kernel panic is not an indication of a software defect in GPFS or the operating system kernel, but rather it is a sign of

1. Network problems (the node is unable to renew its disk lease).
2. Problems accessing the disk device (I/O requests take an abnormally long time to complete). See "MMFS_LONGDISKIO" on page 216.

Chapter 20. Security issues

This topic describes some security issues that you might encounter while using IBM Spectrum Scale.

Encryption issues

The topics that follow provide solutions for problems that may be encountered while setting up or using encryption.

Unable to add encryption policies

If the **mmchpolicy** command fails when you are trying to add encryption policies, perform the following diagnostic steps:

1. Confirm that the `gpfs.crypto` and `gpfs.gskit` packages are installed.
2. Confirm that the file system is at GPFS 4.1 or later and the fast external attributes (**--fastea**) option is enabled.
3. Examine the error messages that are logged in the `mmfs.log.latest` file, which is located at:`/var/adm/ras/mmfs.log.latest`.

Receiving "Permission denied" message

If you experience a "Permission denied" failure while creating, opening, reading, or writing to a file, perform the following diagnostic steps:

1. Confirm that the key server is operational and correctly set up and can be accessed through the network.
2. Confirm that the `/var/mmfs/etc/RKM.conf` file is present on all nodes from which the file is supposed to be accessed. The `/var/mmfs/etc/RKM.conf` file must contain entries for all the RKMs needed to access the file.
3. Verify that the master keys needed by the file and the keys that are specified in the encryption policies are present on the key server.
4. Examine the error messages in the `/var/adm/ras/mmfs.log.latest` file.

"Value too large" failure when creating a file

If you experience a "Value too large to be stored in data type" failure when creating a file, follow these diagnostic steps.

1. Examine error messages in `/var/adm/ras/mmfs.log.latest` to confirm that the problem is related to the extended attributes being too large for the inode. The size of the encryption extended attribute is a function of the number of keys used to encrypt a file. If you encounter this issue, update the encryption policy to reduce the number of keys needed to access any given file.
2. If the previous step does not solve the problem, create a new file system with a larger inode size.

Mount failure for a file system with encryption rules

If you experience a mount failure for a file system with encryption rules, follow these diagnostic steps.

1. Confirm that the `gpfs.crypto` and `gpfs.gskit` packages are installed.
2. Confirm that the `/var/mmfs/etc/RKM.conf` file is present on the node and that the content in `/var/mmfs/etc/RKM.conf` is correct.
3. Examine the error messages in `/var/adm/ras/mmfs.log.latest`.

"Permission denied" failure of key rewrap

If you experience a "Permission denied" failure of a key rewrap, follow these diagnostic steps.

When **mmapplypolicy** is invoked to perform a key rewrap, the command may issue messages like the following:

```
[E] Error on gpfs_enc_file_rewrap_key(/fs1m/sls/test4,KEY-d7bd45d8-9d8d-4b85-a803-e9b794ec0af2:hs21n56_new,KEY-40a0b68b-c86d-4519-9e48-3714d3b71e20:js21n92)
 Permission denied(13)
```

If you receive a message similar to this, follow these steps:

1. Check for syntax errors in the migration policy syntax.
2. Ensure that the new key is not already being used for the file.
3. Ensure that both the original and the new keys are retrievable.
4. Examine the error messages in `/var/adm/ras/mmfs.log.latest` for additional details.

Authentication issues

This topic describes the authentication issues that you might experience while using file and object protocols.

File protocol authentication setup issues

When trying to enable Active Directory Authentication for file (SMB, NFS), the operation might fail due to a timeout. In some cases, the AD server can return multiple IPs that cannot be queried within the allotted timeout period and/or IPs that belong to networks inaccessible by the IBM Spectrum Scale nodes.

You can try the following workarounds to resolve this issue:

- Remove any invalid/unreachable IPs from the AD DNS.

  If you removed any invalid/unreachable IPs, retry the **mmuserauth service create** command that previously failed.
- You can also try to disable any adapters that might not be in use.

  For example, on Windows 2008: **Start** -> **Control Panel** -> **Network and Sharing Center** -> **Change adapter settings** -> Right-click the adapter that you are trying to disable and click **Disable**

  If you disabled any adapters, retry the **mmuserauth service create** command that previously failed.

Protocol authentication issues

You can use a set of GPFS commands to identify and rectify issues that are related to authentication configurations.

To do basic authentication problem determination, perform the following steps:

1. Issue the **mmces state show auth** command to view the current state of authentication.
2. Issue the **mmces events active auth** command to see whether events are currently contributing to make the state of the authentication component unhealthy.
3. Issue the **mmuserauth service list** command to view the details of the current authentication configuration.
4. Issue the **mmuserauth service check -N cesNodes --server-reachability** command to verify the state of the authentication configuration across the cluster.
5. Issue the **mmuserauth service check -N cesNodes --rectify** command to rectify the authentication configuration.

   Note: Server reachability cannot be rectified by using the **--rectify** parameter.

Authentication error events

This topic describes how to verify and resolve Authentication errors.

Following is a list of possible events that may cause a node to go into a failed state and possible solutions for each of the issues. To determine what state a component is in, issue the **mmces command.**

SSD/YPBIND process not running (sssd_down)

Cause

The SSSD or the YPBIND process is not running.

Determination

To learn the authentication current state, run the following command:

```
mmces state show auth
```

To check the active events for authentication, run the following command:

```
mmces events active auth
```

To check the current authentication state, run the following command:

```
mmces state show auth
```

To check the current authentication configuration, run the following command:

```
mmuserauth service list
```

To check the current authentication configuration across the cluster, run the following command:

```
mmuserauth service check -N cesNodes --server-reachability
```

Solution

Rectify the configuration by running the following command:

```
mmuserauth service check -N cesNodes --rectify
```

Note: Server reachability cannot be rectified by using the --rectify flag.

Winbind process not running (wnbd_down)

Cause

The Winbind process is not running.

Determination

Run the same command as recommended in the section above, SSD/YPBIND process not running (sssd_down).

Solution

Follow the steps in the previous section, SSD/YPBIND process not running (sssd_down). Then, run the following command:

```
mmces service stop smb -N <Node on which the problem exists>
mmces service start smb -N <Node on which the problem existed>
```

Authorization issues

You might receive an unexpected "access denied" error either for native access to file system or for using the SMB or NFS protocols. Possible steps for troubleshooting the issue are described here.

Note: ACLs used in the object storage protocols are separate from the file system ACLs, and troubleshooting in that area should be done differently. For more information, see "Object issues" on page 392.

Verify authentication and ID mapping information

As a first step, verify that authentication and ID mapping are correctly configured. For more information, see the *Verifying the authentication services configured in the system* topic in the *IBM Spectrum Scale: Administration Guide*.

Verify authorization limitations

Ensure that Access Control Lists (ACLs) are configured as required by IBM Spectrum Scale. For more information, see the *Authorization limitation* topic in the *IBM Spectrum Scale: Administration Guide*. Also, check for more limitations of the NFSv4 ACLs stored in the file system. For more information, see the *GPFS exceptions and limitations to NFS V4 ACLs* topic in the *IBM Spectrum Scale: Administration Guide*.

Verify stored ACL of file or directory

Read the native ACL stored in the file system by using this command:

```
mmgetacl -k native /path/to/file/or/directory
```

If the output does not report an NFSv4 ACL type in the first line, consider changing the ACL to the NFSv4 type. For more information on how to configure the file system for the recommended NFSv4 ACL type for protocol usage, see the *Authorizing file protocol users* topic in the *IBM Spectrum Scale: Administration Guide*. Also, review the ACL entries for permissions related to the observed "access denied" issue.

Note: ACL entries are evaluated in the listed order for determining whether access is granted, and that the evaluation stops when a "deny" entry is encountered. Also, check for entries that are flagged with "InheritOnly", since they do not apply to the permissions of the current file or directory.

Verify group memberships and ID mappings

Next review the group membership of the user and compare that to the permissions granted in the ACL. If the cluster is configured with Active Directory authentication, first have the user authenticate and then check the group memberships of the user. With Active Directory, authentication is the only reliable way to refresh the group memberships of the user if the cluster does not have the latest and complete list of group memberships:

```
/usr/lpp/mmfs/bin/wbinfo -a 'domainname\username'
 id 'domainname\username'
```

If the cluster is configured with a different authentication method, query the group membership of the user:

```
id 'username'
```

If the user is a member of many groups, compare the number of group memberships with the limitations that are listed in the IBM Spectrum Scale FAQ. For more information, see https://www.ibm.com/support/knowledgecenter/en/STXKQY/gpfsclustersfaq.html#group.

If a group is missing, check the membership of the user in the missing group in the authentication server. Also, check the ID mapping configuration for that group and check whether the group has an ID mapping that is configured and if it is in the correct range. You can query the configured ID mapping ranges by using this command:

```
/usr/lpp/mmfs/bin/mmuserauth service list
```

If the expected groups are missing in the output from the ID command and the authentication method is Active Directory with trusted domains, check the types of the groups in Active Directory. Not all group types can be used in all Active Directory domains.

If the access issue is sporadic, repeat the test on all protocol nodes. Since authentication and ID mapping is handled locally on each protocol node, it might happen that a problem affects only one protocol node, and hence only protocol connections that are handled on that protocol node are affected.

Verify SMB export ACL for SMB export

If the access issue occurs on an SMB export, consider that the SMB export ACL can also cause user access to be denied. Query the current SMB export ACLs and review whether they are set up as expected by using this command:

```
/usr/lpp/mmfs/bin/mmsmb exportacl list
```

Collect trace for debugging

Collect traces as a last step to determine the cause for authorization issues. When the access problem occurs for a user using the SMB protocol, capture the SMB trace first while recreating the problem (the parameter -c is used to specify the IP address of the SMB):

```
/usr/lpp/mmfs/bin/mmprotocoltrace start smb -c x.x.x.x
```

Re-create the access denied issue

```
/usr/lpp/mmfs/bin/mmprotocoltrace stop smb
```

For analyzing the trace, extract the trace and look for the error code NT_STATUS_ACCESS_DENIED in the trace.

If the access issue occurs outside of SMB, collect a file system trace:

```
/usr/lpp/mmfs/bin/mmtracectl –start
```

Re-create the access denied issue

```
/usr/lpp/mmfs/bin/mmtracectl --stop
```

The IBM Security Lifecycle Manager prerequisites cannot be installed

This topic provides troubleshooting references and steps for resolving system errors when the IBM Security Lifecycle Manager prerequisites cannot be installed.

Description

When the user tries to install the IBM Security Lifecycle Manager prerequisites, the system displays the following error:

```
JVMJ9VM011W Unable to load j9dmp24: libstdc++.so.5: cannot open shared
object file: No such file or directory
JVMJ9VM011W Unable to load j9jit24: libstdc++.so.5: cannot open shared
object file: No such file or directory
```

```
JVMJ9VM011W Unable to load j9gc24: libstdc++.so.5: cannot open shared
object file: No such file or directory
JVMJ9VM011W Unable to load j9vrb24: libstdc++.so.5: cannot open shared
object file: No such file or directory
```

Cause

The system displays this error when the system packages are not upgraded.

Proposed workaround

- All system packages must be upgraded, except the kernel that should be 6.3 in order for encryption to work correctly.
- Update all packages excluding kernel:

  ```
  yum update --exclude=kernel*
  ```
- Modify: /etc/yum.conf

  ```
  [main]
  ...
  exclude=kernel* redhat-release*
  ```

IBM Security Lifecycle Manager cannot be installed

This topic provides troubleshooting references and steps for resolving system errors when IBM Security Lifecycle Manager cannot be installed.

Description

When the user tries to install IBM Security Lifecycle Manager, the system displays the following errors:

```
eclipse.buildId=unknownjava.fullversion=JRE 1.6.0 IBM J9 2.4 Linux x86-32
jvmxi3260sr9-20110203_74623 (JIT enabled, AOT enabled)J9VM -
20110203_074623JIT - r9_20101028_17488ifx3GC - 20101027_AABootLoader
constants: OS=linux, ARCH=x86, WS=gtk, NL=enFramework arguments: -toolId
install -accessRights admin input @osgi.install.area/install.xmlCommand-
line arguments: -os linux -ws gtk -arch x86 -toolId install -accessRights
admin input @osgi.install.area/install.xml!ENTRY com.ibm.cic.agent.ui 4 0
2013-07-09 14:11:47.692!MESSAGE Could not load SWT library.
Reasons:/home/tklm-v3/disk1/im/configuration/org.eclipse.osgi/bundles/207/1/
.cp/libswt-pi-gtk-4234.so (libgtk-x11-2.0.so.0: cannot open shared object
file: No such file or directory)
swt-pi-gtk (Not found in java.library.path)/root/.swt/lib/linux/x86/libswt-
pi-gtk-4234.so (libgtk-x11-2.0.so.0: cannot open shared object file: No
such file or directory)
/root/.swt/lib/linux/x86/libswt-pi-gtk.so (/root/.swt/lib/linux/x86/liblib
swt-pi-gtk.so.so:cannot open shared object file: No such file or directory)"
```

Cause

The system displays this error when the system packages are not upgraded.

Proposed workaround

- All system packages must be upgraded, except the kernel that should be 6.3 in order for encryption to work correctly.
- Run through the following checklist before installing IBM Security Lifecycle Manager:

Table 55.

System components	Minimum values	Header
System memory (RAM)	4 GB	4 GB

Table 55. (continued)

System components	Minimum values	Header
Processor speed	Linux and Windows systems	Linux and Windows systems 3.0 GHz dual processors AIX and Sun Solaris systems 1.5 GHz (4-way)
Disk space free for IBM Security Key Lifecycle Manager and prerequisite products such as DB2®	3.0 GHz single processor AIX and Sun Solaris systems 1.5 GHz (2-way)	5 GB
Disk space free in /tmp or C:\temp	5 GB	2 GB
Disk space free in /home directory for DB2	2 GB	6 GB
Disk space free in /var directory for DB2	5 GB 512 MB on Linux and UNIX operating systems	512 MB on Linux and UNIX operating systems

Chapter 21. Protocol issues

This topic describes the protocol-related issues (NFS, SMB, and Object) that you might come across while using IBM Spectrum Scale.

NFS issues

This topic describes some of the possible problems that can be encountered when GPFS interacts with NFS.

For details on how GPFS and NFS interact, see the *NFS and GPFS* topic in the *IBM Spectrum Scale: Administration Guide*.

These are some of the problems encountered when GPFS interacts with NFS:

- "NFS client with stale inode data"
- "NFSV4 problems"

CES NFS failure due to network failure

This topic provides information on how to resolve a CES NFS failure caused by a network failure.

When a network failure occurs because a cable is disconnected, a switch fails, or an adapter fails, CES NFS I/O operations will not complete. To resolve the failure, run the **systemctl restart network** command on the CES node to which the IP is failing back (where the failure occurred). This clears the client suspension and refreshes the network.

NFS client with stale inode data

The NFS client may have stale inode data due to caching and the course of action to be followed to correct this issue.

For performance reasons, some NFS implementations cache file information on the client. Some of the information (for example, file state information such as file size and timestamps) is not kept up-to-date in this cache. The client may view stale inode data (on **ls -l**, for example) if exporting a GPFS file system with NFS. If this is not acceptable for a given installation, caching can be turned off by mounting the file system on the client using the appropriate operating system **mount** command option (for example, **-o noac** on Linux NFS clients).

Turning off NFS caching will result in extra file systems operations to GPFS, and negatively affect its performance.

The clocks of all nodes in the GPFS cluster must be synchronized. If this is not done, NFS access to the data, as well as other GPFS file system operations, may be disrupted. NFS relies on metadata timestamps to validate the local operating system cache. If the same directory is either NFS-exported from more than one node, or is accessed with both the NFS and GPFS mount point, it is critical that clocks on all nodes that access the file system (GPFS nodes and NFS clients) are constantly synchronized using appropriate software (for example, NTP). Failure to do so may result in stale information seen on the NFS clients.

NFSV4 problems

The analysis of NFS V4 issues and suggestions to resolve these issues.

Before analyzing an NFS V4 problem, review this documentation to determine if you are using NFS V4 ACLs and GPFS correctly:

1. The *NFS Version 4 Protocol* paper and other related information that are available in the Network File System Version 4 (nfsv4) section of the IETF Datatracker website (datatracker.ietf.org/wg/nfsv4/documents).
2. The *Managing GPFS access control lists and NFS export* topic in the *IBM Spectrum Scale: Administration Guide*.
3. The *GPFS exceptions and limitations to NFS V4 ACLs* topic in the *IBM Spectrum Scale: Administration Guide*.

The commands **mmdelacl** and **mmputacl** can be used to revert an NFS V4 ACL to a traditional ACL. Use the **mmdelacl** command to remove the ACL, leaving access controlled entirely by the permission bits in the mode. Then use the **chmod** command to modify the permissions, or the **mmputacl** and **mmeditacl** commands to assign a new ACL.

For files, the **mmputacl** and **mmeditacl** commands can be used at any time (without first issuing the **mmdelacl** command) to assign any type of ACL. The command **mmeditacl -k posix** provides a translation of the current ACL into traditional POSIX form and can be used to more easily create an ACL to edit, instead of having to create one from scratch.

NFS mount issues

This topic provides information on how to verify and resolve NFS mount errors.

There are several possible NFS mount error conditions, including

- Mount times out
- NFS mount fails with a "No such file or directory" error
- NFS client cannot mount NFS exports.

Mount times out

Description

The user is trying to do an NFS mount and receives a timeout error.

Verification

When a timeout error occurs, check the following.

1. Check to see whether the server is reachable by issuing either or both of the following commands:

   ```
   ping <server-ip>
   ping <server-name>
   ```

   The expected result is that the server responds.
2. Check to see whether portmapper, NFS, and mount daemons are running on the server.
   a. On a IBM Spectrum Scale CES node, issue the following command:

      ```
      mmces service list
      ```

      The expected results are that the output indicates that the NFS service is running as in this example:

      ```
      Enabled services: SMB NFS
      SMB is running, NFS is running
      ```

   b. On the NFS server node, issue the following command:

      ```
      rpcinfo -p
      ```

      The expected result is that portmapper, mountd, and NFS are running as shown in the following sample output.

```
program  vers  proto  port  service
100000      4   tcp    111  portmapper
100000      4   tcp    111  portmapper
100000      3   tcp    111  portmapper
100000      2   tcp    111  portmapper
100000      4   upd    111  portmapper
100000      3   upd    111  portmapper
100000      2   upd    111  portmapper
100024      1   upd  53111  status
100024      1   tcp  58711  status
100003      3   upd   2049  nfs
100003      3   tcp   2049  nfs
100003      4   upd   2049  nfs
100003      4   tcp   2049  nfs
100005      1   upd  59149  mountd
100005      1   tcp  54013  mountd
100005      3   upd  59149  mountd
100005      3   tcp  54013  mountd
100021      4   upd  32823  nlockmgr
100021      4   tcp  33397  nlockmgr
100011      1   upd  36650  rquotad
100011      1   tcp  36673  rquotad
100011      2   upd  36650  rquotad
100011      2   tcp  36673  rquotad
```

3. Check to see whether the firewall is blocking NFS traffic on Linux systems by issuing the following command on the NFS client and the NFS server:

   ```
   iptables -L
   ```

   Then check whether any hosts or ports that are involved with the NFS connection are blocked (denied).

   If the client and the server are running in different subnets, then a firewall could be running on the router also.
4. Check to see whether the firewall is blocking NFS traffic on the client or router, using the appropriate commands.

NFS mount fails with a “No such file or directory” error

Description

The user is trying to do an NFS mount on Linux and receives this message:

```
No such file or directory
```

Following are the root causes of this error.

Root cause #1 - Access type is none

An NFS export was created on the server without a specified access type. Therefore, for security reasons, the default access is none, mounting does not work.

Solution

On the NFS server, specify an access type (for example, RW for Read and Write) for the export. If the export has been created already, you can achieve this by issuing the **mmnfs export change** command. See the following example. The backslash (\) is a line continuation character:

```
mmnfs export change /mnt/gpfs0/nfs_share1 \
--nfschange "*(Access_Type=RW,Squash=NO_ROOT_SQUASH)"
```

Verification

To verify the access type that is specified for the export, issue the **mmnfs export list** on the NFS server. For example:

```
mmnfs export list --nfsdefs /mnt/gpfs0/nfs_share1
```

The system displays output similar to this:

```
Path               Delegations Clients Access_Type Protocols Transports Squash         Anonymous_uid Anonymous_gid SecType PrivilegedPort Export_id DefaultDelegation Manage_Gids NFS_Commit
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
/mnt/gpfs0/_share1 none        *       RW          3,4       TCP        NO_ROOT_SQUASH -2            -2            KRB5     FALSE          2         none              FALSE       FALSE
```

"NONE" indicates the root cause; the access type is none .

"RO" or "RW" indicates that the solution was successful.

Root cause # 2 - Protocol version that is not supported by the server

Solution

On the NFS server, specify the protocol version needed by the client for export (for example, 3:4). If the export already exists, you can achieve this by issuing the **mmnfs export change command**. For example:

```
mmnfs export change /mnt/gpfs0/nfs_share1 --nfschange "* (Protocols=3:4)"
```

Verification

To verify the protocols that are specified for the export, issue the **mmnfs export change** command. For example:

```
mmnfs export list --nfsdefs /mnt/gpfs0/nfs_share1
```

The system displays output similar to this:

```
Path                Delegations Clients Access_Type Protocols Transports Squash         Anonymous_uid Anonymous_gid SecType PrivilegedPort DefaultDelegations Manage_Gids NFS_Commit
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
/mnt/gpfs0/nfs_share1 none        *       RW          3,4       TCP        NO_ROOT_SQUASH -2            -2            SYS      FALSE          none               FALSE       FALSE
```

NFS client cannot mount NFS exports

Problem

The NFS client cannot mount NFS exports. The **mount** command on the client either returns an error or times out.

Determination

The error itself occurs on the NFS client side. Additionally, and based on the nature of the problem, the server-side NFS logs can provide more details about the origin of the error.

Solution

These are the reasons for client-side mount errors:

- The NFS server is not running
- The firewall is blocking NFS traffic
- The client does not have permissions to mount the export.

1. Ensure that the NFS server is running correctly on all of the CES nodes and that the CES IP address used to mount is active in the CES cluster. To check the CES IP address and the NFS server status run:

   ```
   mmlscluster --ces
   mmces service list -a
   ```

2. Ensure that the firewall allows NFS traffic to pass through. In order for this, the CES NFS service must be configured with explicit NFS ports so that discrete firewall rules can be established. On the client, run:

   ```
   rpcinfo -t <CES_IP_ADDRESS> nfs
   ```

3. Verify that the NFS client is allowed to mount the export. In NFS terms, a definition exists for this client for the export to be mounted. To check NFS export details, enter the following command:

   ```
   mmnfs export list --nfsdefs <NFS_EXPORT_PATH>
   ```

   The system displays output similar to this:

```
Path                 Delegations Clients Access_Type Protocols Transports Squash             Anonymous_uid Anonymous_gid SecType PrivilegedPort DefaultDelegations Manage_Gids NFS_Commit
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
/mnt/gpfs0/nfs_share1 none        *       RW          3,4       TCP        NO_ROOT_SQUASH     -2            -2            SYS     FALSE          none               FALSE       FALSE
```

   On the client, run:

   ```
   showmount -e <CES_IP_ADDRESS>
   ```

NFS error events

This topic provides information on how to verify and resolve NFS errors.

Following is a list of possible events that might cause a node to go into a failed state and possible solutions for each of the issues. To determine what state a component is in, run the **mmces events active nfs** command.

NFS is not active (nfs_not_active)

Cause

Statistics query indicates that CES NFS is not responding.

Determination

Call the CES NFS statistics command with some delay and compare the NFS server time stamp, then determine if the NFS operation counts are increasing. Run this command:

```
/usr/bin/ganesha_stats ; sleep 5 ; /usr/bin/ganesha_stats
Timestamp: Wed Apr 27 19:27:22 201634711407 nsecs
Total NFSv3 ops: 0
Total NFSv4.0 ops: 86449
Total NFSv4.1 ops: 0
Total NFSv4.2 ops: 0
Timestamp: Wed Apr 27 19:27:27 201687146242 nsecs
Total NFSv3 ops: 0
Total NFSv4.0 ops: 105271
Total NFSv4.1 ops: 0
Total NFSv4.2 ops: 0
```

Solution

Restart CES NFS on the local CES node using commands **mmces service stop nfs** and **mmces service start nfs**.

CES NFSD process not running (nfsd_down)

Cause

CES NFS server protocol is no longer running.

Determination

1. Check to see whether the CES NFS daemon is running:
   ```
   ps -C gpfs.ganesha.nfsd
   ```
2. Check whether d-bus is alive. Run:
   ```
   /usr/bin/ganesha_stats
   ```

If either CES NFS or d-bus is down, you will receive an error:

```
ERROR: Can't talk to ganesha service on d-bus. Looks like Ganesh is down.
```

Solution

Restart CES NFS on the local CES node by using commands **mmces service stop nfs** and **mmces service start nfs**.

RPC statd process is not running (statd_down)

This applies only if NFS version 3 is enabled in the CES NFS configuration

Cause

The rpc.statd process is no longer running.

Determination

Check rpc.statd by running:

```
ps -C rpc.statd
```

Solution

Restart CES NFS on the local CES node by using commands **mmces service stop nfs** and **mmces service start nfs.**

Portmapper port 111 is not active (portmapper_down)

Cause

RPC call to port 111 failed or timed out.

Determination

Check portmapper output by running:

```
rpcinfo -n 111 -t localhost portmap
rpcinfo -t localhost nfs 3
rpcinfo -t localhost nfs 4
```

Solution

Check to see whether portmapper is running and if portmapper (rpcbind) is configured to automatically start on system startup.

NFS client cannot mount NFS exports from all protocol nodes

Cause

The NFS client can mount NFS exports from some but not all protocol nodes, because the exports are not seen when doing a **showmount** against those protocol nodes where this problem surfaces.

Determination

The error itself occurs on the NFS server side and is related to a Red Hat problem with netgroup caching which makes caching unreliable.

Solution

Disable caching netgroups in nscd for AD values. For more information on how to disable nscd caching, see the **nsd.conf** man page in https://linux.die.net/man/5/nscd.conf.

For more information on NFS events, see "Events" on page 473.

SELinux in enforcing mode on nodes servicing cNFS/kNFS can cause NFS lock and client monitoring to fail

Cause

If SELinux is configured in enforcing mode, `rpcstat` calls might lack permissions to modify `statd` client node configuration folders. This originates permission-denied responses.

Determination

Check messages similar to the following, in `syslog` on server nodes where cNFS/kNFS is enabled.

```
rpc.statd[<pid>]: Failed to insert: creating /var/lib/nfs/statd/sm/<client node>:Permission denied
rpc.statd[<pid>]: STAT_FAIL to <server node> for SM_MON of <client node>
kernel:lockd: cannot monitor <client node>
```

Solution

SELinux permissions must be granted to `rpcstatd` as under:

```
chcon -t bin_t /usr/sbin/rpc.statd
systemctl restart nfs
```

A node reboot might be required, if this problem persists after NFS service restart.

NFS error scenarios

This topic provides information on how to verify and resolve NFS errors.

NFS client cannot access exported NFS data

Problem

The NFS client cannot access the exported data even though the export is mounted. This often results in errors to occur while writing data, creating files, or traversing the directory hierarchy (permission denied).

Determination

The error itself occurs on the NFS client side. Additionally, and based on the nature of the problem, the server-side NFS logs can provide more details about the origin of the error.

Solution

There are multiple reasons for this problem:

The ACL definition in the file system does not allow the requested operation

The export and/or client definition of the export do not allow that operation (such as a "read only" definition)

1. Verify the ACL definition of the export path in the file system. To check ACL definitions, run:
   ```
   mmgetacl Path
   ```
2. Verify the definition of the export and the client (especially the access type). To check the NFS export details, run:
   ```
   mmnfs export list -n Path
   ```
3. Unmount and remount the file system on the NFS client:
   ```
   umount <Path>
   mount <mount_options> CES_IP_address:<export_path> <mount_point>
   ```

NFS client I/O temporarily stalled

Problem

The NFS client temporarily encounters stalled I/O or access requests to the export. The problem goes away after a short time (about 1 minute.)

Determination

The error itself occurs on the NFS client side, but due to an action on the NFS server side. The server-side NFS logs can provide more details about the origin of the error (such as a restart of the NFS server) along with the CES logs (such as manual move of a CES IP or a failover condition).

Origin

A restart of the NFS server might temporarily suspend further access to the export from the NFS client (depending on the type of request). The suspension occurs because a restart of the NFS server causes the grace period to start. During the grace period, certain NFS operations are not allowed:

1. An explicit restart triggered manually through the CLI by running: **mmces service stop / start ...**
2. An explicit move of CES IPs manually through the CLI by running: **mmces address move ...**
3. **A change in the definition of an existing export.**

   Note: Adding or removing NFS exports does not initiate a restart.
4. The creation of the first export.
5. A critical error condition that triggers CES failover, which in turn causes IP addresses to move.
6. A failback of CES IPs (depending on the setting of the address distribution policy).

Collecting diagnostic data for NFS

This topic describes the procedure for collecting diagnostic data for NFS services.

Diagnostic data can be generated by increasing the logging of the NFS server.

To change the logging temporarily on a single CES node without restarting the server, run the following the command on the CES node where you want to enable the tracing:

```
ganesha_mgr set_log COMPONENT_ALL FULL_DEBUG
```

CES NFS log levels can be user adjusted to select the amount of logging by the server. Every increase in log setting will add additional messages. So the default of "Event" will include messages tagged as EVENT, WARN, CRIT, MAJ, FATAL, but will not show INFO, DEBUG, MID_DEBUG, FULL_DEBUG:

Table 56. CES NFS log levels

Log name	Description
NULL	No logging
FATAL	Only asserts are logged
MAJ	Only major events are logged
CRIT	Only critical events are logged where there is malfunction, that is, for a single request
WARN	Events are logged that may be intended but may otherwise be harmful
EVENT	Default. level, which includes some events that are expected during normal operation (that is, start, grace period),
INFO	Enhanced level
DEBUG	Further enhanced, which includes events-relevant problem determination
MID_DEBUG	Further enhanced, which includes some events for developers
FULL_DEBUG	Maximal logging, which is mainly used for development purposes

These levels can be applied to a single component or to all components.

Note: The ganesha_mgr command requires that the CES NFS server be active on the node where the command is executed.

The CES NFS log file (default is `/var/log/ganesha.log`) will see a lot more updates eventually generating very large files or even filling up all of your disk.

To avoid issues with space usage, revert to the default logging by using the **`ganesha_mgr set_log COMPONENT_ALL EVENT`** command or reduce the set of components by using "FULL_DEBUG" to a reasonable subset of server components. For example, by replacing "COMPONENT_ALL" with "COMPONENT_DISPATCH".

Other possible components can be listed by using **`ganesha_mgr getall_logs`**. The ganesha_mgr changes are not persistent. A server restart will reset these settings to the settings in the cluster configuration as described in the **`mmnfs config list`** command.

Note: The **`mmnfs config list`** will show the persisted log level for all CES nodes (default: EVENT). Any log setting changes by using **`mmnfs config change LOG_LEVEL`** command will automatically do a server restart, possibly preventing to find a cause for the current issue. See the *mmnfs* topic in *IBM Spectrum Scale: Command and Programming Reference* for more information.

SMB issues

This topic describes SMB-related issues that you might come across while using the IBM Spectrum Scale system.

Determining the health of integrated SMB server

There are some IBM Spectrum Scale commands to determine the health of the SMB server.

The following commands can be used to determine the health of SMB services:

- To check the overall CES cluster state, issue the following command:

```
mmlscluster --ces
```

The system displays output similar to this:

```
GPFS cluster information
========================
  GPFS cluster name:         boris.nsd001st001
  GPFS cluster id:           3992680047366063927

Cluster Export Services global parameters
-----------------------------------------
  Shared root directory:                /gpfs/fs0
  Enabled Services:                     NFS SMB
  Log level:                            2
  Address distribution policy:          even-coverage

 Node  Daemon node name            IP address       CES IP address list
-----------------------------------------------------------------------
   4   prt001st001                 172.31.132.1     10.18.24.25 10.18.24.32 10.18.24.34 10.18.24.36 9.11.102.89
   5   prt002st001                 172.31.132.2     9.11.102.90 10.18.24.19 10.18.24.21 10.18.24.23 10.18.24.30
   6   prt003st001                 172.31.132.3     10.18.24.38 10.18.24.39 10.18.24.41 10.18.24.42 9.11.102.43
   7   prt004st001                 172.31.132.4     9.11.102.37 10.18.24.26 10.18.24.28 10.18.24.18 10.18.24.44
   8   prt005st001                 172.31.132.5     9.11.102.36 10.18.24.17 10.18.24.33 10.18.24.35 10.18.24.37
   9   prt006st001                 172.31.132.6     9.11.102.41 10.18.24.24 10.18.24.20 10.18.24.22 10.18.24.40
  10   prt007st001                 172.31.132.7     9.11.102.42 10.18.24.31 10.18.24.27 10.18.24.29 10.18.24.43
```

This shows at a glance whether nodes are failed or whether they host public IP addresses. For successful SMB operation at least one CES node must be HEALTHY and hosting at least one IP address.

- To show which services are enabled, issue the following command:

  ```
  mmces service list
  ```

  The system displays output similar to this:

  ```
  Enabled services: NFS SMB
  NFS is running, SMB is running
  ```

  For successful SMB operation, SMB needs to be enabled and running.

- To determine the overall health state of SMB on all CES nodes, issue the following command:

  ```
  mmces state show SMB -a
  ```

  The system displays output similar to this:

  ```
  NODE        SMB
  prt001st001 HEALTHY
  prt002st001 HEALTHY
  prt003st001 HEALTHY
  prt004st001 HEALTHY
  prt005st001 HEALTHY
  prt006st001 HEALTHY
  prt007st001 HEALTHY
  ```

- To show the reason for a currently active (failed) state on all nodes, issue the following command:

  ```
  mmces events active SMB -a
  ```

  The system displays output similar to this:

  ```
  NODE COMPONENT  EVENT NAME SEVERITY   DETAILS
  ```

  In this case nothing is listed because all nodes are healthy and so there are no active events. If a node was unhealthy it would look similar to this:

  ```
  NODE        COMPONENT EVENT NAME SEVERITY   DETAILS
  prt001st001 SMB       ctdb_down  ERROR      CTDB process not running
  prt001st001 SMB       smbd_down  ERROR      SMBD process not running
  ```

- To show the history of events generated by the monitoring framework, issue the following command

  ```
  mmces events list SMB
  ```

The system displays output similar to this:

```
NODE        TIMESTAMP                              EVENT NAME     SEVERITY  DETAILS
prt001st001 2015-05-27 14:15:48.540577+07:07MST smbd_up        INFO      SMBD process now running
prt001st001 2015-05-27 14:16:03.572012+07:07MST smbport_up     INFO      SMB port 445 is now active
prt001st001 2015-05-27 14:28:19.306654+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 14:28:34.329090+07:07MST ctdb_recovered INFO      CTDB Recovery finished
prt001st001 2015-05-27 14:33:06.002599+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 14:33:19.619583+07:07MST ctdb_recovered INFO      CTDB Recovery finished
prt001st001 2015-05-27 14:43:50.331985+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 14:44:20.285768+07:07MST ctdb_recovered INFO      CTDB Recovery finished
prt001st001 2015-05-27 15:06:07.302641+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 15:06:21.609064+07:07MST ctdb_recovered INFO      CTDB Recovery finished
prt001st001 2015-05-27 22:19:31.773404+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 22:19:46.839876+07:07MST ctdb_recovered INFO      CTDB Recovery finished
prt001st001 2015-05-27 22:22:47.346001+07:07MST ctdb_recovery  WARNING   CTDB Recovery detected
prt001st001 2015-05-27 22:23:02.050512+07:07MST ctdb_recovered INFO      CTDB Recovery finished
```

- To retrieve monitoring state from health monitoring component, issue the following command:

  ```
  mmces state show
  ```

  The system displays output similar to this:

  ```
  NODE        AUTH     NETWORK NFS     OBJECT   SMB      CES
  prt001st001 DISABLED HEALTHY HEALTHY DISABLED DISABLED HEALTHY
  ```

- To check the monitor log, issue the following command:

  ```
  grep smb /var/adm/ras/mmsysmonitor.log | head -n 10
  ```

  The system displays output similar to this:

  ```
  2016-04-27T03:37:12.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:37:27.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:37:42.3 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:37:57.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:38:12.4 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:38:27.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:38:42.5 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:38:57.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:39:12.2 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  2016-04-27T03:39:27.6 prt2st1 I Monitor smb service LocalState:HEALTHY Events:0 Entities:0 - Service.monitor:596
  ```

- The following logs can also be checked:

  ```
  /var/adm/ras/*
  /var/log/messages
  ```

File access failure from an SMB client with sharing conflict

If SMB clients fail to access files with file sharing conflict messages, and no such conflict exists, there can be a mismatch with file locking rules.

File systems that are exported with the CES SMB service, or a customized deployment version of Samba, require the **-D nfs4** flag on the **mmchfs** or **mmcrfs** command. This setting enables NFSv4 and SMB sharing rules.

SMB client on Linux fails with an “NT status logon failure”

This topic describes how to verify and resolve an “NT status logon failure” on the SMB client on Linux.

Description

The user is trying to log on to the SMB client using AD authentication on Linux and receives this message:

```
NT STATUS LOGON FAILURE
```

Following are the root causes of this error.

Description of Root cause #1

The user is trying to log on to the SMB client using AD authentication on Linux and receives this message:

```
Password Invalid
```

Cause

The system did not recognize the specified password.

Verification

Verify the password by running the following command on an IBM Spectrum Scale protocol node:

```
/usr/lpp/mmfs/bin wbinfo -a '<domain>\<user>'
```

The expected result is that the following messages display:

```
plaintext password authentication succeeded.
challenge/response password authentication succeeded.
```

If this message displays:

```
plaintext password authentication failed.
Could not authenticate user USER with plain text password
```

the domain for that user was not specified correctly.

Resolution

To resolve the error, enter the correct password.

If you do not know the correct password, follow your IT procedures to request a new password.

Description of root cause # 2

The user is trying to log on to the SMB client using AD authentication on Linux and receives this message:

```
The Userid is not recognized
```

Cause

The system did not recognize the specified password.

Verification

Verify the password by running the following command on an IBM Spectrum Scale protocol node:

```
/usr/lpp/mmfs/bin wbinfo -a '<domain>\<user>'
```

The expected result is that the following messages display:

```
plaintext password authentication succeeded.
challenge/response password authentication succeeded
```

If this message displays:

```
Could not authenticate user USER with challenge/response password
```

the specified user is not known by the system.

Resolution

To resolve the error, enter the correct userid.

If you think the correct user was specified, contact your IT System or AD Server administrator to get your userid verified.

SMB client on Linux fails with the `NT status password must change` error message

This topic describes how to verify and resolve an `NT status password must change` error on the SMB client on Linux.

Description

The user is trying to access the SMB client on Linux and receives this error message:

```
NT_STATUS_PASSWORD_MUST_CHANGE
```

Cause

The specified password expired.

Verification

Verify the password by running the following command on an IBM Spectrum Scale protocol node:

```
/usr/lpp/mmfs/bin wbinfo -a '<domain>\<user>'
```

The expected result is that the following messages display:

```
plaintext password authentication succeeded.
challenge/response password authentication succeeded.
```

If this message displays:

```
Could not authenticate user mzdom\aduser1 with challenge/response
```

the specified password probably expired.

Resolution

Log on to a Windows client, and when prompted, enter a new password. If the problem persists, ask the AD administrator to unlock the account.

SMB mount issues

This topic describes how to verify and resolve SMB mount errors.

Possible SMB mount error conditions include:

- Mount.CIFS on Linux fails with mount error (13) “Permission denied”
- Mount.CIFS on Linux fails with mount error (127) “Key expired”
- Mount on Mac fails with an authentication error.

If you receive any of these errors, verify your authentication settings. For more information, see “Protocol authentication issues” on page 368

Mount.Cifs on Linux fails with mount error (13) “Permission denied”

Description

The user is trying to mount CIFS on Linux and receives the following error message:

```
Permission Denied
```

The root causes for this error are the same as for “SMB client on Linux fails with an “NT status logon failure”” on page 385.

Mount.Cifs on Linux fails with mount error (127) “Key has expired”

Description

The user is trying to access a CIFS share and receives the following error message:

```
key has expired
```

The root causes for this error are the same as for “SMB client on Linux fails with an “NT status logon failure”” on page 385.

Mount on Mac fails with an authentication error

Description

The user is attempting a mount on a Mac and receives this error message:

```
mount_smbfs: server rejected the connection: Authentication error
```

The root causes for this error are the same as for “SMB client on Linux fails with an “NT status logon failure”” on page 385.

Net use on Windows fails with “System error 86”

This topic describes how to verify and solve a “System error 86” when the user is attempting to access net use on Windows.

Description

While accessing the network the following error message displays:

```
System error 86 has occurred.
The specified password is not correct.
```

Solution

The root causes for this error are the same as that for the failure of SMB client on Linux. For more information on the root cause, see “SMB client on Linux fails with an “NT status logon failure”” on page 385.

Net use on Windows fails with “System error 59” for some users

This topic describes how to resolve a “System error 59” when some users attempt to access net use on Windows.

Description:

Additional symptoms include

```
NT_STATUS_INVALID_PARAMETER
```

errors in the `log.smbd` file when net use command was invoked on the Windows client for the user with this problem.

Solution:

Invalid idmapping entries in gencache might be the cause. To resolve the error, delete these entries in gencache on all nodes. Run the following commands: **net cache del IDMAP/UID2SID/<UID>** and **net cache del IDMAP/SID2XID/<SID>**. You can run the **mmadquery** command to know the <*UID*> and the <*SID*>. Alternatively, you can find the <*SID*> from the `log.smbd` file. See the following message

```
Could not convert sid <SID>: NT_STATUS_INVALID_PARAMETER
```

in the `log.smbd` file. Here, <*SID*> is the SID of the user.

Winbindd causes high CPU utilization

This topic describes the issues that can happen due to the winbindd component.

Cause

One possible reason is that winbind is not able to find domain controllers for a given domain. NT_STATUS_NO_LOGON_SERVERS is seen in log file `log.winbindd-dc-connect` in that case. One possible issue here is that the DNS does not provide this information. Usually the local DCs have to be configured as DNS servers on the protocol nodes, as AD stores additional information for locating DCs in the DNS.

Solution

The problem is also known to go away after upgrading to IBM Spectrum Scale 4.2.2.

SMB error events

This topic describes how to verify and resolve SMB errors.

CTDB process is not running (ctdb_down)

Cause

CTDB process is not running.

Determination

Check `/var/log/messages` for CTDB error messages or crashes.

Solution

Fix any obvious issues and run this command:

```
mmces service stop SMB
mmces service start SMB
```

CTDB recovery detected (ctdb_recovery)

Cause

CTDB status is stuck in Recovery mode for an extended amount of time.

Determination

If the service status is Degraded for a while, there is an issue. The service status should be Transient. Check the logs for a possible issue.

Solution

Run:

```
mmces service stop SMB && mmces service start SMB
```

If still not fixed, run:

```
gpfs.snap
```

and contact IBM support.

CTDB state is not healthy (ctdb_state_down)

Determination

1. Check `/var/log/messages` for errors and correct any that you find.
2. Check CTDB status by running the **`ctdb status`** command.
3. Check the network connectivity.

Solution

After the error is resolved, the CTDB node should recover. If you have not resolved the error, restart SMB by running this command:

```
mmces service stop SMB && mmces service start SMB
```

SMDB process not running

Determination

1. Check `/var/log/messages` and `/var/adm/ras/log.smbd` for errors and correct if found.
2. Restart by running this command:

   ```
   mmces service stop SMB && mmces services start SMB
   ```

SMB port (?) is not active (smbport_down_)

Cause

The SMB port (?) is not listening for connections.

Determination

Check the network connectivity.

Solution

Restart by running:

```
mmces service stop SMB && mmces services start SMB
```

SMB access issues

This topic describes how to analyze and resolve SMB access issues.

The most common issue with ACLs is getting an unexpected Access denied message. Check the following:

1. Export ACLs: Use the MMC tool or **mmsmb exportacl** to see that the share allows access for the logged in user.
2. File system object ACLs: Use the Windows Explorer ACL dialog and/or **mmgetacl** to make sure the correct ACLs are in place on all components in the path.
3. Make sure that the READ_ATTR right is set on folders to be traversed.
4. Keep in mind that even if READ_NAMED and WRITE_NAMED are not enforced by the file system, the SMB server enforces them.
5. Export settings: Check the export settings by running **mmsmb export list --all** so that export options like read only = no or available = no do not restrict access.
6. Make sure your clients try to negotiate a supported protocol level.
7. For smbclient: make sure the option -m SMB2 is used and supported by your version of smbclient (smbclient -L localhost -U<user>%<password> -m SMB2)
8. Windows XP, Windows Server 2003 and older Windows versions are not supported, because they only support SMB1.
9. For the Linux kernel client, make sure you check the version option to use smb2.

Note: For known issues in the Linux kernel client, see the documentation for your Linux distribution.

If the root cause cannot be narrowed down, perform these steps the results of which will help make a more detailed analysis.

1. Provide exact information about what happened.
2. Provide screen captures of Windows ACL dialogs with the problem before and after the issue.
3. Provide the output of **mmgetacl** for all files and folders related to the ACL/permission problem before and after the problematic event.
4. Trace how the client has mounted the share.
5. You can force a re-connect by stopping the smbd process that serves that connection.
6. Describe how the user has mounted the export.
7. List all users and groups that are in the test along with their memberships.
8. Collect export information by running: **mmsmb export list --all**.
9. Provide the version of Windows used for each client.
10. Provide a Samba level 10 trace for the test by running the mmprotocoltrace tool.
11. Provide IBM Spectrum Scale traces for the test by running **mmtracectl --start and --stop**.
12. Collect the network trace of the re-create by running **mmprotocoltrace**.

Slow access to SMB caused by contended access to files or directories

This topic describes the reason behind the slow access to SMB server and the troubleshooting steps to handle it.

If the access through the SMB server is slower than expected, then there might be an issue with the highly contended access to the same file or directory through the SMB server. This happens because of the internal record keeping process of the SMB server. The internal record keeping process requires that the record for each open file or directory must be transferred to different protocol nodes for every open and close operation, which at times, overloads the SMB server. This delay in access is experienced in extreme cases, where many clients are opening and closing the same file or directory. However, note that concurrent access to the same file or directory is handled correctly in the SMB server and it usually causes no problems.

The following procedure can help tracking the files or directories of the contended records in the database statistics using CTDB track. When a "hot" record is detected, it is recorded in the database statistic and a message is printed to syslog.

When this message refers to the locking.tdb database, this can point to the problem of concurrent access to the same file or directory. The same reference might be seen in the ctdb dbstatistics for locking.tdb

```
# ctdb dbstatistics locking.tdb
  DB Statistics locking.tdb
  db_ro_delegations                  0
  db_ro_revokes                      0
  locks
       num_calls                     15
       num_current                   0
       num_pending                   0
       num_failed                    0
  db_ro_delegations                  0
  hop_count_buckets:                 139 40 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  lock_buckets:                      0 9 6 0 0 0 0 0 0 0 0 0 0 0 0 0
  locks_latency      MIN/AVG/MAX     0.002632/0.016132/0.061332 sec out of 15
  vacuum_latency     MIN/AVG/MAX     0.000408/0.003822/0.082142 sec out of 817
  Num Hot Keys:     10
       Count:1 Key:                  6a4128e3ced4681b017c060000000000000000000000000000
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
       Count:0 Key:
```

When ctdb points to a hot record in locking.tdb, then use the "net tdb locking" command to determine the file behind this record:

```
# /usr/lpp/mmfs/bin/net tdb locking 6a4128e3ced4681b017c060000000000000000000000000000
  Share path:            /ibm/fs1/smbexport
  Name:                  testfile
  Number of share modes: 2
```

If this happens on the root directory of an SMB export, then a workaround can be to exclude that from cross-node locking:

```
mmsmb export change smbexport --option fileid:algorithm=fsname_norootdir
```

If this happens on files, the recommendation would be to access that SMB export only through one CES IP address, so that the overhead of transferring the record between the nodes is avoided.

If the SMB export contains only sub directories with home directories where the sub directory names match the user name, the recommended configuration would be an SMB export uses the %U sub situation to automatically map the user with the corresponding home directory:

```
mmsmb export add smbexport /ibm/fs1/%U
```

Object issues

This topic describes some of the Object-related issues that you might come across while using IBM Spectrum Scale.

Getting started with troubleshooting object issues

Use the following checklists to troubleshoot object issues.

Checklist 1

This checklist must be referred to before using an object service.

1. Check the cluster state by running the **mmgetstate -a** command.

   The cluster state must be Active.
2. Check the status of the CES IP by running the **mmlscluster -ces** command.

   The system displays the all the CES nodes along with their assigned IP addresses.
3. Check the service states of the CES by running the **mmces state show -a** or **mmhealth node show ces -N cesnodes** command.

   The overall CES state and object service states must be Healthy.
4. Check the service listing of all the service states by running the **mmces service list -verbose** command.
5. Check the authentication status by running the **mmuserauth service check** command.
6. Check the object auth listing by running the **source $HOME/openrc ; openstack user list**command.

   The system lists all the users IDs.

Checklist 2

This checklist must be referred to before using the keystone service.

1. Check if object authentication has been configured by running the **mmuserauth service list --data-access-method object** command.
2. Check the state of object authentication by running the **mmces state show AUTH_OBJ -a** command.
3. Check if the protocol node is serving the CES IP by running the **mmlscluster --ces** command.
4. Check if the object_database_node tag is present in one of the CES IP by running the **mmces address list** command.
5. Check if httpd is running on all the CES nodes and postgres is running on the node that has CES IP with the object_database_node tag by running the **mmces service list -v -a** command.
6. Check if authentication configuration is correct on all nodes by running the **mmuserauth service check --data-access-method object -N cesNodes** command.
7. If the mmuserauth service check reports an error, run the **mmuserauth service check --data-access-method object --rectify -N <node>** command where *node* is the number of the node on which the error is reported.

Authenticating the object service

This topic provides troubleshooting references and steps for resolving system errors when you are authenticating the object service.

Description

When the user authenticates or runs any create, update, or delete operation, the system displays one of the following errors:

```
{"error": {"message": "An unexpected error prevented the server from fulfilling your request.",
 "code": 500, "title": "Internal Server Error"}}

ERROR: openstack An unexpected error prevented the server from fulfilling your request.
(HTTP 500)(Request-ID: req-11399fd1-a601-4615-8f70-6ba275ec3cd6)
```

Cause

The system displays this error under one or all of the following conditions:

- The authentication service is not running.
- The system is unable to reach the authentication server.
- The user credentials for keystone have been changed or have expired.

Proposed workaround

- Perform all the steps in Checklist 1.
- Check if the IP addresses of the keystone endpoints are correct and reachable. If you are using a local keystone, check if the postgresql-obj service is running.

Authenticating or using the object service

This topic provides troubleshooting references and steps for resolving system errors when you are authenticating or using the object service.

Description

When the user is authenticating the object service or running the create, update, retrieve, and delete operations, the system displays the following error:

```
Error: {"error": {"message": "The request you have made requires authentication.",
 "code": 401, "title": "Unauthorized"}}
```

Cause

The system displays this error under one or all of the following conditions:

- The password, user ID, or service ID that you have entered is incorrect.
- The token that you are using has expired.

Proposed workaround

- Check your user ID and password. All user IDs in the system can be viewed in the OpenStack user list.
- Check if a valid service ID is provided in the `/etc/swift/proxy-server.conf` file, in the filter:authtoken section. Also, check if the password for the service ID is still valid. The service ID can be viewed in the OpenStack service, project, and endpoint lists.

Accessing resources

This topic provides troubleshooting references and steps for resolving system errors when you are accessing resources.

Description

When an unauthorized user is accessing an object resource, the system displays the following error:

```
| Error: Error: HTTP/1.1 403 Forbidden
| Content-Length: 73 Content-Type: text/html; charset=UTF-8 X-Trans-Id: tx90ad4ac8da9242068d111-
| 0056a88ff0 Date: Wed, 27 Jan 09:37:52 GMT <html><h1>Forbidden</h1><p>Access was denied to this
| resource.</p>
```

Cause

The system displays this error under one or all of the following conditions:

- The user is not authorized by the system to access the resources for a certain operation.
- The endpoint, authentication URL, service ID, keystone version, or API version is incorrect.

Proposed workaround

- To gain authorization and access the resources, contact your system administrator.
- Check your service ID. The service ID can be viewed in the OpenStack service, project, and endpoint lists.

Connecting to the object services

This topic provides troubleshooting references and steps for resolving system errors when you are connecting to the object services.

Description

When the user is unable to connect to the object services, the system displays the following error:

```
curl: (7) Failed connect
to spctscl2.in.ibm.com:8080; No route to host
```

Cause

The system displays this error under one or all of the following conditions:

- The firewall is running.
- The firewall rules have been configured incorrectly.

Proposed workaround

Set up the firewall rules correctly in your system.

For more information about the firewall rules, see *Installation prerequisites* in *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

Creating a path

This topic provides troubleshooting references and steps for resolving system errors when you are creating a path.

Description

While you perform a create, update, retrieve, or delete task, if you attempt to create a non-existent path the system displays the following error:

```
| Error: HTTP/1.1 404 Not Found
| Content-Length: 70 Content-Type: text/html; charset=UTF-8 X-Trans-Id:
| tx88ec3b783bc04b78b5608-0056a89b52 Date: Wed, 27 Jan 10:26:26
| GMT <html><h1>Not Found</h1><p>The resource could not be found.</p></html>
```

Cause

The system displays this error because the path you are creating does not exist.

Proposed workaround

Recreate the object or the container before you perform the GET operation.

Constraints for creating objects and containers

This topic provides the constraints that must be kept in mind while creating objects and containers.

Description

When the user is trying to create objects and containers for unified file and object access, the system displays the 400 Bad request error.

Cause

The system displays this error under one or all of the following conditions:

- The name of the container is longer than 255 characters.
- The name of the object is longer than 214 characters.
- The name of any container in the object hierarchy is longer than 214 characters.
- The path name of the object includes successive forward slashes.
- The name of the container and the object is a single period (.) or a double period (..).

Proposed workaround

Keep in mind the following constraints while creating objects and containers for unified file and object access:

- The name of the container must not exceed 255 characters.
- The name of the object must not exceed 214 characters.
- The name of any container in the object hierarchy must not exceed 214 characters.
- The path name of the object must not include successive forward slashes.
- The name of the container and the object must not be a single period (.) or a double period (..). However, a single period or a double period can be part of the name of the container and the object.

The Bind password is used when the object authentication configuration has expired

This topic provides troubleshooting references and steps for resolving system errors when you are using the Bind password and the object authentication configuration has expired.

Description

When object is configured with the AD/LDAP authentication and the bind password is being used for LDAP communication, the system displays the following error:

```
[root@SSClusterNode3 ~]# openstack user list

ERROR: openstack An unexpected error prevented the server from fulfilling your request. (HTTP
500) (Request-ID: req-d2ca694a-31e3-46cc-98b2-93556571aa7d) Authorization Failure. Authorization
failed: An unexpected error prevented the server from fulfilling your request. (HTTP 500)
(Request-ID: req-d6ccba54-baea-4a42-930e-e9576466de3c)
```

Cause

The system displays this error when the Bind password has been changed on the AD/LDAP server.

Proposed workaround

1. Obtain the new password from the AD/LDAP server.
2. Run the following command to restart keystone on all protocol nodes: `mmobj config change --ccrfile keystone.conf --section ldap --property password --value <password>` where password is the new password obtained in Step 1.

Note: This command restarts Keystone on all protocol nodes.

The password used for running the keystone command has expired or is incorrect

This topic provides troubleshooting references and steps for resolving system errors when you are using an expired or incorrect password for running the keystone command.

Description

When the user is trying to run the keystone command by using a password that has expired or is incorrect, the system displays the following error:`[root@SSClusterNode5 ~]# openstack user list`

```
ERROR: openstack The request you have made requires authentication. (HTTP 401) (Request-ID:
req-9e8d91b6-0ad4-42a8-b0d4-797a08150cea)
```

Cause

The system displays this error when the user has changed the password but is still using the expired password to access keystone.

Proposed workaround

Use the correct password to access keystone.

The LDAP server is not reachable

This topic provides troubleshooting references and steps for resolving system errors when you are trying to reach an LDAP server.

Description

When object authentication is configured with AD/LDAP and the user is trying to run the keystone commands, the system displays the following error:`[root@SSClusterNode3 ~]# openstack user list`

```
ERROR: openstack An unexpected error prevented the server from fulfilling your request. (HTTP
500) (Request-ID: req-d3fe863e-da1f-4792-86cf-bd2f4b526023)
```

Cause

The system displays this error under one or all of the following conditions:

- The LDAP server is not reachable due to network issues.
- The LDAP server is not reachable because the system firewall is running.
- The LDAP server has been shut down.

Note:

When the LDAP server is not reachable, the keystone logs can be viewed in the `/var/log/keystone` directory.

The following example is an LDAP error found in `/var/log/keystone/keystone.log`:

```
/var/log/keystone/keystone.log:2016-01-28 14:21:00.663 25720 TRACE keystone.common.wsgi result =
func(*args,**kwargs)2016-01-28 14:21:00.663 25720 TRACE keystone.common.wsgi SERVER_DOWN:
{'desc': "Can't contact LDAP server"}.
```

Proposed workaround

- Check your network settings.
- Configure your firewall correctly.
- Repair the LDAP server.

The TLS certificate has expired

This topic provides troubleshooting references and steps for resolving system errors when the TLS certificate has expired.

Description

When the user is trying to configure object authentication with AD/LDAP by using the TLS certificate for configuration, the system displays the following error:

```
[E] Failed to execute command
ldapsearchldap_start_tls: Connect error (-11)additional info: TLS error -8174:security library
: bad database.mmuserauth service create: Command failed.
Examine previous error messages to determine cause.
```

Cause

The system displays this error because the TLS certificate has expired.

Proposed workaround

1. Update the TLS certificate on the AD/LDAP server.
2. Rerun the command.

The TLS CACERT certificate has expired

This topic provides troubleshooting references and steps for resolving system errors when the TLS CACERT certificate has expired.

Description

When the system is configured with AD/LDAP and TLS, the TLS CACERT has expired after configuration, and the user is trying to run the keystone command, the system displays the following error:

```
[root@SSClusterNode3 ~]# openstack user list
ERROR: openstack An unexpected error prevented the server from fulfilling your request.
(HTTP 500) (Request-ID: req-dfd63d79-39e5-4c4a-951d-44b72e8fd9ef)
Logfile /var/log/keystone/keystone.log2045-01-14 10:50:40.809 30518
TRACE keystone.common.wsgi CONNECT_ERROR:
{'info': "TLS error -8162:The certificate issuer's certificate has expired.
Check your system date and time.", 'desc': 'Connect error'}
```

Note:

The log files for this error can be viewed in /var/log/keystone/keystone.log.

Cause

The system displays this error because the TLS CACERT certificate has expired.

Proposed workaround

1. Obtain the updated TLS CACERT certificate on the system.
2. Rerun the object authentication command.

Note:

If you run the **–idmapdelete** command while performing the workaround steps you might lose existing data.

The TLS certificate on the LDAP server has expired

This topic provides troubleshooting references and steps for resolving system errors when the TLS certificate on the LDAP server has expired.

Description

When the system is configured with AD/LDAP using TLS, and the certificate on AD/LDAP has expired, the system displays the following error when the user is trying to run the keystone commands:

```
[root@SSClusterNode3 ~]# openstack user list
ERROR: openstack An unexpected error prevented the server from fulfilling your request.
(HTTP 500) (Request-ID: req-5b3422a1-fc43-4210-b092-1201e38b8cd5)2017-05-08 22:08:35.443 30518
TRACE keystone.common.wsgi CONNECT_ERROR: {'info': 'TLS error -8157:Certificate extension not found.',
'desc': 'Connect error'}
2017-05-08 22:08:35.443 30518 TRACE keystone.common.wsgi
```

Cause

The system displays this error because the TLS certificate on the LDAP server has expired.

Proposed workaround

Update the TLS certificate on the LDAP server.

The SSL certificate has expired

This topic provides troubleshooting references and steps for resolving system errors when the SSL certificate has expired.

Description

When object authentication is configured with SSL and the user is trying to run the authentication commands, the system displays the following error:

```
[root@SSClusterNode3 ~]# openstack user list
ERROR: openstack SSL exception connecting to https://SSCluster:35357/v3/auth/tokens:
[Errno 1] _ssl.c:504: error:14090086:SSL routines:SSL3_GET_SERVER_CERTIFICATE:certificate verify failed
```

Cause

The system displays this error because the SSL certificate has expired. The user may have used the same certificate earlier for keystone configuration, but now the certificate has expired.

Proposed workaround

1. Remove the object authentication.
2. Reconfigure the authentication with the new SSL certificate.

Note:

Do not run the **mmuserauth service remove --data-access-method object --idmapdelete** command during removing and reconfiguring the authentication.

Users are not listed in the OpenStack user list

This topic provides troubleshooting references and steps for resolving system errors when the user is not listed in the OpenStack user list.

Description

When the authentication type is AD/LDAP, the users are not listed in the OpenStack user list.

Cause

The system displays this error under one or all of the following conditions:

- Only the users under the specified user DN are visible to keystone.
- The users do not have the specified object class.

Proposed workaround

Change the object authentication or modify the AD/LDAP for the users who do not have the specified object class.

The error code signature does not match

This topic provides troubleshooting references and steps for resolving system errors when the error code signature does not match.

Description

When there is an error code signature mismatch, the system displays the following error:

```
<?xml version="1.0" encoding="UTF-8"?><Error> <Code>SignatureDoesNotMatch</Code> <Message>The
request signature we calculated does not match the signature you provided. Check your key and
signing method.</Message> <RequestId>tx48ae6acd398044b5b1ebd-005637c767</RequestId></Error>
```

Cause

The system displays this error when the specified user ID does not exist and the user ID does not have the defined credentials or has not assigned a role to the account.

Proposed workaround

- For role assignments, review the output of these commands to identify the role assignment for the affected user:
  - openstack user list
  - openstack role assignment list
  - openstack role list
  - openstack project list
- For credential issues, review the credentials assigned to that user id:
  - openstack credential list
  - openstack credential show <ID>

The swift-object-info output does not display

This topic provides troubleshooting references and steps for resolving errors when the command **swift-object-info** does not display output.

The swift-object-info does not display output

On IBM Spectrum Scale IBM Spectrum Scale clusters with SELinux enabled and enforced, the system does not display output for the **swift-object-info** command.

Cause

The boolean daemons_use_tty default setting is preventing the **swift-object-info** output to display.

Proposed workaround

Allow daemons to use tty by issuing the **setsebool -P daemons_use_tty 1** command.

Output for subsequent entries of **swift-object-info** now displays correctly.

Swift PUT returns the 202 error and S3 PUT returns the 500 error due to the missing time synchronization

This topic provides troubleshooting references and steps for resolving system errors when Swift PUT returns the 202 error and S3 PUT returns the 500 error due to the missing time synchronization.

Description

The swift object servers require monotonically-increasing timestamps on the PUT requests. If the time between all the nodes is not synchronized, the PUT request can be rejected, resulting in the object server returning a 409 status code that is turned into 202 in the proxy-server. When the swift3 middleware receives the 202 code, it returns a 500 to the client. When enabling DEBUG logging, the system displays the following message:

From the object server:

```
Feb 9 14:41:09 prt001st001 object-server: 10.0.5.6 - - [09/Feb/2016:21:41:09 +0000] "PUT
/z1device119/14886/AUTH_bfd953e691c4481d8fa0249173870a56/mycontainers12/myobjects407"
```

From the proxy server:

```
Feb 9 14:14:10 prt003st001 proxy-server: Object PUT returning 202 for 409: 1455052450.83619 <=
'409 (1455052458.12105)' (txn: txf7611c330872416aabcc1-0056ba56a2) (client_ip:
```

If S3 is used, the following error is displayed from Swift3:

```
Feb 9 14:25:52 prt005st001 proxy-server: 500 Internal Server Error: #012Traceback (most recent
call last):#012 File "/usr/lib/python2.7/site-packages/swift3/middleware.py", line 81, in
__call__#012 resp = self.handle_request(req)#012 File "/usr/lib/python2.7/site-packages/swift3/
middleware.py", line 104, in handle_request#012 res = getattr(controller, req.method)(req)#012
File "/usr/lib/python2.7/site-packages/swift3/controllers/obj.py", line 97, in PUT#012 resp =
req.get_response(self.app)#012 File "/usr/lib/python2.7/site-packages/swift3/request.py", line
825, in get_response#012 headers, body, query)#012 File "/usr/lib/python2.7/site-packages/
swift3/request.py", line 805, in get_acl_response#012 app, method, container, obj, headers,
body, query)#012 File "/usr/lib/python2.7/site-packages/swift3/request.py", line 669, in
_get_response#012 raise InternalError('unexpected status code %d' % status)#012InternalError:
500 Internal Server Error (txn: tx40d4ff7ca5b94b1bb6881-0056ba5960) (client_ip: 10.0.5.1) Feb 9
14:25:52 prt005st001 proxy-server: 500 Internal Server Error: #012Traceback (most recent call
last):#012 File "/usr/lib/python2.7/site-packages/swift3/middleware.py", line 81, in
__call__#012 resp = self.handle_request(req)#012 File "/usr/lib/python2.7/site-packages/swift3/
middleware.py", line 104, in handle_request#012 res = getattr(controller, req.method)(req)#012
```

```
File "/usr/lib/python2.7/site-packages/swift3/controllers/obj.py", line 97, in PUT#012 resp =
req.get_response(self.app)#012 File "/usr/lib/python2.7/site-packages/swift3/request.py", line
825, in get_response#012 headers, body, query)#012 File "/usr/lib/python2.7/site-packages/
swift3/request.py", line 805, in get_acl_response#012 app, method, container, obj, headers,
body, query)#012 File "/usr/lib/python2.7/site-packages/swift3/request.py", line 669, in
_get_response#012 raise InternalError('unexpected status code %d' % status)#012InternalError:
500 Internal Server Error (txn: tx40d4ff7ca5b94b1bb6881-0056ba5960) (client_ip: 10.0.5.1)
```

Cause

The system displays these errors when the time is not in sync.

Proposed workaround

- To check if this problem is occurring, run the **mmdsh date** command.
- Enable the NTPD service on all protocol nodes and have the time synchronized from an NTP server.

Unable to generate the accurate container listing by performing the GET operation for unified file and object access container

This topic provides troubleshooting references and steps for resolving system errors when the system is unable to generate the accurate container listing by performing the GET operation for unified file and object access container.

Description

The system does not display the accurate container listing for a unified file and object access enabled container.

Cause

This error occurs under one or all of the following conditions:

- The ibmobjectizer interval is too long. Therefore, a longer time is taken to update and display the listing.
- The files created on the file system are not supported for objectization.

Proposed workaround

Tune the ibmobjectizer interval configuration by running the **mmobj config change** command.

The following is an example of setting up the objectization interval by using the **mmobj config change**:

```
mmobj config change --ccrfile spectrum-scale-objectizer.conf \
--section DEFAULT --property objectization_interval --value 2400
```

This command sets an interval of 40 minutes between the completion of an objectization cycle and the start of the next cycle.

Fatal error of object configuration during deployment

This topic provides troubleshooting references and steps for resolving fatal system errors in object configuration during deployment.

Description

When the user enables object by using installation toolkit, the system displays the following error:

```
[ FATAL ] Required option 'endpoint_hostname' missing in section:
'object'. To set this, use: ./spectrumscale config object –endpoint

[ FATAL ] Invalid configuration for setting up Object Store.
```

Cause

The system displays this error when the object authentication not completed with the required parameters.

Proposed workaround

Run the **spectrumscale config obj** command with the mandatory arguments.

Object authentication configuration fatal error during deployment

This topic provides troubleshooting references and steps for resolving fatal system errors in object authentication configuration during deployment.

Description

When the user configures the authentication by using the installation toolkit, the system displays the following error:

```
2016-02-16 13:48:07,799 [ FATAL ] <nodename> failure whilst: Configuring object authentication
(SS98)
```

Cause

The system displays this error under one or all of the following conditions:

- Only the users under the specified user DN are visible to Keystone.
- The users do not have the specified object class.

Proposed workaround

Change the object authentication or modify the AD/LDAP for the users who do not have the specified object class.

Fatal error of object authentication during deployment

This topic provides troubleshooting references and steps for resolving fatal errors in object authentication during deployment.

Description

When the user configures authentication by using installation toolkit, the system displays the following error:

```
02-16 13:48:07,799 [ FATAL ] <nodename> failure whilst: Configuring object authentication (SS98)
```

Cause

The system displays this error under one or all of the following conditions:

- IBM Spectrum Scale for the object storage program is currently running.
- Parameters provided in the configuration.txt and authconfig.txt files are incorrect.

- The system is unable to connect to the authentication server with the given credentials or network issues.

Proposed workaround

- Shut down IBM Spectrum Scale for the object storage program before continuing.
- Check the connectivity of protocol nodes with the authentication server with valid credentials.
- Stop the service manually with the **mmces service stop obj -a** command. Manually run the **mmuserauth service create** command to complete the authentication configuration for object.
- Fix the configuration.txt and authconfig.txt files and rerun the IBM Spectrum Scale deployment with the **spectrumscale deploy** command.

Chapter 22. Disaster recovery issues

As with any type of problem or failure, obtain the GPFS log files (**mmfs.log.***) from all nodes in the cluster and, if available, the content of the internal dumps.

For more information, see:

- The *Data mirroring and replication* topic in the *IBM Spectrum Scale: Administration Guide* for detailed information about GPFS disaster recovery
- "Creating a master GPFS log file" on page 198
- "Information to be collected before contacting the IBM Support Center" on page 469

The following two messages might appear in the GPFS log for active/active disaster recovery scenarios with GPFS replication. The purpose of these messages is to record quorum override decisions that are made after the loss of most of the disks:

6027-435 [N]
The file system descriptor quorum has been overridden.

6027-490 [N]
The descriptor replica on disk *diskName* has been excluded.

A message similar to these appear in the log on the file system manager, node every time it reads the file system descriptor with an overridden quorum:

```
...
6027-435 [N] The file system descriptor quorum has been overridden.
6027-490 [N] The descriptor replica on disk gpfs23nsd has been excluded.
6027-490 [N] The descriptor replica on disk gpfs24nsd has been excluded.
...
```

For more information on node override, see the section on *Quorum*, in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*

For PPRC and FlashCopy®-based configurations, more problem determination information can be collected from the ESS log file. This information and the appropriate ESS documentation must be referred while working with various types disk subsystem-related failures. For instance, if users are unable to perform a PPRC failover (or failback) task successfully or unable to generate a FlashCopy of a disk volume, they should consult the subsystem log and the appropriate ESS documentation. For more information, see the following topics:

- *IBM TotalStorage Enterprise Storage Server® Web Interface User's Guide*(publibfp.boulder.ibm.com/epubs/pdf/f2bui05.pdf).

Disaster recovery setup problems

The following setup problems might impact disaster recovery implementation:

1. Considerations of data integrity require proper setup of PPRC consistency groups in PPRC environments. Additionally, when using the FlashCopy facility, make sure to suspend all I/O activity before generating the FlashCopy image. See "Data integrity" on page 346.
2. In certain cases, it might not be possible to restore access to the file system even after relaxing the node and disk quorums. For example, in a three failure group configuration, GPFS tolerates and recovers from a complete loss of a single failure group (and the tiebreaker with a quorum override). However, all disks in the remaining failure group must remain active and usable in order for the file system to continue its operation. A subsequent loss of at least one of the disks in the remaining failure group would render the file system unusable and trigger a forced unmount. In such situations, users

might still be able to perform a restricted mount and attempt to recover parts of their data from the damaged file system. For more information on restricted mounts, see “Restricted mode mount” on page 255.

3. When you issue **mmfsctl syncFSconfig**, you might get an error similar to the following:

   ```
   mmfsctl: None of the nodes in the peer cluster can be reached
   ```

   In such scenarios, check the network connectivity between the peer GPFS clusters and verify their remote shell setup. This command requires full TCP/IP connectivity between the two sites, and all nodes must be able to communicate by using ssh or rsh without the use of a password.

Protocols cluster disaster recovery issues

Sometimes issuing an **mmcesdr** command can cause problems with protocols disaster recovery in IBM Spectrum Scale.

Whenever such an error or problem is encountered, view the Protocols DR log file for more information on the issue. This log file is at `/var/adm/ras/mmcesdr.log` on the node where the command was run.

Other problems with disaster recovery

You might encounter the following issues that are related to disaster recovery in IBM Spectrum Scale:

1. Currently, users are advised to always specify the **all** option when you issue the **mmfsctl syncFSconfig** command, rather than the device name of one specific file system. Issuing this command enables GPFS to detect and correctly resolve the configuration discrepancies that might occur as a result of the manual administrative action in the target GPFS cluster to which the configuration is imported.
2. The optional **SpecFile** parameter to the **mmfsctl syncFSconfig**that is specified with the **-S** flag must be a fully qualified path name that defines the location of the spec data file on nodes in the target cluster. It is not the local path name to the file on the node from which the **mmfsctl** command is issued. A copy of this file must be available at the provided path name on all peer contact nodes that are defined in the **RemoteNodesFile**.

Chapter 23. Performance issues

The performance issues might occur because of the system components or configuration or maintenance issues.

Issues caused by the low-level system components

This section discusses the issues caused by the low-level system components used in the IBM Spectrum Scale cluster.

Suboptimal performance due to high utilization of the system level components

In some cases, the CPU or memory utilization on an IBM Spectrum Scale node is higher than 90%. Such heavy utilization can adversely impact the system performance as it affects the cycles allocated to the IBM Spectrum Scale daemon service.

Problem identification

On the node, issue an Operating System command such as **top** or **dstat** to verify whether the system level resource utilization is higher than 90%. The following example shows the sample output for the **dstat** command:

```
# dstat 1 10
----total-cpu-usage---- -dsk/total- -net/total- ---paging-- ---system--
usr sys idl wai hiq siq| read  writ| recv  send|  in   out | int   csw
  0   0 100   0   0   0|7308k 9236k|   0     0 |   0     0 | 812  3691
  0   0 100   0   0   0|   0     0 |3977B 1038B|   0     0 | 183   317
  1   2  98   0   0   0|   0     0 |2541B  446B|   0     0 | 809   586
  0   1  99   0   0   0|   0     0 |4252B  346B|   0     0 | 427   405
  0   0 100   0   0   0|   0     0 |3880B  346B|   0     0 | 196   349
  0   0 100   0   0   0|   0     0 |3594B  446B|   0     0 | 173   320
  1   1  98   0   0   0|   0     0 |3969B  446B|   0     0 | 692   662
  0   0 100   0   0   0|   0   116k|3120B  346B|   0     0 | 189   312
  0   0 100   0   0   0|   0     0 |3050B  346B|   0     0 | 209   342
  0   0 100   0   0   0|   0  4096B|4555B  346B|   0     0 | 256   376
  0   0 100   0   0   0|   0     0 |3232B  346B|   0     0 | 187   340
```

Problem resolution and verification

If the system level resource utilization is high, determine the process or application that contributes to the performance issue and take appropriate action to minimize the utilization to an acceptable level.

Suboptimal performance due to long IBM Spectrum Scale waiters

Low-level system issues, like slow disks, or slow network, might cause long GPFS waiters. These long waiters cause performance degradation. You can use the **mmdiag --waiters** command to display the `mmfsd` threads waiting for events. This information can help resolve deadlocks and improve the system performance.

Problem identification

On the node, issue the **mmdiag --waiters** command to check whether any long waiters are present. The following example shows long waiters that are contributed by the slow disk, **dm-14**:

```
#mmdiag --waiters
0x7FF074003530 waiting 25.103752000 seconds, WritebehindWorkerThread: for I/O completion on disk dm-14
0x7FF088002580 waiting 30.025134000 seconds, WritebehindWorkerThread: for I/O completion on disk dm-14
```

Problem resolution and verification

Resolve any system-level or software issues that exist. When you verify that no system or software issues are present, issue the **#mmdiag --waiters** command again to verify whether any long waiters exist.

One possible reason for long waiters, among many, can be that Samba lock directory has been configured to be located in GPFS.

Suboptimal performance due to networking issues caused by faulty system components

The system might face networking issues, like significant network packet drops or packet errors, due to faulty system components like NIC, drivers, cables and network switch ports. This can impact the stability and the quality of the GPFS communication between the nodes, degrading the system performance.

Problem identification and verification

If IBM Spectrum Scale is configured over TCP/IP network interfaces like **10GigE** or **40GigE**, can use the **netstat –in** and **ifconfig <GPFS_iface>** commands to confirm whether any significant TX/RX packet errors or drops are happening.

In the following example, the **152326889** TX packets are dropped for the networking interface corresponding to the **ib0** device:

```
# netstat -in
Kernel Interface table
Iface      MTU    RX-OK RX-ERR RX-DRP RX-OVR    TX-OK TX-ERR TX-DRP TX-OVR Flg
ib0      65520    157606763073   0      0 0       165453186948     0     152326889       0 BMRU
```

```
#ifconfig ib0
ib0       Link encap:InfiniBand  HWaddr
80:00:00:49:FE:80:00:00:00:00:00:00:00:00:00:00:00:00:00:00
          inet addr:192.168.1.100  Bcast:192.168.1.255
Mask:255.255.255.0
          inet6 addr: fe80::f652:1403:10:bb72/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:65520  Metric:1
          RX packets:157606763073 errors:0 dropped:0 overruns:0 frame:0
          TX packets:165453186948 errors:0 dropped:152326889 overruns:0
carrier:0
```

Problem resolution and verification

Resolve low-level networking issues like bad NIC cable, or improper driver setting. If possible, shut down GPFS on the node with networking issues until the low-level networking problem is resolved. This is done so that GPFS operations on other nodes are not impacted. Issue the **# netstat -in** command to verify that the networking issues are resolved. Issue the **mmstartup** command to start GPFS on the node again. Monitor the network interface to ensure that it is operating optimally.

In the following example, no packet errors or drops corresponding to the **ib0** network interface exist.

```
# netstat -in
```

```
Kernel Interface table
Iface       MTU Met    RX-OK RX-ERR RX-DRP RX-OVR    TX-OK TX-ERR TX-DRP TX-OVR Flg
ib0       65520   0 313534358      0      0      0 301875166      0    0      0 BMRU
```

#ifconfig ib0

```
ib0       Link encap:InfiniBand  HWaddr 80:00:00:03:FE:80:00:00:00:00:00:00:00:00:00:00:00:00:00:00
                    inet addr:10.168.3.17  Bcast:10.168.255.255  Mask:255.255.0.0
                    inet6 addr: fe80::211:7500:78:a42a/64 Scope:Link
                    UP BROADCAST RUNNING MULTICAST  MTU:65520  Metric:1
                    RX packets:313534450 errors:0 dropped:0 overruns:0 frame:0
                    TX packets:301875212 errors:0 dropped:0 overruns:0 carrier:0
                    collisions:0 txqueuelen:256
                    RX bytes:241364128830 (224.7 GiB)  TX bytes:197540627923 (183.9 GiB)
```

Issues caused by the suboptimal setup or configuration of the IBM Spectrum Scale cluster

This section discusses the issues caused due to the suboptimal setup or configuration of the IBM Spectrum Scale cluster.

Suboptimal performance due to unbalanced architecture and improper system level settings

The system performance depends on the IBM Spectrum Scale cluster architecture components like servers, network, storage, disks, topology, and balance-factor. The performance is also dependent on the performance of the low-level components like network, node, and storage subsystems that make up the IBM Spectrum Scale cluster.

Problem identification

Verify whether all the layers of the IBM Spectrum Scale cluster are sized properly to meet the necessary performance requirements. The things to be considered in the IBM Spectrum Scale cluster include:

- The servers
- The network connectivity and the number of connections between the NSD client and servers
- The I/O connectivity and number of connections between the servers to the storage controller or subsystem
- The storage controller
- The disk type and the number of disks in the storage subsystem

In addition, get the optimal values for the low-level system components used in the IBM Spectrum Scale stack from the vendor, and verify whether these components are set to their optimal value. The low-level components must be tuned according to the vendor specifications for better performance.

Problem resolution and verification

It is recommended that the customer involves an IBM Spectrum Scale architect during the setup to ensure that the underlying layers of IBM Spectrum Scale cluster are capable of delivering the necessary I/O performance for the expected I/O workload.

The IBM Spectrum Scale wiki has recommendation for tuning the clusters in System X. These recommendations that are available in the following link can be used as a reference for low-level component tunings:
System X Cluster Tuning Recommendations.
However, these recommendations might not list the tuning information for all the low-level system components.

Use the low-level components benchmark values to verify that the performance of the low-level components is optimal. For example, some of the common benchmarks are xdd for block device test, OFED performance micro-benchmarks for InfiniBand, and GPFS nsdperf tool to assess the network performance.

Suboptimal performance due to low values assigned to IBM Spectrum Scale configuration parameters

Most GPFS configuration parameters have default values. For example, in IBM Spectrum Scale version 4.2 and above, the **pagepool** attribute defaults to either one-third of the physical memory on the node or 1 GiB (whichever is smaller), maxMBpS defaults to 2048 and maxFilesToCache defaults to 4000. However, if the IBM Spectrum Scale configuration parameters are explicitly set to values lower than their default values by the user, it can impact the I/O performance.

Problem identification

On the GPFS node, issue the **mmdiag --config** command to display and verify the values of the GPFS configuration parameters. Check whether these values match the optimal values set for IBM Spectrum Scale system configuration. For more information on optimal values for configuration parameter see Tuning Parameters.

Problem resolution and verification

Issue the **mmchconfig Attribute=value -i** command to set the configuration parameters to their optimal values based on the best practices followed for an IBM Spectrum Scale system configuration.

You might need to restart GPFS for certain configuration parameter values to take effect. Issue the **mmshutdown** command, followed by the **mmstartup** command to restart GPFS. Issue the **mmdiag --config** command to verify the configuration changes and updates.

Suboptimal performance due to new nodes with default parameter values added to the cluster

When new nodes are added to the IBM Spectrum Scale cluster, ensure that the GPFS configuration parameter values on the new nodes are not set to default values, unless explicitly set so by the user based on the GPFS node class. Instead, the GPFS configuration parameter values on the new nodes must be similar to the values of the existing nodes of similar type for optimal performance. The necessary system level component settings, like BIOS, network and others on the new nodes, also need to match the system level component settings of the existing nodes.

Problem identification

The **mmlsconfig** command can be used to display and verify the configuration values for a IBM Spectrum Scale cluster.

Issue the **mmdiag --config** command on the newly added GPFS nodes to verify whether the configuration parameter values for the new nodes are same as values for the existing nodes. If the newly added nodes have special roles or higher capability, then the configuration values must be adjusted accordingly.

Certain applications like SAS benefit from a larger GPFS page pool. The GPFS page pool is used to cache user file data and file system metadata. The default size of the GPFS page pool is 1 GiB in GPFS version 3.5 and higher. For SAS application, a minimum of 4 GiB page pool size is recommended. When new SAS application nodes are added to the IBM Spectrum Scale cluster, ensure that the **pagepool** attribute is set to at least 4 GiB. If left to its default value, the **pagepool** attribute is set to 1 GiB. This negatively impacts the application performance.

Problem resolution and verification

Issue the **mmchconfig Attribute=value –N <new_nodes> -i** command to set the configuration parameters to either their optimal values based on the best practices, or values similar to the existing nodes. It might be necessary to restart the GPFS daemon for the values to take effect. Issue the **mmshutdown** command, followed by the **mmstartup** command to restart the GPFS daemon. Verify the changes by running the **mmlsconfig** on a node that is part of the GPFS cluster, and the **mmdiag --config** command on the new nodes.

The following sample output shows that the value for the **pagepool** attribute on the existing application nodes **c25m3n03-ib** and **c25m3n04-ib** is set to 2G.

Note: Here **Application node** refers to NSD or SAN GPFS client nodes where applications are executed. These nodes have GPFS RPM installed for good performance.

```
#mmlsconfig
[c25m3n03-ib,c25m3n04-ib]
pagepool 2G
```

If you add new application nodes **c25m3n05-ib** and **c25m3n06-ib** to the cluster, the **pagepool** attribute and other GPFS parameter values for the new node must be set according to the corresponding parameter values for the existing nodes **c25m3n03-ib** and **c25m3n04-ib**. Therefore, the **pagepool** attribute on these new nodes must also be set to 2G by using the **mmchconfig** command.

```
mmchconfig  pagepool=2G -N c25m3n05-ib,c25m3n06-ib –i
```

Note: The **-i** option specifies that the changes take effect immediately and are permanent. This option is valid only for certain attributes. For more information on block allocation, see the *mmchconfig command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Issue the **mmlsconfig** command to verify whether all the nodes have similar values. The following sample output shows that all the nodes have **pagepool** attribute set to 2G:

```
[c25m3n03-ib,c25m3n04-ib,c25m3n05-ib,c25m3n06-ib]
pagepool 2G
```

Note: If the **pagepool** attribute is set to a custom value (2G for this example), then the **pagepool** attribute value is listed when you issue the **mmlsconfig** command. If the **pagepool** attribute is set to a default value (1G) then this will be listed when you issue the **mmlsconfig pagepool** command.

On the new node, issue the **mmdiag --config** command to verify that the new values are in effect. The sample output displays that the **pagepool** attribute value has been effectively set to 2G for the nodes **c25m3n03-ib**, **c25m3n04-ib**,**c25m3n05-ib**, **c25m3n06-ib**:

```
! pagepool 2147483648
```

Note: The exclamation (!) in the front of the parameter denotes that the value of this parameter was set by the user, and is not the default value for the parameter.

Suboptimal performance due to low value assigned to QoSIO operation classes

If Quality of Service for I/O (QoSIO) feature is enabled on the file system, verify whether any of the storage pools are assigned low values for **other** and **maintenance** class. Assigning low values for **other** and **maintenance** class can impact the performance when I/O is performed on that specific storage pool.

Problem identification

On the GPFS node, issue the **mmlsqos <fs>** command and check the **other** and **maintenance** class settings. In the sample output below, the **maintenance** class IOPS for **datapool1** storage pool is set to 200 IOPS, and the **other** class IOPS for **datapool2** storage pool is set to 400 IOPS. This IOPS value might be low for an environment with high performing storage subsystem.

```
# mmlsqos gpfs1b
QOS config::     enabled -- pool=*,other=inf,maintenance=inf:pool=datapool1,other=inf,
maintenance=200Iops:pool=datapool2,other=400Iops,maintenance=inf
QOS values::     pool=system,other=inf,maintenance=inf:pool=datapool1,other=inf,
maintenance=200Iops:pool=datapool2,other=400Iops,maintenance=inf
QOS status::     throttling active, monitoring active
```

Problem resolution and verification

On the GPFS node, issue the **mmchqos** command to change the QoS values for a storage pool in the file system. Issue the **mmlsqos** command to verify whether the changes are reflected in the QoS settings.

For example, if the IOPS corresponding to datapool2 **other** class must be set to `unlimited` then issue the following command.

```
mmchqos gpfs1b --enable pool=datapool2,other=unlimited
```

Issue the **# mmlsqos gpfs1b** command to verify whether the change is reflected.

```
# mmlsqos gpfs1b
QOS config::     enabled -- pool=*,other=inf,maintenance=inf:pool=datapool1,
other=inf,maintenance=200Iops:pool=datapool2,
other=inf,maintenance=inf
QOS values::     pool=system,other=inf,maintenance=inf:pool=datapool1,
other=inf,maintenance=200Iops:pool=datapool2,
other=inf,maintenance=inf
QOS status::     throttling active, monitoring active
```

Suboptimal performance due to improper mapping of the file system NSDs to the NSD servers

The NSDs in a file system need to be optimally assigned to the NSD servers so that the client I/O is equally distributed across all the NSD servers. For example, consider a file system with 10 NSDs and 2 NSD servers. The NSD-to-server mapping must be done in such a way that each server acts as the **primary** server for 5 of the NSD in the file system. If the NSD-to-server mapping is unbalanced, it can result in hot spots in one or more of the NSD servers. Presence of hot spots within a system can cause performance degradation.

Problem identification

Issue the **mmlsnsd** command, and verify that the primary NSD server allocated to a file system is evenly distributed.

Note: The primary server is the first server listed under the **NSD server** column for a particular file system.

On the NSD client, issue the **mmlsdisk <fs> -m** command to ensure that the NSD client I/O is distributed evenly across all the NSD servers.

In the following sample output, all the NSDs are assigned to the same primary server **c80f1m5n03ib0**.

```
# mmlsnsd
```

```
File system   Disk name    NSD servers
------------------------------------------------------------------------
 gpfs2         Perf2a_NSD01 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD02 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD03 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD04 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD05 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD06 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD07 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD08 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD09 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD10 c80f1m5n03ib0,c80f1m5n02ib0
```

In this case, all the NSD client I/O for the **gpfs2** file system are processed by the single NSD server **c80f1m5n03ib0**, instead of being equally distributed across both the NSD servers **c80f1m5n02ib0** and **c80f1m5n03ib0**. This can be verified by issuing the **mmlsdisk <fs> -m** command on the NSD client, as shown in the following sample output:

```
# mmlsdisk gpfs2 -m
Disk name     IO performed on node     Device             Availability
------------  -----------------------  -----------------  ------------
Perf2a_NSD01  c80f1m5n03ib0            -                  up
Perf2a_NSD02  c80f1m5n03ib0            -                  up
Perf2a_NSD03  c80f1m5n03ib0            -                  up
Perf2a_NSD04  c80f1m5n03ib0            -                  up
Perf2a_NSD05  c80f1m5n03ib0            -                  up
Perf2a_NSD06  c80f1m5n03ib0            -                  up
Perf2a_NSD07  c80f1m5n03ib0            -                  up
Perf2a_NSD08  c80f1m5n03ib0            -                  up
Perf2a_NSD09  c80f1m5n03ib0            -                  up
Perf2a_NSD10  c80f1m5n03ib0            -                  up
```

Problem resolution and verification

If the NSD-to-primary mapping is unbalanced, issue the **mmchnsd** command to balance the NSD distribution across the NSD servers. Issue the **mmlsnsd** command or the **mmlsdisk <fs> -m** command on the NSD client to ensure that the NSD distribution across the servers is balanced.

In the following sample output, there are 10 NSDs in the **gpfs2** file system. The NSDs are evenly distributed between the two servers, such that both servers, **c80f1m5n03ib0** and **c80f1m5n02ib0** act as primary servers for 5NSDs each.

```
# mmlsnsd
File system   Disk name    NSD servers
------------------------------------------------------------------------
 gpfs2         Perf2a_NSD01 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD02 c80f1m5n02ib0,c80f1m5n03ib0
 gpfs2         Perf2a_NSD03 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD04 c80f1m5n02ib0,c80f1m5n03ib0
 gpfs2         Perf2a_NSD05 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD06 c80f1m5n02ib0,c80f1m5n03ib0
 gpfs2         Perf2a_NSD07 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD08 c80f1m5n02ib0,c80f1m5n03ib0
 gpfs2         Perf2a_NSD09 c80f1m5n03ib0,c80f1m5n02ib0
 gpfs2         Perf2a_NSD10 c80f1m5n02ib0,c80f1m5n03ib0
```

The NSD client I/O is also evenly distributed across the two NSD servers, as seen in the following sample output:

```
# mmlsdisk gpfs2 -m
```

```
Disk name     IO performed on node     Device             Availability
------------  -----------------------  -----------------  ------------
Perf2a_NSD01  c80f1m5n03ib0            -                  up
Perf2a_NSD02  c80f1m5n02ib0            -                  up
Perf2a_NSD03  c80f1m5n03ib0            -                  up
Perf2a_NSD04  c80f1m5n02ib0            -                  up
Perf2a_NSD05  c80f1m5n03ib0            -                  up
Perf2a_NSD06  c80f1m5n02ib0            -                  up
Perf2a_NSD07  c80f1m5n03ib0            -                  up
Perf2a_NSD08  c80f1m5n02ib0            -                  up
Perf2a_NSD09  c80f1m5n03ib0            -                  up
Perf2a_NSD10  c80f1m5n02ib0            -                  up
```

Suboptimal performance due to incompatible file system block allocation type

In some cases, proof-of-concept (POC) is done on a smaller setup that consists of clusters with eight or fewer nodes and file system with eight or fewer disks. When the necessary performance requirements are met, the production file system is deployed on a larger cluster and storage setup. It is possible that on a larger cluster, the file performance per NSD is less compared to the smaller POC setup, even if all the cluster and storage component are healthy and performing optimally. In such cases, it is likely that the file system is configured with the default **cluster** block allocation type during the smaller POC setup and the larger file system setup are configured with **scatter** block allocation type.

Problem identification

Issue the **mmlsfs** command to verify the block allocation type that is in effect on the smaller and larger setup file system.

In the sample output below, the **Block allocation type** for the **gpfs2** file system is set to **scatter**.

```
# mmlsfs gpfs2 | grep 'Block allocation type'
 -j                 scatter                 Block allocation type
```

Problem resolution and verification

`layoutMap={scatter|cluster}` specifies the block allocation map type. When allocating blocks for a file, GPFS first uses a round robin algorithm to spread the data across all disks in the storage pool. After a disk is selected, the location of the data block on the disk is determined by the block allocation map type.

For cluster block allocation map type, GPFS attempts to allocate blocks in clusters. Blocks that belong to a particular file are kept adjacent to each other within each cluster. For scatter block allocation map type, the location of the block is chosen randomly. For production setup, where performance consistency throughout the life time of the file system is paramount, scatter block allocation type is recommended. The IBM Spectrum Scale storage I/O performance sizing also needs to be performed by using the scatter block allocation.

The cluster allocation method might provide better disk performance for some disk subsystems in relatively small installations. However, the benefits of clustered block allocation diminish when the number of nodes in the cluster or the number of disks in a file system increases, or when the file system's free space becomes fragmented. The cluster allocation is the default allocation method for GPFS clusters with eight or fewer nodes and for file systems with eight or fewer disks.

The scatter allocation method provides more consistent file system performance by averaging out performance variations. This is so because for many disk subsystems, the location of the data relative to the disk edge has a substantial effect on the performance. This allocation method is appropriate in most cases and is the default allocation type for GPFS clusters with more than eight nodes or file systems with more than eight disks.

The block allocation map type cannot be change after the storage pool is created. For more information on block allocation, see the *mmcrfs command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Attention: Scatter block allocation is recommended for a production setup where performance consistency is paramount throughout the lifetime of the file system. However, in an FPO environments (Hadoop or Big Data), cluster block allocation is recommended.

Issues caused by the unhealthy state of the components used

This section discusses the issues caused due to the unhealthy state of the components used in the IBM Spectrum Scale stack

Suboptimal performance due to failover of NSDs to secondary server - NSD server failure

In a shared storage configuration, failure of an NSD server might result in the failover of its NSDs to the secondary server, if the secondary server is active. This can reduce the total number of NSD servers actively serving the file system, which in turn impacts the file system's performance.

Problem identification

In IBM Spectrum Scale, the system-defined node class "nsdnodes" contains all the NSD server nodes in the IBM Spectrum Scale cluster. Issue the **mmgetstate –N nsdnodes** command to verify the state of the GPFS daemon. The GPFS file system performance might degrade if one or more NSD servers are in the **down** or **arbitrating** or **unknown** state.

The following example displays two nodes: one in **active** state and the other in **down** state

```
# mmgetstate -N nsdnodes
 Node number  Node name        GPFS state
          ------------------------------------------
                1      c25m3n07-ib      active
                2      c25m3n08-ib      down
```

Problem resolution and verification

Resolve any system-level or software issues that exist. For example, confirm that NSD server have no network connectivity problems, or that the GPFS portability modules are correctly built for the kernel that is running. Also, perform necessary low-level tests to ensure that both the NSD server and the communication to the node are healthy and stable.

Verify that no system or software issues exist, and start GPFS on the NSD server by using the **mmstartup –N <NSD_server_to_revive>** command. Use the **mmgetstate –N nsdnodes** command to verify that the GPFS daemon is in `active` state as shown:

```
# mmgetstate -N nsdnodes
Node number  Node name        GPFS state
         ------------------------------------------
               1      c25m3n07-ib      active
               2      c25m3n08-ib      active
```

Suboptimal performance due to failover of NSDs to secondary server - Disk connectivity failure

In a shared storage configuration, disk connectivity failure on an NSD server might result in failover of its NSDs to the secondary server, if the secondary server is active. This can reduce the total number of NSD servers actively serving the file system, which in turn impacts the overall performance of the file system.

Problem identification

The **mmlsnsd** command displays information about the currently defined disks in a cluster. In the following sample output, the NSD client is configured to perform file system I/O on the primary NSD server **c25m3n07-ib** for odd-numbered NSDs like **DMD_NSD01**, **DMD_NSD03**. In this case, **c25m3n08-ib** acts as a secondary server.

The NSD client is configured to perform file system I/O on the NSD server **c25m3n08-ib** for even-numbered NSDs like **DMD_NSD02**,**DMD_NSD04**. In this case, **c25m3n08-ib** is the primary server, while **c25m3n07-ib** acts as the secondary server

Issue the **#mmlsnsd** command to display the NSD server information for the disks in a file system. The following sample output shows the various disks in the **gpfs1b** file system and the NSD servers that are supposed to act as primary and secondary servers for these disks.

```
# mmlsnsd
 File system   Disk name    NSD servers
---------------------------------------------------------------------------
            gpfs1b        DMD_NSD01    c25m3n07-ib,c25m3n08-ib
            gpfs1b        DMD_NSD02    c25m3n08-ib,c25m3n07-ib
            gpfs1b        DMD_NSD03    c25m3n07-ib,c25m3n08-ib
            gpfs1b        DMD_NSD04    c25m3n08-ib,c25m3n07-ib
            gpfs1b        DMD_NSD05    c25m3n07-ib,c25m3n08-ib
            gpfs1b        DMD_NSD06    c25m3n08-ib,c25m3n07-ib
            gpfs1b        DMD_NSD07    c25m3n07-ib,c25m3n08-ib
            gpfs1b        DMD_NSD08    c25m3n08-ib,c25m3n07-ib
            gpfs1b        DMD_NSD09    c25m3n07-ib,c25m3n08-ib
            gpfs1b        DMD_NSD10    c25m3n08-ib,c25m3n07-ib
```

However, the **mmlsdisk <fsdevice> -m** command that is issued on the NSD client indicates that the NSD client is currently performing all the file system I/O on a single NSD server, **c25m3n07-ib**.

```
# mmlsdisk <fsdevice> -m
Disk name     IO performed on node     Device             Availability
          ------------  -----------------------  -----------------  ------------
            DMD_NSD01     c25m3n07-ib              -                  up
            DMD_NSD02     c25m3n07-ib              -                  up
            DMD_NSD03     c25m3n07-ib              -                  up
            DMD_NSD04     c25m3n07-ib              -                  up
            DMD_NSD05     c25m3n07-ib              -                  up
            DMD_NSD06     c25m3n07-ib              -                  up
            DMD_NSD07     c25m3n07-ib              -                  up
            DMD_NSD08     c25m3n07-ib              -                  up
            DMD_NSD09     c25m3n07-ib              -                  up
            DMD_NSD10     c25m3n07-ib              -                  up
```

Problem resolution and verification

Resolve any system-level or disk-level software issues that exist. For example, storage connectivity issues on the NSD server, or driver issues. Rediscover the NSD disk paths by using the **mmnsddiscover –a –N all** command. On the NSD client, first issue the **mmlsnsd** command to obtain the primary NSD server

configured for the NSD pertaining to a file system. The **echo "NSD-Name Primary-NSD-Server"; mmlsnsd | grep <fsdevice> | awk** command parses the output that is generated by the **mmlsnsd** command and displays the primary NSD server for each of the NSDs. Perform file I/O on the NSD client and issue the **mmlsdisk <fs> -m** command to verify that the NSD client is performing file system I/O by using all the configured NSD servers. On the NSD client, first issue the **mmlsnsd** command to obtain the configured primary NSD server for the NSD pertaining to a file system. The **# echo "NSD-Name Primary-NSD-Server"; mmlsnsd | grep <fsdevice> | awk** command parses the output that is generated by the **mmlsnsd** command and displays the primary NSD server for each of the NSDs.

```
# echo "NSD-Name Primary-NSD-Server"; mmlsnsd | grep <gpfs1b> | awk -F ',' '{print $1}' | awk
'{print $2 " " $3}'
NSD-Name   Primary-NSD-Server
DMD_NSD01   c25m3n07-ib
DMD_NSD02   c25m3n08-ib
DMD_NSD03   c25m3n07-ib
DMD_NSD04   c25m3n08-ib
DMD_NSD05   c25m3n07-ib
DMD_NSD06   c25m3n08-ib
DMD_NSD07   c25m3n07-ib
DMD_NSD08   c25m3n08-ib
DMD_NSD09   c25m3n07-ib
DMD_NSD10   c25m3n08-ib
```

Suboptimal performance due to file system being fully utilized

As a file system nears full utilization, it becomes difficult to find free space for new blocks. This impacts the performance of the write, append, and create operations.

Problem identification

On the GPFS node, issue the **mmdf <fs>** command to determine the available space.

```
# mmdf gpfs1b
disk                disk size  failure holds    holds              free KB             free KB
name                    in KB    group metadata data        in full blocks        in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
Disks in storage pool: system (Maximum disk size allowed is 18 TB)
DMD_NSD01          1756094464      101 Yes      Yes      1732298752 ( 99%)         18688 ( 0%)
DMD_NSD09          1756094464      101 Yes      Yes      1732296704 ( 99%)         13440 ( 0%)
DMD_NSD03          1756094464      101 Yes      Yes      1732304896 ( 99%)         17728 ( 0%)
DMD_NSD07          1756094464      101 Yes      Yes      1732300800 ( 99%)         14272 ( 0%)
DMD_NSD05          1756094464      101 Yes      Yes      1732298752 ( 99%)         13632 ( 0%)
DMD_NSD06          1756094464      102 Yes      Yes      1732300800 ( 99%)         13632 ( 0%)
DMD_NSD04          1756094464      102 Yes      Yes      1732300800 ( 99%)         15360 ( 0%)
DMD_NSD08          1756094464      102 Yes      Yes      1732294656 ( 99%)         13504 ( 0%)
DMD_NSD02          1756094464      102 Yes      Yes      1732302848 ( 99%)         18688 ( 0%)
DMD_NSD10          1756094464      102 Yes      Yes      1732304896 ( 99%)         18560 ( 0%)
               -------------                         -------------------- -------------------
(pool total)      17560944640                          17323003904 ( 99%)        157504 ( 0%)

               =============                         ==================== ===================
(total)           17560944640                          17323003904 ( 99%)        157504 ( 0%)

Inode Information
-----------------
Number of used inodes:            4048
Number of free inodes:          497712
Number of allocated inodes:     501760
Maximum number of inodes:     17149440
```

The UNIX command **df** also can be used to determine the use percentage (Use%) of a file system. The following sample output displays a file system with 2% capacity used.

```
# df -h
Filesystem        Size  Used Avail Use% Mounted on
/dev/gpfs1b        17T  227G   17T   2% /mnt/gpfs1b
```

Problem resolution and verification

Use the **mmadddisk** command to add new disks or NSDs to increase the GPFS file system capacity. You can also delete unnecessary files from the file system by using the **rm** command in UNIX environments to free up space.

In the sample output below, the **df –h** and **mmdf** commands show the file system use percentage to be around 2%. This indicates that the file system has sufficient capacity available.

```
# df -h
Filesystem        Size  Used Avail Use% Mounted on
/dev/gpfs1b        17T  211G   17T   2% /mnt/gpfs1b
```

```
# mmdf gpfs1b
disk                disk size  failure holds    holds              free KB             free KB
         name                    in KB    group metadata data        in full blocks        in fragments
         --------------- ------------- -------- -------- ----- -------------------- -------------------
         Disks in storage pool: system (Maximum disk size allowed is 18 TB)
         DMD_NSD01          1756094464      101 Yes      Yes      1734092800 ( 99%)         12992 ( 0%)
         DMD_NSD09          1756094464      101 Yes      Yes      1734094848 ( 99%)         14592 ( 0%)
         DMD_NSD03          1756094464      101 Yes      Yes      1734045696 ( 99%)         15360 ( 0%)
         DMD_NSD07          1756094464      101 Yes      Yes      1734043648 ( 99%)         10944 ( 0%)
         DMD_NSD05          1756094464      101 Yes      Yes      1734053888 ( 99%)         11584 ( 0%)
         DMD_NSD06          1756094464      102 Yes      Yes      1734103040 ( 99%)         11584 ( 0%)
         DMD_NSD04          1756094464      102 Yes      Yes      1734096896 ( 99%)         10048 ( 0%)
         DMD_NSD08          1756094464      102 Yes      Yes      1734053888 ( 99%)         14592 ( 0%)
         DMD_NSD02          1756094464      102 Yes      Yes      1734092800 ( 99%)         13504 ( 0%)
         DMD_NSD10          1756094464      102 Yes      Yes      1734062080 ( 99%)         13632 ( 0%)
               -------------                         -------------------- -------------------
         (pool total)      17560944640                          17340739584 ( 99%)        128832 ( 0%)

                           =============                         ==================== ===================
         (total)           17560944640                          17340739584 ( 99%)        128832 ( 0%)

Inode Information
-----------------
Number of used inodes:            4075
Number of free inodes:          497685
Number of allocated inodes:     501760
Maximum number of inodes:     17149440
```

CAUTION:
Exercise extreme caution when you delete files. Ensure that the files are no longer required for any purpose or are backed up before you delete them.

Suboptimal performance due to VERBS RDMA being inactive

IBM Spectrum Scale for Linux supports InfiniBand Remote Direct Memory Access (RDMA) using the Verbs API for data transfer between an NSD client and the NSD server. If InfiniBand (IB) VERBS RDMA is enabled on the IBM Spectrum Scale cluster, and if there is drop in the file system performance, verify whether the NSD client nodes are using VERBS RDMA for communication to the NSD server nodes. If the nodes are not using RDMA, then the communication switches to using the GPFS node's TCP/IP interface, which can cause performance degradation.

Problem identification

Issue the **mmlsconfig | grep verbsRdma** command to verify whether VERBS RDMA is enabled on the IBM Spectrum Scale cluster.

```
# mmlsconfig | grep verbsRdma
```

```
verbsRdma enable
```

If VERBS RDMA is enabled, check whether the status of VERBS RDMA on a node is **Started** by running the **mmfsadm test verbs status** command.

```
# mmfsadm test verbs status
VERBS RDMA status: started
```

The following sample output shows the various disks in the **gpfs1b** file system and the NSD servers that are supposed to act as primary and secondary servers for these disks.

```
# mmlsnsd
File system   Disk name    NSD servers
---------------------------------------------------------------------------
 gpfs1b        DMD_NSD01    c25m3n07-ib,c25m3n08-ib
 gpfs1b        DMD_NSD02    c25m3n08-ib,c25m3n07-ib
 gpfs1b        DMD_NSD03    c25m3n07-ib,c25m3n08-ib
 gpfs1b        DMD_NSD04    c25m3n08-ib,c25m3n07-ib
 gpfs1b        DMD_NSD05    c25m3n07-ib,c25m3n08-ib
 gpfs1b        DMD_NSD06    c25m3n08-ib,c25m3n07-ib
 gpfs1b        DMD_NSD07    c25m3n07-ib,c25m3n08-ib
 gpfs1b        DMD_NSD08    c25m3n08-ib,c25m3n07-ib
 gpfs1b        DMD_NSD09    c25m3n07-ib,c25m3n08-ib
 gpfs1b        DMD_NSD10    c25m3n08-ib,c25m3n07-ib
```

Issue the **mmfsadm test verbs conn** command to verify whether the NSD client node is communicating with all the NSD servers that use VERBS RDMA. In the following sample output, the NSD client node has VERBS RDMA communication active on only one of the two NSD servers.

```
# mmfsadm test verbs conn
RDMA Connections between nodes:
  destination idx cook sta cli peak cli RD cli WR  cli RD KBcli WR KB srv wait serv RD  serv WR  serv RD KB  serv WR KB   vrecv  vsend  vrecv KB  vsend KB
  ----------- --- ---  --- --- --- ------ -------- --------   --------- --- --- -------- -------- ----------- ----------- -------  -----  ---------  --------
  c25m3n07-ib  1    2  RTS   0  24   198   16395     12369    34360606  0   0    0        0          0           0          0      0       0        0
```

Problem resolution

Resolve any low-level IB RDMA issue like loose IB cables or IB fabric issues. When the low-level RDMA issues are resolved, issue system commands like **ibstat** or **ibv_devinfo** to verify whether the IB **port state** is **active**. The following system output displays the output for a **ibstat** command issued. In the sample output, the port state for **Port 1** is **Active**, while that for **Port 2** is **Down**.

```
# ibstat
CA 'mlx5_0'
        CA type: MT4113
        Number of ports: 2
        Firmware version: 10.100.6440
        Hardware version: 0
        Node GUID: 0xe41d2d03001fa210
        System image GUID: 0xe41d2d03001fa210
        Port 1:
                State: Active
                Physical state: LinkUp
                Rate: 56
                Base lid: 29
                LMC: 0
                SM lid: 1
                Capability mask: 0x26516848vverify
                Port GUID: 0xe41d2d03001fa210
                Link layer: InfiniBand
        Port 2:
                State: Down
```

```
Physical state: Disabled
Rate: 10
Base lid: 65535
LMC: 0
SM lid: 0
Capability mask: 0x26516848
Port GUID: 0xe41d2d03001fa218
Link layer: InfiniBand
```

Restart GPFS on the node and check whether the status of VERBS RDMA on a node is **Started** by running the **mmfsadm test verbs status** command.

In the following sample output, the NSD client (c25m3n03-ib) and the two NSD servers all show VERBS RDMA status as **started**.

```
# mmdsh -N nsdnodes,c25m3n03-ib '/usr/lpp/mmfs/bin/mmfsadm test verbs status'
c25m3n03-ib:  VERBS RDMA status: started
c25m3n07-ib:  VERBS RDMA status: started
c25m3n08-ib:  VERBS RDMA status: started
```

Perform a large I/O activity on the NSD client, and issue the **mmfsadm test verbs conn** command to verify whether the NSD client node is communicating with all the NSD servers that use VERBS RDMA.

In the sample output below, the NSD client node has VERBS RDMA communication active on all the active NSD servers.

```
# mmfsadm test verbs conn
RDMA Connections between nodes:
  destination  idx cook sta cli peak cli RD  cli WR cli RD KB cli WR KB srv wait serv RD serv WR  serv RD KB  serv WR KB vrecv vsend  vrecv KB  vsend KB
  ------------ --- ---  --- --- --- ------- ------- --------- --------- --- ---  ------- ------- ----------- ----------- ----- ------ --------- ---------
  c25m3n08-ib    0   3  RTS   0  13   8193   8205    17179930  17181212   0   0     0       0           0           0     0      0         0         0
  c25m3n07-ib    1   2  RTS   0  14    8192  8206    17179869  17182162   0   0     0       0           0           0     0       0         0         0
```

Issues caused by the use of configurations or commands related to maintenance and operation

This section discusses the issues caused due to the unhealthy state of the components used in the IBM Spectrum Scale stack

Suboptimal performance due to maintenance commands in progress

When in progress, long-running GPFS maintenance operations like **mmrestripefs**, **mmapplypolicy**, **mmadddisk**, and **mmdeldisk**, consume some percentage of the system resources. Significant consumption of the system resources can impact the I/O performance of the application.

Problem identification

Check the GPFS log file `/var/adm/ras/mmfs.log.latest` on the File System Manager node **mmlsmgr** to verify whether any GPFS maintenance operations are in progress.

The following sample output shows that the **mmrestripefs** operation was initiated on Jan 19 at 14:32:41, and the operation was successfully completed at 14:45:42. The I/O performance of the application is impacted during this time frame due to the execution of the **mmrestripefs** command.

```
Tue Jan 19 14:32:41.625 2016: [I] Command: mmrestripefs /dev/gpfs2 -r -N all
Tue Jan 19 14:45:42.975 2016: [I] Command: successful mmrestripefs /dev/gpfs2 -r -N all
```

Problem resolution and verification

The Quality of Service (QoS) feature for I/O operations in IBM Spectrum Scale 4.2 and higher versions is used to allocate appropriate maintenance IOPS to reduce the impact of the maintenance operation on the application. In the following sample output, the file system consists of a single storage pool – the default 'system' pool. The QoS feature is disabled and inactive.

```
# mmlsqos gpfs1a
QOS config::     disabled
QOS status::     throttling inactive, monitoring inactive
```

You can use the **mmchqos** command to allocate appropriate maintenance IOPS to the IBM Spectrum Scale system. For example, consider that the storage system has 100 K IOPS. If you want to allocate 1000 IOPS to the long running GPFS maintenance operations for the system storage pool, use the **mmchqos** command to enable the QoS feature, and allocate the IOPS as shown:

```
# mmchqos gpfs1a --enable pool=system,maintenance=1000IOPS
Adjusted QOS Class specification: pool=system,other=inf,maintenance=1000Iops
QOS configuration has been installed and broadcast to all nodes.
```

Verify the QoS setting and values on a file system by using the **mmlsqos** command.

```
# mmlsqos gpfs1aQOS config:: enabled --
          pool=system,other=inf,maintenance=1000IopsQOS status::     throttling active,
          monitoring active
```

Note: Allocating a small share of IOPS, for example 1000 IOPS, to the long running GPFS maintenance operations can increase the maintenance command execution times. So depending on the operation's needs, the IOPS assigned to the 'other' and 'maintenance' class must be adjusted by using the **mmchqos** command. This balances the application as well as the I/O requirements for the GPFS maintenance operation.

For more information on setting the QoS for I/O operations, see the *mmlsqos command* section in the *IBM Spectrum Scale: Command and Programming Reference* and *Setting the Quality of Service for I/O operations (QoS)* section in the *IBM Spectrum Scale: Administration Guide*.

Suboptimal performance due to frequent invocation or execution of maintenance commands

When the GPFS maintenance operations like **mmbackup**, **mmapplypolicy**, **mmdf**, **mmcrsnapshot**, **mmdelsnapshot**, and others are in progress, they can consume some percentage of system resources. This can impact the I/O performance of applications. If these maintenance operations are scheduled frequently, for example within every few seconds or minutes, the performance impact can be significant, unless the I/O subsystem is sized adequately to handle both the application and the maintenance operation I/O load.

Problem identification

Check the GPFS log file `/var/adm/ras/mmfs.log.latest` on the file system manager node **mmlsmgr** to verify whether any GPFS maintenance operations are being invoked frequently by a cron job or other cluster management software like **Nagios**.

In the sample output below, the **mmdf** command is being invoked periodically every 3-4 seconds.

```
Tue Jan 19 15:13:47.389 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:13:47.518 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:13:51.109 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:13:51.211 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:13:54.816 2016: [I] Command: mmdf /dev/gpfs2
```

```
Tue Jan 19 15:13:54.905 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:13:58.481 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:13:58.576 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:02.164 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:02.253 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:05.850 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:05.945 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:09.536 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:09.636 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:13.210 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:13.299 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:16.886 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:16.976 2016: [I] Command: successful mmdf /dev/gpfs2
Tue Jan 19 15:14:20.557 2016: [I] Command: mmdf /dev/gpfs2
Tue Jan 19 15:14:20.645 2016: [I] Command: successful mmdf /dev/gpfs2
```

Problem resolution and verification

Adjust the frequency of the GPFS maintenance operations so that it does not impact the applications performance. The I/O subsystem must be designed in such a way that it is able to handle both the application and the maintenance operation I/O load.

You can also use the **mmchqos** command to allocate appropriate maintenance IOPS, which can reduce the impact of the maintenance operations on the application.

Suboptimal performance when a tracing is active on a cluster

Tracing is usually enabled on the IBM Spectrum Scale cluster for troubleshooting purposes. However, running a trace on a node might cause performance degradation.

Problem identification

Issue the **mmlsconfig** command and verify whether GPFS tracing is configured. The following sample output displays a cluster in which tracing is configured:

```
# mmlsconfig | grep trace
trace all 4 tm 2 thread 1 mutex 1 vnode 2 ksvfs 3 klockl 2 io 3 pgalloc 1 mb 1 lock 2 fsck 3
tracedevOverwriteBufferSize 1073741824
tracedevWriteMode overwrite 268435456
```

Issue the **# ps -aux | grep lxtrace | grep mmfs** command to determine whether GPFS tracing process is running on a node. The following sample output shows that GPFS tracing process is running on the node:

```
# ps -aux | grep lxtrace | grep mmfs
root     19178  0.0  0.0  20536   128 ?        Ss   14:06   0:00
/usr/lpp/mmfs/bin/lxtrace-3.10.0-229.el7.x86_64 on
/tmp/mmfs/lxtrace.trc.c80f1m5n08ib0 --overwrite-mode --buffer-size
268435456
```

Problem resolution and verification

When the traces have met their purpose and are no longer needed, use one of the following commands to stop the tracing on all nodes:

- Use this command to stop tracing:

  mmtracectl --stop –N all
- Use this command to clear all the trace setting variables and stop the tracing:

  mmtracectl --off –N all

Suboptimal performance due to replication settings being set to 2 or 3

The file system write performance depends on the write performance of the storage volumes and its RAID configuration. However, in case the backend storage write performance is on par with its read performance, but the file system write performance is just 50% (half) or 33% (one-third) of the read performance, check if the file system replication is enabled.

Problem identification

When file system replication is enabled and set to 2, effective write performance becomes 50% of the raw write performance, since for every write operation, there are two internal write operation due to replication. Similarly, when file system replication is enabled and set to 3, effective write performance becomes approximately 33% of the raw write performance, since for every write operation, there are three internal write operation.

Issue the **mmlsfs** command, and verify the default number of metadata and data replicas enabled on the file system. In the following sample output the metadata and data replication on the file system is set to 2:

```
# mmlsfs <fs> | grep replica | grep -i default
 -m                2                        Default number of metadata replicas
 -r                2                       Default number of data replicas
```

Issue the **mmlsattr** command to check whether replication is enabled at file level

```
# mmlsattr -L largefile.foo | grep replication
 metadata replication: 2 max 2
data replication:     2 max 2
```

Problem resolution and verification

The GPFS placement policy can be enforced to set the replication factor of temporary files for non-critical datasets to one. For example, temporary files like log files that can be recreated if necessary.

Follow these steps to set the replication value for log files to 1:

1. Create a `placement_policy.txt` file by using the following rule:
   ```
   rule 'non-replicate-log-files' SET POOL 'SNCdata' REPLICATE (1) where lower(NAME) like '%.log'
   rule 'default' SET POOL 'SNCdata'
   ```
2. Install the placement policy on the file system by using the following command:
   ```
   mmchpolicy <fs> placement_policy.txt
   ```

   Note: You can test the placement policy before installing it by using the following command:
   ```
   mmchpolicy <fs> placement_policy.txt -I test
   ```
3. Issue one of the following commands to remount the file system for the policy to take effect:
   Remount the file system on all the nodes by using one of the following commands:
   - **mmumount <fs> -N all**
   - **mmmount <fs> -N all**
4. Issue the **mmlspolicy <fs> -L** command to verify whether the output is as shown:
   ```
   rule 'non-replicate-log-files' SET POOL 'SNCdata' REPLICATE (1) where lower(NAME) like '%.log'
   rule 'default' SET POOL 'SNCdata'
   ```

Suboptimal performance due to updates made on a file system or fileset with snapshot

If a file is modified after its snapshot creation, the system can face performance degradation due to the copy-on-write property enforced on updated data files.

Problem identification

Updating a file that has a snapshot might create unnecessary load on a system because each application update or write operation goes through the following steps:

1. Read the original data block pertaining to the file region that must be updated.
2. Write the data block read in the step 1 above to the corresponding snapshot location.
3. Perform the application write or update operation on the desired file region.

Issue the **mmlssnapshot** to verify whether the snapshot was created before the file data update operation.

In the following sample output, the **gpfs2** file system contains a snapshot.

```
# mmlssnapshot gpfs2
Snapshots in file system gpfs2:
Directory               SnapId    Status  Created
snap1                   2         Valid   Mon Jan 25 12:42:30 2016
```

Problem resolution and verification

Use the **mmdelsnapshot** command to delete the file system snapshot, if it is no longer necessary. For more information on the **mmdelsnapshot** command, see the *mmdelsnapshot command* in the *IBM Spectrum Scale: Command and Programming Reference*.

Delays and deadlocks

The first item to check when a file system appears hung is the condition of the networks including the network used to access the disks.

Look for increasing numbers of dropped packets on all nodes by issuing:

- The **netstat -D** command on an AIX node.
- The **ifconfig** *interfacename* command, where *interfacename* is the name of the interface being used by GPFS for communication.

When using subnets (see the *Using remote access with public and private IP addresses* topic in the *IBM Spectrum Scale: Administration Guide* .), different interfaces may be in use for intra-cluster and intercluster communication. The presence of a hang or dropped packed condition indicates a network support issue that should be pursued first. Contact your local network administrator for problem determination for your specific network configuration.

If file system processes appear to stop making progress, there may be a system resource problem or an internal deadlock within GPFS.

Note: A deadlock can occur if user exit scripts that will be called by the **mmaddcallback** facility are placed in a GPFS file system. The scripts should be placed in a local file system so they are accessible even when the networks fail.

To debug a deadlock, do the following:

1. Check how full your file system is by issuing the **mmdf** command. If the **mmdf** command does not respond, contact the IBM Support Center. Otherwise, the system displays information similar to:

```
disk                disk size  failure holds    holds              free KB             free KB
name                    in KB    group metadata data        in full blocks        in fragments
--------------- ------------- -------- -------- ----- -------------------- -------------------
Disks in storage pool: system (Maximum disk size allowed is 1.1 TB)
dm2                 140095488        1 yes      yes       136434304 ( 97%)        278232 ( 0%)
dm4                 140095488        1 yes      yes       136318016 ( 97%)        287442 ( 0%)
dm5                 140095488     4000 yes      yes       133382400 ( 95%)        386018 ( 0%)
dm0nsd              140095488     4005 yes      yes       134701696 ( 96%)        456188 ( 0%)
dm1nsd              140095488     4006 yes      yes       133650560 ( 95%)        492698 ( 0%)
dm15                140095488     4006 yes      yes       140093376 (100%)            62 ( 0%)
                -------------                         -------------------- -------------------
(pool total)        840572928                             814580352 ( 97%)       1900640 ( 0%)

                =============                         ==================== ===================
(total)             840572928                             814580352 ( 97%)       1900640 ( 0%)

Inode Information
-----------------
Number of used inodes:            4244
Number of free inodes:          157036
Number of allocated inodes:     161280
Maximum number of inodes:       512000
```

GPFS operations that involve allocation of data and metadata blocks (that is, file creation and writes) will slow down significantly if the number of free blocks drops below 5% of the total number. Free up some space by deleting some files or snapshots (keeping in mind that deleting a file will not necessarily result in any disk space being freed up when snapshots are present). Another possible cause of a performance loss is the lack of free inodes. Issue the **mmchfs** command to increase the number of inodes for the file system so there is at least a minimum of 5% free. If the file system is approaching these limits, you may notice the following error messages:

6027-533 [W]
Inode space *inodeSpace* in file system *fileSystem* is approaching the limit for the maximum number of inodes.

operating system error log entry
Jul 19 12:51:49 node1 mmfs: Error=MMFS_SYSTEM_WARNING, ID=0x4DC797C6, Tag=3690419: File system warning. Volume fs1. Reason: File system fs1 is approaching the limit for the maximum number of inodes/files.

2. If automated deadlock detection and deadlock data collection are enabled, look in the latest GPFS log file to determine if the system detected the deadlock and collected the appropriate debug data. Look in **/var/adm/ras/mmfs.log.latest** for messages similar to the following:

```
Thu Feb 13 14:58:09.524 2014: [A] Deadlock detected: 2014-02-13 14:52:59: waiting 309.888 seconds on node
p7fbn12: SyncHandlerThread 65327: on LkObjCondvar, reason 'waiting for RO lock'
Thu Feb 13 14:58:09.525 2014: [I] Forwarding debug data collection request to cluster manager p7fbn11 of
cluster cluster1.gpfs.net
Thu Feb 13 14:58:09.524 2014: [I] Calling User Exit Script gpfsDebugDataCollection: event deadlockDebugData,
Async command /usr/lpp/mmfs/bin/mmcommon.
Thu Feb 13 14:58:10.625 2014: [N] sdrServ: Received deadlock notification from 192.168.117.21
Thu Feb 13 14:58:10.626 2014: [N] GPFS will attempt to collect debug data on this node.
mmtrace: move /tmp/mmfs/lxtrace.trc.p7fbn12.recycle.cpu0
/tmp/mmfs/trcfile.140213.14.58.10.deadlock.p7fbn12.recycle.cpu0
mmtrace: formatting /tmp/mmfs/trcfile.140213.14.58.10.deadlock.p7fbn12.recycle to
/tmp/mmfs/trcrpt.140213.14.58.10.deadlock.p7fbn12.gz
```

This example shows that deadlock debug data was automatically collected in **/tmp/mmfs**. If deadlock debug data was not automatically collected, it would need to be manually collected.

To determine which nodes have the longest waiting threads, issue this command on each node:

```
/usr/lpp/mmfs/bin/mmdiag --waiters waitTimeInSeconds
```

For all nodes that have threads waiting longer than *waitTimeInSeconds* seconds, issue:

```
mmfsadm dump all
```

Notes:

a. Each node can potentially dump more than 200 MB of data.
b. Run the **mmfsadm dump all** command only on nodes that you are sure the threads are really hung. An **mmfsadm dump all** command can follow pointers that are changing and cause the node to crash.

3. If the deadlock situation cannot be corrected, follow the instructions in "Additional information to collect for delays and deadlocks" on page 470, then contact the IBM Support Center.

Chapter 24. GUI issues

The topics listed in this section provide the list of most frequent and important issues reported with the IBM Spectrum Scale GUI.

Related concepts:

Chapter 2, “Monitoring system health using IBM Spectrum Scale GUI,” on page 99

“Collecting diagnostic data through GUI” on page 235
IBM Support might ask you to collect logs, trace files, and dump files from the system to help them resolve a problem. You can perform this task from the management GUI or by using the **gpfs.snap** command. Use the **Settings** > **Diagnostic Data** page in the IBM Spectrum Scale GUI to collect details of the issues reported in the system.

Understanding GUI support matrix and limitations

It is important to understand the supported versions and limitations to analyze whether you are facing a real issue in the system.

The IBM Spectrum Scale FAQ in IB Knowledge Center contains the GUI support matrix. The IBM Spectrum Scale FAQ is available at http://www.ibm.com/support/knowledgecenter/STXKQY/gpfsclustersfaq.html.

To know more about GUI limitations, see *GUI limitations* in *IBM Spectrum Scale: Administration Guide*.

Known GUI issues

There are a set of known issues in the GUI. Most of the problems get fixed when you upgrade the GUI to the version that contains the fix. The list of known issues in the GUI are available at: IBM Spectrum Scale GUI known issues.

The following topics also cover some of the examples for the most frequent GUI issues and their resolutions.

GUI fails to start

This issue is primarily because of the database issue. In ideal scenarios, the service script automatically initializes and starts postgreSQL. However, in rare cases, the database might be either inconsistent or corrupted.

If the postgreSQL database is corrupted, it might be because of the following reasons:

- The additional (non-distro) postgreSQL package is installed and it occupies the port 5432.
- Details that are stored in the `/etc/hosts` file are corrupted so the "localhost" is not listed as the first item for the IP`127.0.0.1`.
- An incompatible schema exists in the database from a previous release.

If the GUI logs show any of the database errors, try the following steps:

1. Issue **systemctl stop gpfsgui** to stop GUI services.
2. Issue `'su postgres -c 'psql -d postgres -c "DROP SCHEMA FSCC CASCADE"''`.
3. If the previous step does not help, issue `'rm -rf /var/lib/pgsql/data'`.
4. Issue **systemctl start gpfsgui** to start GUI.

If the problem still persists, it might be because of a corrupted GUI installation, missing GUI dependency, or some other unknown issue. In this scenario, you can remove and reinstall the GUI rpm. For more information on how to install and uninstall GUI rpms, see *Manually installing IBM Spectrum Scale management GUI* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide*.

You can collect the logs that are available in the `/var/log/cnlog/mgtsrv` folder to investigate further. You can also use the **gpfs.snap** command as shown in the following example to collect logs and dumps in case of a GUI issue:

```
gpfs.snap -N GUI_MGMT_SERVERS
```

Collecting logs and dumps through the **gpfs.snap** command also collects the GPFS logs. So, manually getting the logs from the folder `/var/log/cnlog/mgtsrv` is quicker and provides only the required data that is required to search for the details of the GUI issue.

GUI login page does not open

The management GUI is accessible through the following URL after the installation: `https://<ip or host name>`.

If the GUI login page does not open, try out the following:

1. Issue the following command to verify the status:

   ```
   systemctl status gpfsgui
   ```

2. Check the status of java components by issuing the following command:

   ```
   netstat -lnp | grep java
   ```

   The system must display the following output in the ideal scenarios:

   There can be more lines in the output as given in the following example but the port 443 is the most important for the GUI service:

   ```
   tcp6 0 0 :::47080 :::* LISTEN 13924/java
   tcp6 0 0 :::47443 :::* LISTEN 13924/java
   tcp6 0 0 127.0.0.1:4444 :::* LISTEN 13924/java
   ```

   Note:

   - The IBM Spectrum Scale GUI WebSphere® Java process no longer runs as root but as a user named *scalemgmt*. The GUI process now runs on port 47443 and 47080 and uses iptables rules to forward port 443 to 47443 and 80 to 47080.
   - The port 4444 is used by the GUI CLI to interact with the GUI back-end service. Other ports that are listed here are used by Java internally.

   If you find that the port 443 is not opened by WebSphere Liberty, restart the GUI service by issuing the **systemctl restart gpfsgui** command. The GUI uses the default HTTPS port 443. If some other application or process listens to this port, it causes a port conflict and the GUI does not work.

GUI performance monitoring issues

The sensor gets the performance data for the collector. The collector application that is called *pmcollector* runs on every GUI node to display the performance details in the GUI. A sensor application is running on every node of the system.

If GUI is not displaying the performance data, the following might be the reasons:

1. Collectors are not enabled
2. Sensors are not enabled
3. NTP failure

Collectors are not enabled

Do the following to verify whether collectors are working properly:

1. Issue **systemctl status pmcollector** on the GUI node to confirm that the collector is running.
2. If collector service is not started already, start the collector on the GUI nodes by issuing the **systemctl restart pmcollector** command. Depending on the system requirement, the *pmcollector* service can be configured to be run on the nodes other than GUI nodes. You need to verify the status of *pmcollector* service on all nodes where collector is configured.
3. If you cannot start the service, verify its log file that is located at `/var/log/zimon/ZIMonCollector.log` to see whether it logs any other details of the issues related to the collector service status.
4. Use a sample CLI query to test if data collection works properly. For example:

   ```
   mmperfmon query cpu_user
   ```

Note: After migrating from release 4.2.0.x or later to 4.2.1 or later, you might see the *pmcollector* service critical error on GUI nodes. In this case, restart the *pmcollector* service by running the **systemctl restart pmcollector** command on all GUI nodes.

Sensors are not enabled

The following table lists sensors that are used to get the performance data for each resource type:

Table 57. Sensors available for each resource type

Resource type	Sensor name	Candidate nodes
Network	Network	All
System Resources	CPU	All
	Load	
	Memory	
NSD Server	GPFSNSDDisk	NSD Server nodes
IBM Spectrum Scale Client	GPFSFilesystem	IBM Spectrum Scale Client nodes
	GPFSVFS	
	GPFSFilesystemAPI	
NFS	NFSIO	Protocol nodes running NFS service
SMB	SMBStats	Protocol nodes running SMB service
	SMBGlobalStats	
CTDB	CTDBStats	Protocol nodes running SMB service
Object	SwiftAccount	Protocol nodes running Object service
	SwiftContainer	
	SwiftObject	
	SwiftProxy	
Transparent Cloud Tiering	MCStoreGPFSStats	Cloud gateway nodes
	MCStoreIcstoreStats	
	MCStoreLWEStats	
Capacity	DiskFree	All nodes
	GPFSFilesetQuota	Only a single node
	GPFSDiskCap	Only a single node

Do the following to verify whether sensors are working properly:

1. Confirm that the sensor is configured correctly by issuing the **mmperfmon config show** command. This command lists the content of the sensor configuration that is located at `/opt/IBM/zimon/ZIMonSensors.cfg`.
2. The configuration must point to the node where the collector is running and all the expected sensors must be enabled. An enabled sensor has a period greater than 0 in the same configuration file. After the configuration file is updated, the *pmsensor* service needs to be restarted.
3. Issue **systemctl start pmsensors** to start the service if it is stopped.

If sensors and collectors are properly configured and enabled, you can issue the **mmperfmon** and **mmpmon** commands to see whether performance data is really generated.

You can query the data displayed in the performance charts through CLI as well. For more information on how to query performance data displayed in GUI, see "Querying performance data shown in the GUI through CLI" on page 95.

NTP failure

The performance monitoring fails if the clock is not properly synchronized in the cluster. Issue the **ntpq -c peers** command to verify the NTP state.

Related concepts:

"Performance monitoring using IBM Spectrum Scale GUI" on page 87
The IBM Spectrum Scale GUI provides a graphical representation of the status and historical trends of the key performance indicators. This helps the users to make decisions easily without wasting time.

Chapter 23, "Performance issues," on page 407
The performance issues might occur because of the system components or configuration or maintenance issues.

GUI is showing "Server was unable to process the request" error

The GUI might not respond on user actions or it might show "Server was unable to process the request" error. This might be because of an issue in the JavaScript layer, which runs on the browser. JavaScript errors are not collected in the diagnostic data. The IBM Support might need the JavaScript error details to troubleshoot this issue.

The location where the JavaScript console can be accessed depends on the web browser.

- **For Google Chrome:** Select menu item **Tools > Javascript Console**.
- **For Mozilla Firefox:** Install and run the firebug plug-in to get the JavaScript console.

GUI is displaying outdated information

The IBM Spectrum Scale GUI caches configuration data in an SQL database. Refresh tasks update the cached information. Many refresh tasks are invoked by events when the configuration is changed in the cluster. In those cases, the GUI pages reflect changes in a minute. For certain types of data, events are not raised by itself to invoke the refresh tasks. In such cases, the system must poll the data on a regular interval to reflect up-to-date information in the GUI pages. All the refresh tasks run on a schedule. The system also polls the data frequently even for those tasks that are triggered by events.

If the GUI shows stale data and the user does not want to wait until the next issue of refresh task, you can run those refresh tasks manually as shown in the following example:

```
/usr/lpp/mmfs/gui/cli/runtask <task_name>
```

Note: Many file system-related tasks require the corresponding file system to be mounted on the GUI to collect data.

The following table lists the details of the available GUI refresh tasks.

Table 58. GUI refresh tasks

Refresh task	Frequency	Collected information	Prerequisite - File system must be mounted	Invoked by event	CLI commands used
AFM_FILESET_STATE	60	The AFM fileset status	Yes	Any event for component AFM	`mmafmctl getstate -Y`
AFM_NODE_MAPPING	720	The AFM target map definitions	No	On execution of mmafmconfig	`mmafmconfig show -Y`
ALTER_HOST_NAME	12 h	Host names and IP addresses in Monitor > Nodes page			`mmremote networkinfo`
CALLBACK	6 h	Checks and registers callbacks used by GUI		Yes	`mmlscallback` and `mmaddcallback`
CALLHOME	1 h	Call home configuration data	No	On execution of `mmcallhome` command	`mmcallhome capability list -Y`, `mmcallhome info list -Y`, `mmcallhome proxy list -Y`, and `mmcallhome group list -Y`
CALLHOME_STATUS	1 m	Callhome status information	No		`mmcallhome status list -Y`
CES_ADDRESS	1 h	CES IP addresses in Monitor > Nodes page		Yes	`mmces node list`
CES_STATE	10 min	CES state in Monitor > Nodes		Yes	`mmces state show -N cesNodes` `mmces events active -N cesNodes` (used for the information field)
CES_SERVICE_STATE	1 h	CES service state in Monitor > Nodes page		Yes	`mmces service list -N cesNodes -Y`
CES_USER_AUTH_ SERVICE	1 h	Not displayed		Yes	`mmuserauth service list -Y`
CHECK_FIRMWARE	6 h	Monitor > Events page			Checks whether the reported firmware is up to date
CLUSTER_CONFIG	1 h	List of nodes and node classes in Monitoring > Nodes		Yes	`mmsdrquery` and `mmlsnodeclass`
CONNECTION_ STATUS	10 min	Connections status in Monitoring > Nodes page			Nodes reachable through SSH
DAEMON_ CONFIGURATION	1 h	Not displayed		Yes	`mmlsconfig`
DF	1 h	Not directly displayed; used to generate low space events	Yes	Yes	`df`, `df -i`, `mmlspool`
DIGEST_NOTIFICATION_ TASK	Once a day at 04:15AM	Sends daily event reports if configured	No		
DISK_USAGE	3:00 AM	Disk usage information. Not directly displayed; used to generate low space events	Yes		`mmdf`, `mmsdrquery` (`mmlsnsd` and `mmremote getdisksize` for non-GNR-NSDs that is not assigned to the file system)
DISKS	1 h	NSD list in Monitoring > NSDs		Yes	`mmsqrquery`, `mmlsnsd`, and `mmlsdisk`
FILESYSTEM_MOUNT	1 h	Mount state in Files > File Systems		Yes	`mmlsmount`
FILESYSTEMS	1 h	List of file systems in Files > File Systems	Yes	Yes	`mmsdrquery`, `mmlsfs`, , and `mmlssnapdir`
GUI_CONFIG_CHECK	12 h	Checks that cluster configuration is compatible with GUI requirements		Yes	`mmsdrquery`, `mmgetstate`, and `getent`

HEALTH_STATES	10 min	Health events in Monitoring > Events		Yes	`mmhealth node show {component} -v -N {nodes} -Y`
					
					`mmhealth node eventlog -Y`

HOST_STATES	1 h	GPFS state in Monitoring > Nodes		Yes	`mmgetstate`

Table 58. GUI refresh tasks (continued)

Refresh task	Frequency	Collected information	Prerequisite - File system must be mounted	Invoked by event	CLI commands used
HOST_STATES_CLIENTS	3 h	Information about GPFS clients	No	On quorum-related events	`mmgetstate -N 'clientLicense' -Y`
LOG_REMOVER	6 h	Deletes aged database entries	No		
MASTER_GUI_ELECTION	1 m	Checks if all GUIs in the cluster are running and elects a new master GUI if needed.	No		HTTP call to other GUIs
MOUNT_CONFIG	12 h	Mount configuration	No	On execution of any mm*fs command	Internal commands
NFS_EXPORTS	1 h	Exports in Protocols > NFS Exports		Yes	`mmcesservice list` and `mmcesnfslsexport`
NFS_EXPORTS_ DEFAULTS	1 h	Not displayed		Yes	`mmcesservice list` and `mmcesnfslscfg`
NFS_SERVICE	1 h	NFS settings in Settings > NFS Service		Yes	`mmcesservice list` and `mmcesnfslscfg`
NODECLASS	6 h	Node classes in Monitor>Nodes		Yes	`mmlsnodeclass`
NODE_LICENSE	6 h	Node license information	No	On execution of `mmchlicense` command	`mmlslicense -Y`
OBJECT_STORAGE_ POLICY	6 h	Storage policies of containers in Object > Accounts		Yes	`mmobj policy list`
OS_DETECT	6 h	Information about operating system, cpu architecture, hardware vendor, type, serial in Monitoring > Nodes		Yes	`mmremote nodeinfo`
PM_MONITOR	10 min	Checks if the performance collector is up and running and also checks the CPU data for each node			`systemctl status pmcollector` and `zimon query`
PM_SENSORS	6 h	Performance monitoring sensor configuration	No	On execution of `mmperfmon`	`mmperfmon config show`
PM_TOPOLOGY	1 h	Performance data topology	No		perfmon query
POLICIES	1 h	Policies in Files > Information Lifecycle	Yes	Yes	`mmlspolicy`
QUOTA	2:15 AM	Quotas in Files > Quota Fileset capacity in Monitoring > Capacity	Yes	Yes	`mmrepquota` and `mmlsdefaultquota`
QUOTA_MAIL	Once a day at 05:00 AM	Sends daily quota reports if configured	No		
RDMA_INTERFACES	12 h	Information about the RDMA interfaces	No		`mmremote ibinfo`
REMOTE_CLUSTER	10 m	General information about remote clusters	No		REST API call to remote GUIs
REMOTE_CONFIG	1 h	Not displayed		Yes	`mmauth, gets and parses sdr file`
REMOTE_FILESETS	1 h	Information about filesets of remote clusters	No		REST API call to remote GUIs
REMOTE_GPFS_CONFIG	3 h	The GPFS configuration of remote clusters	No		REST API call to remote GUIs
REMOTE_HEALTH _STATES	15 m	The health states of remote clusters	No		REST API call to remote GUIs
SMB_GLOBALS	1 h	SMB settings in Settings > SMB Service		Yes	`mmcessmblsconfig`
SMB_SHARES	1 h	Shares in Protocols > SMB Shares		Yes	`mmcessmblsexport`
SNAPSHOTS	1 h	Snapshots in Files > Snapshots	Yes	Yes	`mmlssnapshot`

Table 58. GUI refresh tasks (continued)

Refresh task	Frequency	Collected information	Prerequisite - File system must be mounted	Invoked by event	CLI commands used
SNAPSHOT_MANAGER	1 m	Creates scheduled snapshots			`mmcrsnapshot` and `mmdelsnapshot`
SYSTEMUTIL_DF	1 h	Used to generate warnings if nodes run out of local disk space			Checks local disk space of node
STORAGE_POOL	1 h	Pool properties in Files > File Systems		Yes	`mmlspool <device> all -L -Y`
TCT_ACCOUNT	1 h	Information about TCT accounts	No	On execution of `mmcloudgateway` command	`mmcloudgateway account list`
TCT_CLOUD_SERVICE	1 h	Information about TCT cloud services	No	On execution of `mmcloudgateway` command	`mmcloudgateway service list`
TCT_NODECLASS	1 h	Information about TCT node classes	No	On execution of `mmchnode` command	`mmlsnodeclass`
THRESHOLDS	3 h	Information about the configured thresholds	No	On execution of `mmhealth thresholds`	`mmhealth thres`

Capacity information is not available in GUI pages

The IBM Spectrum Scale management GUI does not display the capacity information on various GUI pages if *GPFSDiskCap* and *GPFSFilesetQuota* sensors are disabled and quota is disabled on the file system.

The following table provides the solution for the capacity data display issues in the corresponding GUI pages.

Table 59. Troubleshooting details for capacity data display issues in GUI

GUI page	Solution
Files > **File SystemsStorage** > **Pools**	Verify whether the *GPFSPool* sensor is enabled on at least one node and ensure that the file system is mounted on this node. The health subsystem might have enabled this sensor already. The default period for the *GPFSPool* sensor is 300 seconds (5 minutes).
Files > **Filesets** does not display fileset capacity details.	In this case, the quota is not enabled for the file system that hosts this fileset. Go to **Files** > **Quotas** page and enable quotas for the corresponding file system. By default, the quotas are disabled for all file systems.
Monitoring > **Statistics**	Verify whether the *GPFSDiskCap* and *GPFSFilesetQuota* sensors are enabled and quota is enabled for the file systems. For more information on how to enable performance monitoring sensors, see "Configuring performance monitoring options in GUI" on page 89.

Chapter 25. AFM issues

The following table lists the common questions in AFM.

Table 60. Common questions in AFM with their resolution

Question	Answer / Resolution
How do I flush requeued messages?	Sometimes, requests in the AFM messages queue on the gateway node get requeued because of errors at the home cluster. For example, if space is not available at the home cluster to perform a new write, a write message that is queued is not successful and gets requeued. The administrator views the failed message being requeued on the Primary gateway. Add more space to the home cluster and run **mmafmctl resumeRequeued** so that the requeued messages are executed at home again. If **mmafmctl resumeRequeued** is not run by an administrator, AFM executes the message in the regular order of message executions from the cache cluster to the home cluster. Running the **mmfsadm saferdump afm all** command on the gateway node displays the queued messages. The requeued messages are displayed in the dumps. An example: `c12c4apv13.gpfs.net: Normal Queue: (listed by execution order) (state: Active) c12c4apv13.gpfs.net: Write [612457.552962] requeued file3 (43 @ 293) chunks 0 bytes 0 0`
Why is a fileset in the Unmounted or Disconnected state when parallel I/O is set up?	Filesets that are using a mapping target go to the Disconnected mode if the NFS server of the Primary gateway is unreachable, even if NFS servers of all participating gateways are reachable. The NFS server of the Primary gateway must be checked to fix this problem.
How do I activate an inactive fileset?	The **mmafmctl prefetch** command without options, where prefetch statistics are procured, activates an inactive fileset.
How do I reactivate a fileset in the Dropped state?	The **mmafmctl prefetch** command without options, where prefetch statistics are procured, activates a fileset in a dropped state.
How to clean unmount the home filesystem if there are caches using GPFS protocol as backend?	To have a clean unmount of the home filesystem, the filesystem must first be unmounted on the cache cluster where it is remotely mounted and the home filesystem must be unmounted. Unmounting the remote file system from all nodes in the cluster might not be possible until the relevant cache cluster is unlinked or the local file system is unmounted. Force unmount, shutdown, or crash of the remote cluster results in panic of the remote filesystem at the cache cluster and the queue is dropped. The next access to the fileset runs the recovery. However, this should not affect the cache cluster.
What should be done if the **df** command hangs on the cache cluster?	On RHEL 7.0 or later, **df** does not support hidden NFS mounts. As AFM uses regular NFS mounts on the gateway nodes, this change causes commands like **df** to hang if the secondary gets disconnected. The following workaround can be used that allows NFS mounts to continue to be hidden: Remove /etc/mtab symlink, and create a new file /etc/mtab and copy /proc/mounts to /etc/mtab file during the startup. In this solution, the mtab file might go out of synchronization with /proc/mounts.
What happens when the hard quota is reached in an AFM cache?	Like any filesystem that reaches the hard quota limit, requests fail with `E_NO_SPACE`.

Table 60. Common questions in AFM with their resolution (continued)

Question	Answer / Resolution
When are inodes deleted from the cache?	After an inode is allocated, it is never deleted. The space remains allocated and they are re-used.
If inode quotas are set on the cache, what happens when the inode quotas are reached?	Attempts to create new files fail, but cache eviction is not triggered. Cache eviction is triggered only when block quota is reached, not the inode quotas.
How can the cache use more inodes than the home?	One way is for file deletions. If a file is renamed at the home site, the file in cache is deleted and created again in cache. This results in the file being assigned a different inode number at the cache site. Also, if a cache fileset is LU mode or SW mode, then there can be changes made at the cache that cause it to be bigger than the home.
Why does fileset go to Unmounted state even if home is accessible on the cache cluster?	Sometimes, it is possible that the same home is used by multiple clusters, one set of filesets doing a quiesce turn the home unresponsive to the second cluster's filesets, which show home as unmounted
What could be impact of not running **mmafmconfig** command despite having a GPFS home?	Sparse file support is not present even if home is GPFS. Recovery and many AFM functions do not work. Crashes can happen for `readdir` or `lookup`, if the backend is using NSD protocol and remote mount is not available at the gateway node.
What should be done if there are cluster wide waiters but everything looks normal, such as home is accessible from gateway nodes, applications are in progress on the cache fileset?	This can happen when the application is producing requests at a faster pace. Check **iohist** to check disk rates.
Read seems to be stuck/inflight for a long time. What should be done?	Restart **nfs** at home to see if error resolves. Check the status of the fileset using **mmafmctl getstate** command to see if you fileset is in unmounted state.
The `mmfs.log` show errors during read such as error 233 :	These are temporary issues during read:`Tue Feb 16 03:32:40.300 2016: [E] AFM: Read file system fs1 fileset newSanity-160216-020201-KNFS-TC8-SW file IDs [58195972.58251658.-1.-1,R] name file-3G remote error 233` These go away automatically and read should be successful.
Can the home have different sub-directories exported using unique FSIDs, while parent directory is also exported using an FSID?	This is not a recommended configuration.
I have a non-GPFS home, I have applications running in cache and some requests are requeued with the following error: `SetXAttr file system fs1 fileset sw_gpfs file IDs [-1.1067121.-1.-1,N] name local error 124`	**mmafmconfig** is not setup at home. Running **mmafmconfig** command at home and relinking cache should resolve this issue.
During failover process, some gateway nodes might show error 233 in `mmfs.log`.	This error is harmless. The failover completes successfully.
Resync fails with `No buffer space available` error, but **mmdiag --memory** shows that memory is available.	Increase **afmHardMemThreshold**.

Table 60. Common questions in AFM with their resolution (continued)

Question	Answer / Resolution
How can I change the mode of a fileset?	The mode of an AFM client cache fileset cannot be changed from local-update mode to any other mode; however, it can be changed from read-only to single-writer (and vice versa), and from either read-only or single-writer to local-update. Complete the following steps to change the mode: 1. Ensure that fileset status is active and that the gateway is available. 2. Unmount the file system 3. Unlink the fileset. 4. Run the **mmchfileset** command to change the mode. 5. Mount the file system again. 6. Link the fileset again.
Why are **setuid** or **setgid** bits in a single-writer cache reset at home after data is appended?	The **setuid** or **setgid** bits in a single-writer cache are reset at home after data is appended to files on which those bits were previously set and synced. This is because over NFS, a write operation to a **setuid** file resets the **setuid** bit.
How can I traverse a directory that is not cached?	On a fileset whose metadata in all subdirectories is not cached, any application that optimizes by assuming that directories contain two fewer subdirectories than their hard link count do not traverse the last subdirectory. One such example is find; on Linux, a workaround for this is to use find **-noleaf** to correctly traverse a directory that has not been cached
What extended attribute size is supported?	For an operating system in the gateway whose Linux kernel version is below 2.6.32, the NFS max **rsize** is 32K, so AFM does not support an extended attribute size of more than 32K on that gateway.
What should I do when my file system or fileset is getting full?	The `.ptrash` directory is present in cache and home. In some cases, where there is a conflict that AFM cannot resolve automatically, the file is moved to `.ptrash` at cache or home. In cache the `.ptrash` gets cleaned up when eviction is triggered. At home, it is not cleared automatically. When the administrator is looking to clear some space, the `.ptrash` must be cleaned up first.
How to restore an unmounted AFM fileset that uses GPFS™ protocol as backend?	If the NSD mount on the gateway node is unresponsive, AFM does not synchronize data with home. The filesystem might be unmounted at the gateway node. A message `AFM: Remote filesystem` *remotefs* `is panicked due to unresponsive messages on fileset <`*fileset_name*`>,re-mount the filesystem after it becomes responsive. mmcommon preunmount invoked. File system: fs1 Reason: SGPanic` is written to `mmfs.log`. After the home is responsive, you must restore the NSD mount on the gateway node.

Chapter 26. AFM DR issues

This topic lists the answers to the common AFM DR questions.

Table 61. Common questions in AFM DR with their resolution

Issue	Resolution
How do I flush requeued messages?	Sometimes, requests in the AFM messages queue on the gateway node get requeued due to errors at the home cluster. For example, if space is not available at the home cluster to perform a new write, a write message that is queued is not successful and gets requeued. The administrator views the failed message being requeued on the MDS. Add more space to the home cluster and run **mmafmctl resumeRequeued** so that the requeued messages are executed at home again. If **mmafmctl resumeRequeued** is not run by an administrator, AFM executes the message in the regular order of message executions from the cache cluster to the home cluster. Running **mmfsadm saferdump afm all** on the gateway node displays the queued messages. The requeued messages are displayed in the dumps. An example: `c12c4apv13.gpfs.net: Normal Queue: (listed by execution order) (state: Active)c12c4apv13.gpfs.net: Write [612457.552962] requeued file3 (43 @ 293) chunks 0 bytes 0 0`
Why is a fileset in the Unmounted or Disconnected state when parallel I/O is set up?	Filesets that are using a mapping target go to the Disconnected mode if the NFS server of the MDS is unreachable, even if NFS servers of all participating gateways are reachable. The NFS server of the MDS must be checked to fix this problem.
How to clean unmount of the secondary filesystem fails if there are caches using GPFS protocol as backend?	To have a clean unmount of secondary filesystem, the filesystem should first be unmounted on the primary cluster where it has been remotely mounted and then the secondary filesystem should be unmounted. It might not be possible to unmount the remote file system from all nodes in the cluster until the relevant primary is unlinked or the local file system is unmounted. Force unmount/shutdown/crash of remote cluster results panic of the remote filesystem at primary cluster and queue gets dropped, next access to fileset runs recovery. However this should not affect primary cluster.

Table 61. Common questions in AFM DR with their resolution (continued)

Issue	Resolution
'DF' command hangs on the primary cluster	On RHEL 7.0 or later, **df** does not support hidden NFS mounts. As AFM uses regular NFS mounts on the gateway nodes, this change causes commands like **df** to hang if the secondary gets disconnected. The following workaround can be used that allows NFS mounts to continue to be hidden: Remove `/etc/mtab` symlink, and create new file `/etc/mtab` and copy `/proc/mounts` to `/etc/mtab` file during startup. In this solution, mtab file might go out of sync with `/proc/mounts`
What does `NeedsResync` state imply ?	`NeedsResync` state does not necessarily mean a problem. If this state is during a conversion or recovery, the problem gets automatically fixed in the subsequent recovery. You can monitor the **mmafmctl $fsname getstate** to check if its queue number is changing. And also can check the `gpfs` logs and for any errors, such as unmounted.
Is there a single command to delete all RPO snapshots from a primary fileset?	No. All RPOs need to be manually deleted.
Suppose there are more than two RPO snapshots on the primary. Where did these snapshots come from?	Check the queue. Check if recovery happened in the recent past. The extra snapshots will get deleted during subsequent RPO cycles.
How to restore an unmounted AFM DR fileset that uses GPFS™ protocol as backend?	If the NSD mount on the gateway node is unresponsive, AFM DR does not synchronize data with secondary. The filesystem might be unmounted at the gateway node. A message `AFM: Remote filesystem` *remotefs* `is panicked due to unresponsive messages on fileset <`*fileset_name*`>,re-mount the filesystem after it becomes responsive. mmcommon preunmount invoked. File system: fs1 Reason: SGPanic` is written to `mmfs.log`. After the secondary is responsive, you must restore the NSD mount on the gateway node.

Chapter 27. Transparent cloud tiering issues

This topic describes the common issues (along with workarounds) that you might encounter while using Transparent cloud tiering.

Migration/Recall failures

If a migration or recall fails, simply retry the policy or CLI command that failed two times after clearing the condition causing the failure. This works because the Transparent cloud tiering service is idempotent.

mmcloudgateway: Internal Cloud services returned an error: MCSTG00098I: Unable to reconcile /ibm/fs1 - probably not an space managed file system.

This typically happens if administrator has tried the **mmcloudgateway account delete** command before and has not restarted the service prior to invoking the migrate, reconcile, or any other similar commands. If the migration, reconcile, or any other Cloud services command fails with such a message, restart the Cloud services once by using the **mmcloudgateway service restart {-N node-class}** and retry the command.

Starting or stopping Transparent cloud tiering service fails with `the Transparent cloud tiering seems to be in startup phase` message

This is typically caused if the Gateway service is killed manually by using the **kill** command, without the graceful shutdown by using the **mmcloudgateway service stop** command.

Adding a cloud account to configure IBM Cloud Object Storage fails with this error, `56: Cloud Account Validation failed. Invalid credential for Cloud Storage Provider. Details: Endpoint URL Validation Failed, invalid username or password.`

Ensure that the appropriate user role is set through IBM Cloud Object StoragedsNet Manager GUI.

`HTTP Error 401 Unauthorized` exception while you configure a cloud account

This issue happens when the time between the object storage server and the Gateway node is not synced up.

Sync up the time with an NTP server and retry the operation.

Account creation command fails after a long wait and IBM Cloud Object Storage displays an error message saying that the vault cannot be created; but the vault is created

When you look at the IBM Cloud Object Storage manager UI, you see that the vault exists. This problem can occur if Transparent cloud tiering does not receive a successful return code from IBM Cloud Object Storage for the vault creation request.

The most common reason for this problem is that the threshold setting on the vault template is incorrect. If you have 6 IBM Cloud Object Storage slicestors and the write threshold is 6, then IBM Cloud Object Storage expects that all the slicestors are healthy. Check the IBM Cloud Object Storage manager UI. If any slicestors are in a warning or error state, update the threshold of the vault template.

Account creation command fails with error MCSTG00065E, but the data vault and the metadata vault exist

The full error message for this error is as follows:

```
MCSTG00065E: Command Failed with following reason: Error checking existence of, or creating,
cloud container container_name or cloud metadata container container_name.meta.
```

But the data vault and the metadata vault are visible on the IBM Cloud Object Storage UI.

This error can occur if the metadata vault was created but its name index is disabled. To resolve this problem, do one of the follow actions:

- Enter the command again with a new vault name and vault template.
- Delete the vault on the IBM Cloud Object Storage UI and run the command again with the correct *--metadata-location*.

Note: It is a good practice to disable the name index of the data vault. The name index of the metadata vault must be enabled.

File or metadata transfer fails with `koffLimitedRetryHandler:logError - Cannot retry after server error, command has exceeded retry limit`, followed by `RejectingHandler:exceptionCaught - Caught an exception com.ibm.gpfsconnector.messages.GpfsConnectorException: Unable to migrate`

This is most likely caused by a network connectivity and/or bandwidth issue.

Make sure that the network is functioning properly and retry the operation.

For policy-initiated migrations, IBM Spectrum Scale policy scan might automatically retry the migration of the affected files on a subsequent run.

gpfs.snap: An Error was detected on node XYZ while invoking a request to collect the snap file for Transparent cloud tiering: (return code: 137).

If the **gpfs.snap** command fails with this error, increase the value of the *timeout* parameter by using the **gpfs.snap --timeout Seconds** option.

Note: If the Transparent cloud tiering log collection fails after the default timeout period expires, you can increase the timeout value and collect the TCT logs. The default timeout is 300 seconds (or 5 minutes).

Migration fails with error: MCSTG00008E: Unable to get fcntl lock on inode. Another MCStore request is running against this inode.

This happens because some other application might be having the file open, while Cloud services are trying to migrate it.

Connect: No route to host Cannot connect to the Transparent Cloud Tiering service. Please check that the service is running and that it is reachable over the network. Could not establish a connection to the MCStore server

During any data command, if this error is observed, it is due to abrupt shutdown of Cloud services on one of the nodes. This happens when Cloud services is not stopped on a node explicitly using the **mmcloudgateway service stop** command, but power of a node goes down or IBM Spectrum Scale daemon is taken down. This causes node IP address to be still considered as an active Cloud services node and, the data commands routed to it fail with this error.

"Generic_error" in the `mmcloudgateway service status` output

This error indicates that the cloud object storage is unreachable. Ensure that there is outbound connectivity to the cloud object storage. Logs indicate an exception about the failure.

An unexpected exception occurred during directory processing : Input/output error

You might encounter this error while migrating files to the cloud storage tier. To fix this, check the status of NSDs and ensure that the database/journal files are not corrupted and can be read from the file system.

It is marked for use by Transparent Cloud Tiering

You might encounter this error when you try to remove a Cloud services node from a cluster. To resolve this, use the --force option with the **mmchnode** command as follows:

```
mmchnode --cloud-gateway-disable -N nodename --cloud-gateway-nodeclass nodeclass --force
```

Chapter 28. File audit logging issues

The following topics discuss issues that you might encounter in file audit logging.

Failure of mmaudit because of the file system level

The following problem can occur if you upgrade **spectrumscale** without completing the permanent cluster upgrade.

```
# mmaudit TestDevice enable
[E] File system device TestDevice is not at a supported file system level for audit logging.
    The File Audit Logging command associated with device: TestDevice cannot be completed.
    Choose a device that is at the minimum supported file system level
    and is accessible by all nodes associated with File Audit Logging and try the command again.
mmaudit: Command failed. Examine previous error messages to determine cause.
```

See *Completing the upgrade to a new level of IBM Spectrum Scale* in the *IBM Spectrum Scale: Concepts, Planning, and Installation Guide* to get the cluster and file system to enable new functionality.

JSON reporting issues in file audit logging

This topic describes limitations that the user might observe in the file audit logging JSON logs for the **spectrumscale** protocols.

SMB

- For a created file, there can be many **OPEN** and **CLOSE** events.
- For a deleted file, a **DESTROY** event might never get logged. Instead, the JSON might show an **UNLINK** event only.

NFS

- For NFSv3, upon file creation, audit logs only get an **OPEN** event without a corresponding **CLOSE** event. The **CLOSE** event only occurs when a subsequent operation is done on the file (for example, **RENAME**).
- The file path name, NFS IP for kNFS might not always be available.

Object

Object is not supported for file audit logging in IBM Spectrum Scale 5.0.0 or 5.0.1.

Note: In addition, file activity in the primary object fileset does not generate events.

Unified file

Unified file is not supported for file audit logging in IBM Spectrum Scale 5.0.0 or 5.0.1.

Miscellaneous limitations

The file path for the **DESTROY** event might be NULL at times. A **DESTROY** event is preceded by an **UNLINK** event. The file path will be reported for the **UNLINK** event. However, for the **DESTROY** event, the file path will only be reported if there are other links pointing to the file.

Events are not being audited after disabling and re-enabling the message queue

This topic describes what to do if events are no longer being written to the file audit logging fileset after you disabled and re-enabled the message queue.

Symptoms:

- You have recently disabled the message queue and re-enabled it.

- New events are not being written to the file audit logging fileset.
- You see warning messages on multiple nodes from the `/var/adm/ras/mmaudit.log`, which state that the Kafka producer is **DEGRADED** as well as an increasing number of messages that state that they cannot be logged.

Resolution:

Perform a rolling restart of the GPFS daemon on every node by logging into each node and running the **mmshutdown** command followed by the **mmstartup** command. Perform this on one node at a time. Let it complete before progressing to the next node.

Chapter 29. Maintenance procedures

The directed maintenance procedures (DMPs) assist you to fix certain issues reported in the system.

Directed maintenance procedures available in the GUI

The directed maintenance procedures (DMPs) assist you to repair a problem when you select the action **Run fix procedure** on a selected event from the **Monitoring** > **Events** page. DMPs are present for only a few events reported in the system.

The following table provides details of the available DMPs and the corresponding events.

Table 62. DMPs

DMP	Event ID
Start NSD	disk_down
Start GPFS daemon	gpfs_down
Increase fileset space	inode_error_high and inode_warn_high
Synchronize Node Clocks	time_not_in_sync
Start performance monitoring collector service	pmcollector_down
Start performance monitoring sensor service	pmsensors_down
Activate AFM performance monitoring sensors	afm_sensors_inactive
Activate NFS performance monitoring sensors	nfs_sensors_inactive
Activate SMB performance monitoring sensors	smb_sensors_inactive

Start NSD

The Start NSD DMP assists to start NSDs that are not working.

The following are the corresponding event details and the proposed solution:

- **Event ID:** disk_down
- **Problem:** The availability of an NSD is changed to “down”.
- **Solution:** Recover the NSD

The DMP provides the option to start the NSDs that are not functioning. If multiple NSDs are down, you can select whether to recover only one NSD or all of them.

The system issues the **mmchdisk** command to recover NSDs as given in the following format:

```
/usr/lpp/mmfs/bin/mmchdisk <device> start -d <disk description>
```

For example: `/usr/lpp/mmfs/bin/mmchdisk r1_FS start -d G1_r1_FS_data_0`

Start GPFS daemon

When the GPFS daemon is down, GPFS functions do not work properly on the node.

The following are the corresponding event details and the proposed solution:

- **Event ID:** gpfs_down
- **Problem:** The GPFS daemon is down. GPFS is not operational on node.

- **Solution:** Start GPFS daemon.

The system issues the **mmstartup -N** command to restart GPFS daemon as given in the following format:
`/usr/lpp/mmfs/bin/mmstartup -N <Node>`

For example: `usr/lpp/mmfs/bin/mmstartup -N gss-05.localnet.com`

Increase fileset space

The system needs inodes to allow I/O on a fileset. If the inodes allocated to the fileset are exhausted, you need to either increase the number of maximum inodes or delete the existing data to free up space.

The procedure helps to increase the maximum number of inodes by a percentage of the already allocated inodes. The following are the corresponding event details and the proposed solution:

- **Event ID:** inode_error_high and inode_warn_high
- **Problem:** The inode usage in the fileset reached an exhausted level
- **Solution:** increase the maximum number of inodes

The system issues the **mmchfileset** command to recover NSDs as given in the following format:
`/usr/lpp/mmfs/bin/mmchfileset <Device> <Fileset> --inode-limit <inodesMaxNumber>`

For example: `/usr/lpp/mmfs/bin/mmchfileset r1_FS testFileset --inode-limit 2048`

Synchronize node clocks

The time must be in sync with the time set on the GUI node. If the time is not in sync, the data that is displayed in the GUI might be wrong or it does not even display the details. For example, the GUI does not display the performance data if time is not in sync.

The procedure assists to fix timing issue on a single node or on all nodes that are out of sync. The following are the corresponding event details and the proposed solution:

- **Event ID:** time_not_in_sync
- **Limitation:** This DMP is not available in sudo wrapper clusters. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter `GPFS_USER=<user name>`, which is available in the file `/usr/lpp/mmfs/gui/conf/gpfsgui.properties`.
- **Problem:** The time on the node is not synchronous with the time on the GUI node. It differs more than 1 minute.
- **Solution:** Synchronize the time with the time on the GUI node.

The system issues the **sync_node_time** command as given in the following format to synchronize the time in the nodes:
`/usr/lpp/mmfs/gui/bin/sync_node_time <nodeName>`

For example: `/usr/lpp/mmfs/gui/bin/sync_node_time c55f06n04.gpfs.net`

Start performance monitoring collector service

The collector services on the GUI node must be functioning properly to display the performance data in the IBM Spectrum Scale management GUI.

The following are the corresponding event details and the proposed solution:

- **Event ID:** pmcollector_down
- **Limitation:** This DMP is not available in sudo wrapper clusters when a remote *pmcollector* service is used by the GUI. A remote *pmcollector* service is detected in case a different value than localhost is specified in the ZIMonAddress in file, which is located at: `/usr/lpp/mmfs/gui/conf/`

gpfsgui.properties. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter GPFS_USER=<user name>, which is available in the file /usr/lpp/mmfs/gui/conf/gpfsgui.properties.

- **Problem:** The performance monitoring collector service *pmcollector* is in inactive state.
- **Solution:** Issue the **systemctl status pmcollector** to check the status of the collector. If *pmcollector* service is inactive, issue **systemctl start pmcollector**.

The system restarts the performance monitoring services by issuing the **systemctl restart pmcollector** command.

The performance monitoring collector service might be on some other node of the current cluster. In this case, the DMP first connects to that node, then restarts the performance monitoring collector service.

```
ssh <nodeAddress> systemctl restart pmcollector
```

For example: `ssh 10.0.100.21 systemctl restart pmcollector`

In a sudo wrapper cluster, when collector on remote node is down, the DMP does not restart the collector services by itself. You need to do it manually.

Start performance monitoring sensor service

You need to start the sensor service to get the performance details in the collectors. If sensors and collectors are not started, the GUI and CLI do not display the performance data in the IBM Spectrum Scale management GUI.

The following are the corresponding event details and the proposed solution:

- **Event ID:** pmsensors_down
- **Limitation:** This DMP is not available in sudo wrapper clusters. In a sudo wrapper cluster, the user name is different from 'root'. The system detects the user name by finding the parameter GPFS_USER=<user name>, which is available in the file /usr/lpp/mmfs/gui/conf/gpfsgui.properties.
- **Problem:** The performance monitoring sensor service *pmsensor* is not sending any data. The service might be down or the difference between the time of the node and the node hosting the performance monitoring collector service *pmcollector* is more than 15 minutes.
- **Solution:** Issue **systemctl status pmsensors** to verify the status of the sensor service. If *pmsensor* service is inactive, issue **systemctl start pmsensors**.

The system restarts the sensors by issuing **systemctl restart pmsensors** command.

For example: `ssh gss-15.localnet.com systemctl restart pmsensors`

Activate AFM performance monitoring sensors

The activate SMB performance monitoring sensors DMP assists to activate the inactive SMB sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** afm_sensors_inactive
- **Problem:** The AFM performance cannot be monitored because one or more of the performance sensors GPFSAFMFS, GPFSAFMFSET, and GPFSAFM are offline.
- **Solution:** Activate the AFM sensors.

The DMP provides the option to activate the AFM monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate AFM sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update <<sensor_name>>.restrict=<<afm_gateway_nodes>>
/usr/lpp/mmfs/bin/mmperfmon config update <<sensor_name>>.period=<<seconds>>
```

For example:

```
/usr/lpp/mmfs/bin/mmperfmon config update GPFSAFM.restrict=gss-41
/usr/lpp/mmfs/bin/mmperfmon config update GPFSAFM.period=30
```

Activate NFS performance monitoring sensors

The activate NFS performance monitoring sensors DMP assists to activate the inactive NFS sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** nfs_sensors_inactive
- **Problem:** The NFS performance cannot be monitored because the performance monitoring sensor NFSIO is inactive.
- **Solution:** Activate the SMB sensors.

The DMP provides the option to activate the NFS monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate the sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update NFSIO.restrict=cesNodes NFSIO.period=<<seconds>>
```

For example: `/usr/lpp/mmfs/bin/mmperfmon config update NFSIO.restrict=cesNodes NFSIO.period=10`

Activate SMB performance monitoring sensors

The activate SMB performance monitoring sensors DMP assists to activate the inactive SMB sensors.

The following are the corresponding event details and the proposed solution:

- **Event ID:** smb_sensors_inactive
- **Problem:** The SMB performance cannot be monitored because either one or both the SMBStats and SMBGlobalStats sensors are inactive.
- **Solution:** Activate the SMB sensors.

The DMP provides the option to activate the SMB monitoring sensor and select a data collection interval that defines how frequently the sensors must collect data. It is recommended to select a value that is greater than or equal to 10 as the data collection frequency to reduce the impact on the system performance.

The system issues the **mmperfmon** command to activate the sensors as given in the following format:

```
/usr/lpp/mmfs/bin/mmperfmon config update SMBStats.restrict=cesNodes SMBStats.period=<<seconds>>
```

For example: `/usr/lpp/mmfs/bin/mmperfmon config update SMBStats.restrict=cesNodes NFSIO.period=10`

Directed maintenance procedures for tip events

The directed maintenance procedures (DMPs) assist you to repair a problem when you select the action **Run fix procedure** on a selected event from the **GUI** > **Monitoring** > **Events page**. DMPs are present for the following tip events reported in the system.

Attention:

If you run these DMPs manually on the command line, the tip event will not reset immediately.

Table 63. Tip events list

Reporting component	Event Name	Prerequisites	Conditions	Fix Procedure

gpfs	gpfs_pagepool_ small
	
	gpfs_pagepool_ ok
			The actively used GPFS pagepool setting (`mmdiag --config | grep pagepool`) is lower than or equal to 1 GB.	• To change the value and make it effective immediately, use the following command::
				`mmchconfig pagepool=<value> -i`
				where **<value>** is a value higher than 1GB.
				• To change the value and make it effective after next GPFS recycle, use the following command::
				`mmchconfig pagepool=<value>`
				where **<value>** is a value higher than 1GB.
				• To ignore the event, use the following command:
				`mmhealth event hide gpfs_pagepool_small`

AFM component	afm_sensors_ inactive
	
	afm_sensors_ active
		Verify that the node has a gateway designation and a perfmon designation using the **mmlscluster** command.	The period for at least one of the following AFM sensors' is set to 0: GPFSAFM, GPFSAFMFS, GPFSAFMFSET.	• To change the period when the sensors are defined in the perfmon configuration file, use the following command:
				`mmperfmon config update <sensor_name>.period=<interval>`
				Where **<sensor_name>** is one of the AFM sensors *GPFSAFM*, *GPFSAFMFS*, or *GPFSAFMFSET*, and **<interval>** is the time in seconds that the sensor waits to gather the different sensors' metrics again.
				• To change the period when the sensors are not defined in the perfmon configuration file, create a sensorsFile with input using the following command:
				`sensors = {`
				`name = <sensor_name>`
				`period = <interval>`
				`type = "Generic"`
				`}`
				`mmperfmon config add --sensors <path_to_tmp_cfg_file>`
				• To ignore the event, use the following command:
				`mmhealth event hide afm_sensors_inactive`

NFS component	nfs_sensors_ inactive
	
	nfs_sensors_ active
		Verify that the node is NFS enabled, and has a perfmon designation using the **mmlscluster** command.	The NFS sensor NFSIO has a period of 0.	• To change the period when the sensors are defined in the perfmon configuration file, use the following command:
				`mmperfmon config update <sensor_name>.period=<interval>`
				Where **<sensor_name>** is the NFS sensor *NFSIO*, and **<interval>** is the time in seconds that the sensor waits to gather the different sensors' metrics again.
				• To change the period when the sensors are not defined in the perfmon configuration file, use the following command:
				`mmperfmon config add --sensors /opt/IBM/zimon/defaults/GaneshaProxy.conf`
				• To ignore the event, use the following command:
				`mmhealth event hide nfs_sensors_inactive`

Table 63. Tip events list (continued)

Reporting component	Event Name	Prerequisites	Conditions	Fix Procedure

SMB component	smb_sensors_ inactive
	
	smb_sensors_ active
		Verify that the node is SMB enabled, and has a perfmon designation using the **mmlscluster** command.	The period of at least one of the following SMB sensors' is set to 0: SMBStats, SMBGlobalStats .	• To change the period when the sensors are defined in the perfmon configuration file, use the following command:
				`mmperfmon config update <sensor_name>.period=<interval>`
				Where **<sensor_name>** is one of the SMB sensors *SMBStats* or *SMBGlobalStats*, and **<interval>** is the time in seconds that the sensor waits to gather the different sensors' metrics again.
				• To change the period when the sensors are not defined in the perfmon configuration file, use the following command:
				`mmperfmon config add --sensors /opt/IBM/zimon/defaults/ZIMonSensors_smb.cfg`
				• To ignore the event, use the following command:
				`mmhealth event hide smb_sensors_inactive`

gpfs	gpfs_maxfilestocache_ small
	
	gpfs_maxfilestocache _ok
		Verify that the node is in the cesNodes node class using the
		`mmlsnodeclass --all`
		
		command.
			The actively used GPFS maxFilesToCache (`mmdiag --config | grep maxFilesToCache`) setting has a value smaller than or equal to 100,000.	• To change the value, use the following command:
				`mmchconfig maxFilesToCache=<value>; mmshutdown; mmstartup`
				where **<value>** is a value higher than 100,000
				• To ignore the event, use the following command:
				`mmhealth event hide gpfs_maxfilestocache_small`

gpfs	gpfs_maxstatcache _high
	
	gpfs_maxstatcache _ok
		Verify that the node is a Linux node.	The actively used GPFS maxStatCache (`mmdiag --config | grep maxStatCache`) value is higher than 0.	• To change the value, use the following command:
				`mmchconfig maxStatCache=0; mmshutdown; mmstartup`
				• To ignore the event, use the following command:
				`mmhealth event hide gpfs_maxstatcache_high`

gpfs	callhome_not_enabled callhome_enabled	Verify that the node is the Cluster Manager using the **mmlsmgr -c** command.	Callhome is not enabled on the cluster.	• To install the callhome packages: 1. Install the **gpfs.callhome-ecc-client-{version-number}.noarch.rpm** package for the ecc client. 2. Install the **gpfs.callhome-{version}.{type}.noarch.rpm** package for the callhome code. first install the ecc client, then the other one. • To configure the callhome package that are installed but not configured: 1. Issue the **mmcallhome capability enable** command to initialize the configuration. 2. Issue the **mmcallhome info change** command to add personal information. 3. Issue the **mmcallhome pro** command to include a proxy if needed. 4. Issue the **mmcallhome group add** or **mmcallhome group auto** command to create callhome groups . • To enable callhome once the callhome package is installed and the groups are configured, issue the **mmcallhome capability enable** command.

For information on tip events, see "Event type and monitoring status for system health" on page 112.

Note: Since the TIP state is only checked once every hour, it might take up to an hour for the change to reflect in the output of the **mmhealth** command.

Chapter 30. Recovery procedures

You need to perform certain procedures to recover the system to minimize the impact of the issue reported in the system and to bring the system back to the normal operating state. The procedure re-creates the system by using saved configuration data or by restarting the affected services.

Restoring data and system configuration

You can back up and restore the configuration data for the system after preliminary recovery tasks are completed.

You can maintain your configuration data for the system by completing the following tasks:

1. Backing up the configuration data
2. Restoring the configuration data
3. Deleting unwanted backup configuration data files

The following topics describes how to perform backup and restore data and configuration in the IBM Spectrum Scale system:

- *Protocols cluster disaster recovery* in *IBM Spectrum Scale: Administration Guide*
- *Restore procedure with SOBAR* in *IBM Spectrum Scale: Administration Guide*
- *Encryption and backup/restore* in *IBM Spectrum Scale: Administration Guide*
- *Backup and restore with storage pools* in *IBM Spectrum Scale: Administration Guide*
- *Restoring quota files* in *IBM Spectrum Scale: Administration Guide*
- *Backing up and restoring protocols and CES configuration information* in *IBM Spectrum Scale: Administration Guide*
- *Failback or restore steps for object configuration* in *IBM Spectrum Scale: Administration Guide*

Automatic recovery

The IBM Spectrum Scale recovers itself from certain issues without manual intervention.

The following automatic recovery options are available in the system:

- Failover of CES IP addresses to recover from node failures. That is, if any important service or protocol service is broken on a node, the system changes the status of that node as *Failed* and moves the public IPs to healthy nodes in the cluster.

  These failovers get triggered due to following conditions:

  1. If the spectrum scale monitoring service detects a critical problem in any of the CES components such as NFS,SMB, or OBJ, then the CES state is set to FAILED and this triggers a failover.
  2. If the IBM Spectrum Scale daemon detects a problem with the node or cluster such as expel node, or quorumloss, then it executes callbacks and a failover is triggered.
  3. The CES framework also triggers a failover during the distribution of IP addresses as per the distribution policy.
- In case of any errors with the SMB and Object protocol services, the system restarts the corresponding daemons. If restarting the protocol service daemons does not resolve the issue and the maximum retry count is reached, the system changes the status of the node as *Failed*. The protocol service restarts are logged in the event log. Issue either **mmhealth node eventlog** commands to view the details of such events.

It the system detects multiple problems at once, then it starts the recovery procedure such as automatic restart, and addresses the issue of the highest priority event first. Once the recovery actions are completed for the highest priority event, the system health is monitored again and then the recovery actions for the next priority event is started. Similarly, issues with each event are handled based on their priority state until all failure events have been resolved or the retry count is reached. For example, if the system has two failure events as smb_down and ctdb_down, then since the ctdb_down event has a higher priority, so the ctdb service is restarted first. Once the recovery actions for ctdb_down event is completed, the system health is monitored again. If the ctdb_down issue is resolved, then the recovery actions for the smb_down event is started.

Upgrade recovery

Use this information to recover from a failed upgrade.

A failed upgrade might leave a cluster in a state of containing multiple code levels. It is important to analyze console output to determine which nodes or components were upgraded prior to the failure and which node or component was in the process of being upgraded when the failure occurred.

Once the problem has been isolated, a healthy cluster state must be achieved prior to continuing the upgrade. Use the **mmhealth** command in addition to the **mmces state show -a** command to verify that all services are up. It might be necessary to manually start services that were down when the upgrade failed. Starting the services manually helps achieve a state in which all components are healthy prior to continuing the upgrade.

For more information about verifying service status, see **mmhealth command** and **mmces state show command** in *IBM Spectrum Scale: Command and Programming Reference*.

Recovery procedure for a broken cluster when no CCR backup is available

The following procedure describes how to recover from a broken cluster state, when no intact CCR can be found on the quorum nodes, and no CCR backup is available from which the broken cluster can be recovered.

This procedure does not guarantee to recover the most recent state of all the configuration files in the CCR. Instead, it brings the CCR back into a consistent state with the most recent available version of each configuration file.

The procedure consists of the following major steps:

1. Diagnose if the CCR is broken on a majority of the quorum nodes.
2. Evaluate the CCR's most recent Paxos state file.
3. Patch the CCR's Paxos state file.
4. Verify that the CCR state is intact after patching the Paxos state file and copying back to the CCR directory.

The procedure can be done for a single node cluster as well as a multi-node cluster.

Recovery procedure for a broken single node cluster

The CCR becomes unavailable on a quorum node due to the presence of corrupted files in the CCR committed directory. The presence of corrupted files in the CCR committed directory might be the result of a node crash or a hard power-off.

The following code block provides information regarding the cluster used for this procedure:

```
GPFS cluster information
========================
  GPFS cluster name:         gpfs-cluster-1.localnet.com
  GPFS cluster id:           317908494312539041
  GPFS UID domain:           localnet.com
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           CCR

GPFS cluster configuration servers:
-----------------------------------
  Primary server:    node-11.localnet.com (not in use)
  Secondary server:  (none)

 Node  Daemon node name      IP address   Admin node name       Designation
--------------------------------------------------------------------------
   1   node-11.localnet.com  10.0.100.11  node-11.localnet.com  quorum
```

Since almost all mm-commands use the CCR, the mm-commands start failing when the CCR becomes unavailable. Follow these steps to recover the broken single node cluster:

1. Run the **mmgetstate** command to display the status of the nodes.

   ```
   [root@node-11 ~]# mmgetstate -a
   get file failed: Maximum number of retries reached (err 801)
   gpfsClusterInit: Unexpected error from ccr fget mmsdrfs.  Return code: 149
   mmgetstate: Command failed. Examine previous error messages to determine cause.
   ```

2. Run the **mmccr check** command to get the CCR state of the quorum node.

   ```
   [root@node-11 ~]# mmccr check -Y -e;echo $?
   mmccr::HEADER:version:reserved:reserved:NodeId:CheckMnemonic:
   ErrorCode:ErrorMsg:ListOfFailedEntities:ListOfSucceedEntities:Severity:
   mmccr::0:1:::1:CCR_CLIENT_INIT:0:::/var/mmfs/ccr,
   /var/mmfs/ccr/committed,/var/mmfs/ccr/ccr.nodes,Security:OK:
   mmccr::0:1:::1:FC_CCR_AUTH_KEYS:0:::/var/mmfs/ssl/authorized_ccr_keys:OK:
   mmccr::0:1:::1:FC_CCR_PAXOS_CACHED:0:::/var/mmfs/ccr/cached,
   /var/mmfs/ccr/cached/ccr.paxos:OK:
   mmccr::0:1:::1:FC_CCR_PAXOS_12:0:::/var/mmfs/ccr/ccr.paxos.1,
   /var/mmfs/ccr/ccr.paxos.2:OK:
   mmccr::0:1:::1:PC_LOCAL_SERVER:0:::node-11.localnet.com:OK:
   mmccr::0:1:::1:PC_IP_ADDR_LOOKUP:0:::node-11.localnet.com,0.000:OK:
   mmccr::0:1:::1:PC_QUORUM_NODES:0:::10.0.100.11:OK:
   mmccr::0:1:::1:FC_COMMITTED_DIR:5:Files in committed directory
    missing or corrupted:1:7:WARNING:
   mmccr::0:1:::1:TC_TIEBREAKER_DISKS:0::::OK:
   149
   ```

   Note: The **mmccr check** command works on the local node, except the `mmccr check mnemonic PC_QUORUM_NODES` option, which is used to ping other quorum nodes. The **mmccr check** command has no dependencies to other mm-commands.

   The **mmccr check** command lists the corrupted or missing files in the committed directory, `/var/mmfs/ccr/committed`, and returns a nonzero exit code. A list of files that are committed to the CCR can be seen along with their metadata information when reading one of the CCR Paxos state files. The metadata includes information like file ID, file version, update ID, and Cyclic Redundancy Check (CRC):

   ```
   [root@node-11 ~]# mmccr readpaxos /var/mmfs/ccr/ccr.paxos.1 | awk '/files/,0'
   files: 8, max deleted version 0
     1 = version 1 uid ((n1,e0),0) crc 6E0689F3
     2 = version 1 uid ((n1,e0),1) crc FFFFFFFF
     3 = version 1 uid ((n1,e0),2) crc FFFFFFFF
     4 = version 1 uid ((n1,e0),3) crc 4CC9CEC4
     5 = version 1 uid ((n1,e0),4) crc CEE097C7
     6 = version 2 uid ((n1,e1),15) crc 0CDFC737
     7 = version 11 uid ((n1,e2),39) crc 35026185
     8 = version 4 uid ((n1,e2),42) crc 90406D69
   ```

CRC is used to verify the proper file transfer from one node to the other through the network. However, you need to know the current CRC of the files to identify a corrupted one. The cksum tool can be used to calculate the CRC independent from the CCR. The file name of the files in the committed directory has one of these patterns:

- <FILE_NAME>.<FILE_ID>.<FILE_VERSION>.<FILE_CRC>.<UPDATE_ID>
- <FILE_NAME>.<FILE_ID>.<FILE_VERSION>.<FILE_CRC>.<UPDATE_ID>.bad.<PROCESS_ID>.<THREAD_ID>.<TIME_STAMP>

If the CRC returned by the cksum tool does not match the FILE_CRC value for a file, then that file has become corrupted. For example, consider the CRC values of the following files:

```
[root@node-11 ~]# cksum /var/mmfs/ccr/committed/* | awk
 '{ printf "%x %s\n", $1, $3 }'
ffffffff /var/mmfs/ccr/committed/ccr.disks.2.1.ffffffff.010001
6e0689f3 /var/mmfs/ccr/committed/ccr.nodes.1.1.6e0689f3.010000
90406d69 /var/mmfs/ccr/committed/clusterEvents.8.3.90406d69.010229
ffffffff /var/mmfs/ccr/committed/
clusterEvents.8.4.90406d69.01022a.bad.21345.2476373824.2018-02-26_15:11:13.731+0100
4cc9cec4 /var/mmfs/ccr/committed/genKeyData.4.1.4cc9cec4.010003
4cc9cec4 /var/mmfs/ccr/committed/genKeyDataNew.6.1.4cc9cec4.010005
cdfc737 /var/mmfs/ccr/committed/genKeyDataNew.6.2.cdfc737.01010f
ffffffff /var/mmfs/ccr/committed/mmLockFileDB.3.1.ffffffff.010002
dcc79030 /var/mmfs/ccr/committed/mmsdrfs.7.10.dcc79030.010223
35026185 /var/mmfs/ccr/committed/mmsdrfs.7.11.35026185.010227
cee097c7 /var/mmfs/ccr/committed/mmsysmon.json.5.1.cee097c7.010004
```

In the output, the file `clusterEvents.8.4.90406d69.01022a.bad.21345.2476373824.2018-02-26_15:11:13.731+0100` is the only file with a CRC mismatch in the CCR's committed directory. For a cluster with multiple quorum nodes, this corrupted file is recovered by copying an intact version from another quorum node during GPFS startup. But it is not possible for a single cluster node, because there is no other intact copy of this file on another quorum node.

3. Back up the entire CCR folder on the quorum node for further analysis by archiving the entire CCR directory using the tar-command.

```
[root@node-11 ~]# tar -cvf
CCR_archive_node-11_20180226151240.tar /var/mmfs/ccr
tar: Removing leading `/' from member names
/var/mmfs/ccr/
/var/mmfs/ccr/ccr.noauth
/var/mmfs/ccr/ccr.paxos.1
/var/mmfs/ccr/committed/
/var/mmfs/ccr/committed/clusterEvents.8.3.90406d69.010229
/var/mmfs/ccr/committed/genKeyDataNew.6.1.4cc9cec4.010005
/var/mmfs/ccr/committed/genKeyData.4.1.4cc9cec4.010003
/var/mmfs/ccr/committed/mmLockFileDB.3.1.ffffffff.010002
/var/mmfs/ccr/committed/mmsysmon.json.5.1.cee097c7.010004
/var/mmfs/ccr/committed/ccr.disks.2.1.ffffffff.010001
/var/mmfs/ccr/committed/ccr.nodes.1.1.6e0689f3.010000
/var/mmfs/ccr/committed/mmsdrfs.7.10.dcc79030.010223
/var/mmfs/ccr/committed/genKeyDataNew.6.2.cdfc737.01010f
/var/mmfs/ccr/committed/
clusterEvents.8.4.90406d69.01022a.bad.21345.2476373824.2018-02-26_15:11:13.731+0100
/var/mmfs/ccr/committed/mmsdrfs.7.11.35026185.010227
/var/mmfs/ccr/ccr.disks
/var/mmfs/ccr/cached/
/var/mmfs/ccr/cached/ccr.paxos
/var/mmfs/ccr/ccr.nodes
/var/mmfs/ccr/ccr.paxos.2

[root@node-11 ~]# ls -al CCR_archive_node-11_20180226151240.tar
-rw-r--r-- 1 root root 51200 Feb 26 15:12 CCR_archive_node-11_20180226151240.tar
```

4. Run the **mmshutdown** command to shut down IBM Spectrum Scale to avoid any issues that can be caused by an active mmfsd daemon.

```
[root@node-11 ~]# mmshutdown
Mon Feb 26 15:14:03 CET 2018: mmshutdown: Starting force unmount of GPFS file systems
Mon Feb 26 15:14:08 CET 2018: mmshutdown: Shutting down GPFS daemons
Mon Feb 26 15:15:15 CET 2018: mmshutdown: Finished
```

5. Run the **mmcommon** command to stop the mmsdrserv daemon and the startup scripts.

```
[root@node-11 ~]# mmcommon killCcrMonitor
```

Note: You can check whether all the GPFS daemons and monitor scripts have stopped on the quorum node using the following command:

```
[root@node-11 ~]# ps -C mmfsd,mmccrmonitor,mmsdrserv
  PID TTY          TIME CMD
```

6. Run the **mmccr readpaxos** command to get the most recent CCR Paxos state file on the quorum node.

```
[root@node-11 ~]# mmccr readpaxos /var/mmfs/ccr/ccr.paxos.1 | grep seq
dblk: seq 43, mbal (0.0), bal (0.0), inp ((n0,e0),0):(none)
:-1:None, leaderChallengeVersion 0
[root@node-11 ~]# mmccr readpaxos /var/mmfs/ccr/ccr.paxos.2 | grep seq
dblk: seq 42, mbal (1.1), bal (1.1), inp ((n1,e2),42):fu:4:
['clusterEvents',8,87,90406d69], leaderChallengeVersion 0
```

Ensure that you use the path to the most recent Paxos state file while using the **mmccr readpaxos** command. The CCR has two Paxos state files in its /var/mmfs/ccr directory, ccr.paxos.1 and ccr.paxos.2. CCR writes alternately to these two files. Maintaining dual copies allows the CCR to always have a copy intact in case the write to one the file fails for some reason and makes this file corrupt. The most recent file is the file with the higher sequence number in it.

In the example above the CCR Paxos state file ccr.paxos.1 is the most recent one. The ccr.paxos.1 file has the sequence number 43, while ccr.paxos.2 has sequence number 42. In case of a multi nodes cluster, not all quorum nodes must have the same set of sequence numbers, based on how many updates the CCR has seen until the **readpaxos** command is invoked.

The ccr.paxos.1 acts as the input file to patch the exiting CCR state based on the most recent Paxos state file and the files in the committed directory. The patched output file is specified as /var/mmfs/ccr/myPatched_ccr.paxos.1.

7. Run the **mmccr patchpaxos** command.

The output of the **mmccr patchpaxos** command gives the following information:

- Initial Paxos state. It is the state before the **patchpaxos** command is run.
- Files in the committed directory.
- Statistics of the changes made by the **patchpaxos** command.
- Final Paxos state. It is the state after the **patchpaxos** command is run.

```
[root@node-11 ~]# mmccr patchpaxos /var/mmfs/ccr/ccr.paxos.1
 /var/mmfs/ccr/committed/ /var/mmfs/ccr/myPatched_ccr.paxos.1
Committed state found in /var/mmfs/ccr/ccr.paxos.1:
config: minNodes: 1 version 0
  nodes: [(N1,S0,V0,L1)]
  disks: []
leader: id 1 version 1
updates: horizon -1
  {(n1,e0): 5, (n1,e1): 33, (n1,e2): 42}
values: 1, max deleted version 9
  mmRunningCommand = version 3 ""
files: 8, max deleted version 0
  1 = version 1 uid ((n1,e0),0) crc 6E0689F3
  2 = version 1 uid ((n1,e0),1) crc FFFFFFFF
  3 = version 1 uid ((n1,e0),2) crc FFFFFFFF
  4 = version 1 uid ((n1,e0),3) crc D3ED62E3
  5 = version 1 uid ((n1,e0),4) crc CEE097C7
  6 = version 2 uid ((n1,e1),15) crc FC913FFC
  7 = version 11 uid ((n1,e2),38) crc DF4822D6
  8 = version 4 uid ((n1,e2),42) crc 90406D69

Comparing to content of '/var/mmfs/ccr/committed/':
```

```
  match file: name: 'ccr.nodes' suffix: '1.1.6e0689f3.010000' id: 1
version: 1 crc: 6e0689f3 uid: ((n1,e0),0) and file list entry: 1.1.6e0689f3.010000
  match file: name: 'ccr.disks' suffix: '2.1.ffffffff.010001' id: 2
version: 1 crc: ffffffff uid: ((n1,e0),1) and file list entry: 2.1.ffffffff.010001
  match file: name: 'mmLockFileDB' suffix: '3.1.ffffffff.010002' id: 3
version: 1 crc: ffffffff uid: ((n1,e0),2) and file list entry: 3.1.ffffffff.010002
  match file: name: 'genKeyData' suffix: '4.1.d3ed62e3.010003' id: 4
version: 1 crc: d3ed62e3 uid: ((n1,e0),3) and file list entry: 4.1.d3ed62e3.010003
  match file: name: 'mmsysmon.json' suffix: '5.1.cee097c7.010004' id: 5
version: 1 crc: cee097c7 uid: ((n1,e0),4) and file list entry: 5.1.cee097c7.010004
  match file: name: 'genKeyDataNew' suffix: '6.2.fc913ffc.01010f' id: 6
version: 2 crc: fc913ffc uid: ((n1,e1),15) and file list entry: 6.2.fc913ffc.01010f
  match file: name: 'mmsdrfs' suffix: '7.11.df4822d6.010226' id: 7
version: 11 crc: df4822d6 uid: ((n1,e2),38) and file list entry: 7.11.df4822d6.010226
  older:   name: 'clusterEvents' suffix: '8.3.90406d69.010229' id: 8
version: 3 crc: 90406d69 uid: ((n1,e2),41)
  *** reverting committed file list version 4 uid ((n1,e2),42)

Found 7 matching, 0 deleted, 0 added, 0 updated, 1 reverted, 0 reset

Verifying update history

Writing 1 changes to /var/mmfs/ccr/myPatched_ccr.paxos.1
config: minNodes: 1 version 0
  nodes: [(N1,S0,V0,L1)]
  disks: []
leader: id 1 version 1
updates: horizon -1
  {(n1,e0): 5, (n1,e1): 33, (n1,e2): 42}
values: 1, max deleted version 9
  mmRunningCommand = version 3 ""
files: 8, max deleted version 0
  1 = version 1 uid ((n1,e0),0) crc 6E0689F3
  2 = version 1 uid ((n1,e0),1) crc FFFFFFFF
  3 = version 1 uid ((n1,e0),2) crc FFFFFFFF
  4 = version 1 uid ((n1,e0),3) crc D3ED62E3
  5 = version 1 uid ((n1,e0),4) crc CEE097C7
  6 = version 2 uid ((n1,e1),15) crc FC913FFC
  7 = version 11 uid ((n1,e2),38) crc DF4822D6
  8 = version 3 uid ((n1,e2),41) crc 90406D69
```

Note:

In this example, the `clusterEvents` file reverts to the previous version, which is still intact. The file list in the patched Paxos state file now displays its version number as 3 for the file with the ID no 8 (clusterEvents).

8. Copy the patched CCR Paxos state file into its working copies.

```
[root@node-11 ~]# cp /var/mmfs/ccr/myPatched_ccr.paxos.1 /var/mmfs/ccr/ccr.paxos.1
cp: overwrite '/var/mmfs/ccr/ccr.paxos.1'? y

[root@node-11 ~]# cp /var/mmfs/ccr/myPatched_ccr.paxos.1 /var/mmfs/ccr/ccr.paxos.2
cp: overwrite '/var/mmfs/ccr/ccr.paxos.2'? y
```

9. Restart the `mmsdrserv` daemon.

```
[root@node-11 ~]# mmcommon startCcrMonitor
```

10. Verify that the `mmsdrserv` daemon and its monitor script have restarted.

```
[root@node-11 ~]# ps -C mmfsd,mmccrmonitor,mmsdrserv
  PID TTY          TIME CMD
23355 ?        00:00:00 mmccrmonitor
23576 ?        00:00:00 mmsdrserv
23710 ?        00:00:00 mmccrmonitor
```

11. Run the **mmccr check** command to check the state of CCR.

```
mmccr::HEADER:version:reserved:reserved:NodeId:CheckMnemonic:ErrorCode:ErrorMsg:ListOfFailedEntities:ListOfSucceedEntities:Severi
mmccr::0:1:::1:CCR_CLIENT_INIT:0:::/var/mmfs/ccr,/var/mmfs/ccr/committed,/var/mmfs/ccr/ccr.nodes,Security:OK:
mmccr::0:1:::1:FC_CCR_AUTH_KEYS:0:::/var/mmfs/ssl/authorized_ccr_keys:OK:
mmccr::0:1:::1:FC_CCR_PAXOS_CACHED:0:::/var/mmfs/ccr/cached,/var/mmfs/ccr/cached/ccr.paxos:OK:
mmccr::0:1:::1:FC_CCR_PAXOS_12:0:::/var/mmfs/ccr/ccr.paxos.1,/var/mmfs/ccr/ccr.paxos.2:OK:
mmccr::0:1:::1:PC_LOCAL_SERVER:0:::node-11.localnet.com:OK:
mmccr::0:1:::1:PC_IP_ADDR_LOOKUP:0:::node-11.localnet.com,0.000:OK:
mmccr::0:1:::1:PC_QUORUM_NODES:0:::10.0.100.11:OK:
mmccr::0:1:::1:FC_COMMITTED_DIR:0::0:8:OK:
mmccr::0:1:::1:TC_TIEBREAKER_DISKS:0::::OK:
0
```

After the CCR check command returns without any error on this quorum node, the mm-commands work again.

```
[root@node-11 ~]# mmgetstate

 Node number  Node name        GPFS state
------------------------------------------
       1      node-11          down

[root@node-11 ~]# mmstartup
Mon Feb 26 15:59:47 CET 2018: mmstartup: Starting GPFS ...

[root@node-11 ~]# mmgetstate

 Node number  Node name        GPFS state
------------------------------------------
       1      node-11          active
```

The master copy of the GPFS configuration file can be corrupted. The CCR patch command rolls back to the latest available intact version of a corrupted file. This means that for the `mmsdrfs` file, you lose the configuration changes made between the corrupted and the previous intact version. In such cases, it might be necessary to reboot the quorum node to cleanup the cached memory and all drivers in case GPFS shows different errors in the administration log during startup.

Recovery procedure for a broken multi node cluster

For a multi node cluster containing more than one quorum node, the recovery procedure is similar to a single node cluster with just one quorum node. However the number of possible CCR states is different, as the quorum nodes can have different CCR states.

The step to evaluate the CCR's most recent Paxos state file based on the available number of CCR states is also different. The final patched CCR state and the consolidated committed files must be copied to every quorum node in the cluster to bring back the cluster into a working state.

The following output provides information regarding the cluster used for this procedure. This procedure is for a cluster with three quorum nodes. In case of more than three quorum nodes the procedure must be altered accordingly.

```
[root@node-11 ~]# mmlscluster

GPFS cluster information
========================
  GPFS cluster name:         gpfs-cluster-1.localnet.com
  GPFS cluster id:           317908494312547875
  GPFS UID domain:           localnet.com
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           CCR

GPFS cluster configuration servers:
-----------------------------------
  Primary server:    node-11.localnet.com (not in use)
  Secondary server:  (none)
```

```
 Node  Daemon node name      IP address   Admin node name       Designation
-------------------------------------------------------------------------
   1   node-11.localnet.com  10.0.100.11  node-11.localnet.com  quorum
   2   node-12.localnet.com  10.0.100.12  node-12.localnet.com  quorum
   3   node-13.localnet.com  10.0.100.13  node-13.localnet.com  quorum
   4   node-14.localnet.com  10.0.100.14  node-14.localnet.com
   5   node-15.localnet.com  10.0.100.15  node-15.localnet.com
```

Follow these steps to recover the broken cluster:

1. Run the **mmccrcheck** command to find the missing or corrupted files on all the three quorum nodes.

   ```
   [root@node-11 ~]# mmgetstate -a
   ```

   The command gives output similar to the following:

   ```
   [root@node-11 ~]# mmdsh -N quorumnodes "mmccr check -Y -e" | grep
   "mmdsh\|FC_COMMITTED_DIR"
   node-11.localnet.com:  mmccr::0:1:::1:FC_COMMITTED_DIR:5:Files in
   committed directory missing or corrupted:1:7:WARNING:
   mmdsh: node-11.localnet.com remote shell process had return code 149.
   node-12.localnet.com:  mmccr::0:1:::2:FC_COMMITTED_DIR:5:Files in
   committed directory missing or corrupted:1:7:WARNING:
   mmdsh: node-12.localnet.com remote shell process had return code 149.
   node-13.localnet.com:  mmccr::0:1:::3:FC_COMMITTED_DIR:5:Files in
   committed directory missing or corrupted:1:7:WARNING:
   mmdsh: node-13.localnet.com remote shell process had return code 149.
   ```

   Most mm-commands executed on the different quorum nodes will fail. Run the **mmgetstate** command to view the error pattern of the failed commands:

   ```
   [root@node-11 ~]# mmgetstate -a
   ```

   The command gives output similar to the following:

   ```
   get file failed: Maximum number of retries reached (err 801)
   gpfsClusterInit: Unexpected error from ccr fget mmsdrfs.  Return code: 149
   mmgetstate: Command failed. Examine previous error messages to determine cause.
   ```

2. Shutdown GPFS on all nodes in the cluster to prevent any issues that can be caused by an active mmfsd daemon. The **mmshutdown -a** command also fails in case the CCR is unavailable. Run the **mmdsh** command to bypass the CCR that is still unavailable in such cases.

   ```
   [root@node-11 ~]# mmdsh -N all mmshutdown
   ```

   The command gives output similar to the following:

   ```
   node-13.localnet.com:  Mon Feb 26 17:01:18 CET 2018:mmshutdown: Starting force unmount of GPFS file systems
   node-12.localnet.com:  Mon Feb 26 17:01:17 CET 2018:mmshutdown: Starting force unmount of GPFS file systems
   node-15.localnet.com:  Mon Feb 26 17:01:17 CET 2018:mmshutdown: Starting force unmount of GPFS file systems
   node-14.localnet.com:  Mon Feb 26 17:01:17 CET 2018:mmshutdown: Starting force unmount of GPFS file systems
   node-11.localnet.com:  Mon Feb 26 17:01:18 CET 2018:mmshutdown: Starting force unmount of GPFS file systems
   node-15.localnet.com:  Mon Feb 26 17:01:22 CET 2018:mmshutdown: Shutting down GPFS daemons
   node-13.localnet.com:  Mon Feb 26 17:01:23 CET 2018:mmshutdown: Shutting down GPFS daemons
   node-12.localnet.com:  Mon Feb 26 17:01:22 CET 2018:mmshutdown: Shutting down GPFS daemons
   node-14.localnet.com:  Mon Feb 26 17:01:22 CET 2018:mmshutdown: Shutting down GPFS daemons
   node-11.localnet.com:  Mon Feb 26 17:01:23 CET 2018:mmshutdown: Shutting down GPFS daemons
   node-15.localnet.com:  Mon Feb 26 17:02:11 CET 2018:mmshutdown: Finished
   node-13.localnet.com:  Mon Feb 26 17:02:12 CET 2018:mmshutdown: Finished
   node-11.localnet.com:  Mon Feb 26 17:02:12 CET 2018:mmshutdown: Finished
   node-14.localnet.com:  Mon Feb 26 17:02:11 CET 2018:mmshutdown: Finished
   node-12.localnet.com:  Mon Feb 26 17:02:12 CET 2018:mmshutdown: Finished
   ```

   Note: Use the **mmdsh** command to stop the mmsdrserv daemon and the startup scripts on all quorum nodes in the cluster:

   ```
   [root@node-11 ~]# mmdsh -N quorumnodes "mmcommon killCcrMonitor"
   ```

   Use the following command to check if all the GPFS daemons and the monitor scripts have been stopped on all the quorum nodes:

```
[root@node-11 ~]# mmdsh -N quorumnodes
"ps -C mmfsd,mmccrmonitor,mmsdrserv"
```

The command gives output similar to the following:

```
node-11.localnet.com:    PID TTY          TIME CMD
mmdsh: node-11.localnet.com remote shell process had return code 1.
node-12.localnet.com:    PID TTY          TIME CMD
mmdsh: node-12.localnet.com remote shell process had return code 1.
node-13.localnet.com:    PID TTY          TIME CMD
mmdsh: node-13.localnet.com remote shell process had return code 1.
```

3. Back up the entire CCR state of the three quorum nodes using the tar-command:

```
[root@node-11 ~]#tar -cvf CCR_archive_node 11_20180226170307.tar /var/mmfs/ccr
```

The command gives output similar to the following:

```
/var/mmfs/ccr/
/var/mmfs/ccr/ccr.noauth
/var/mmfs/ccr/ccr.paxos.1
/var/mmfs/ccr/committed/
/var/mmfs/ccr/committed/mmsysmon.json.3.1.cee097c7.010002
/var/mmfs/ccr/committed/clusterEvents.8.12.963fe8ed.010232.bad.26086.4229216064.2018-02-26_16:59:25.250+0100
/var/mmfs/ccr/committed/ccr.nodes.1.1.e7e9c9f0.010000
/var/mmfs/ccr/committed/clusterEvents.8.11.963fe8ed.010231
/var/mmfs/ccr/committed/clusterEvents.8.12.963fe8ed.010232.bad.24040.4169168640.2018-02-26_16:57:39.226+0100
/var/mmfs/ccr/committed/genKeyData.5.1.a043b58e.010004
/var/mmfs/ccr/committed/mmLockFileDB.4.1.ffffffff.010003
/var/mmfs/ccr/committed/ccr.disks.2.1.ffffffff.010001
/var/mmfs/ccr/committed/mmsdrfs.7.10.e29fc7cd.010226
/var/mmfs/ccr/committed/clusterEvents.8.12.963fe8ed.010232.bad.22517.4083746624.2018-02-26_16:57:02.857+0100
/var/mmfs/ccr/committed/genKeyDataNew.6.1.a043b58e.010005
/var/mmfs/ccr/committed/genKeyDataNew.6.2.94f88a51.01010f
/var/mmfs/ccr/committed/clusterEvents.8.12.963fe8ed.010232.bad.27281.1088599808.2018-02-26_16:59:59.681+0100
/var/mmfs/ccr/committed/mmsdrfs.7.11.bf35437.01022a
/var/mmfs/ccr/ccr.disks
/var/mmfs/ccr/cached/
/var/mmfs/ccr/cached/ccr.paxos
/var/mmfs/ccr/ccr.nodes
/var/mmfs/ccr/ccr.paxos.2
```

Note: This example only shows the output for the first quorum node. The command must be executed on all the quorum nodes as needed.

4. Create some temporary directories to store the collected CCR state files.

Two sub-directories must be created inside this `CCRtemp` directory to collect the committed files from all quorum nodes. Of the two sub-directories, one acts as the intermediate directory and the other as the final directory. The committed or final directory keeps the intact files used in the final step to copy back the patched Paxos state. The `committedTemp` or intermediate directory keeps the files only from the current quorum node that are processed during the procedure.

```
[root@node-11 ~]# mkdir -p /root/CCRtemp/committed /root/CCRtemp/committedTemp
[root@node-11 ~]# cd /root/CCRtemp/
```

5. Copy the `/var/mmfs/ccr/ccr.paxos.1` and `/var/mmfs/ccr/ccr.paxos.2` files from every quorum node in the cluster to the current temporary directory, `/root/CCRtemp`, using the following command:

```
[root@node-11 CCRtemp]# scp root@node-11:/var/mmfs/ccr/ccr.paxos.1 ./ccr.paxos.1.node-11
ccr.paxos.1
[root@node-11 CCRtemp]# scp root@node-11:/var/mmfs/ccr/ccr.paxos.2 ./ccr.paxos.2.node-11
ccr.paxos.2
```

Note: You can see the directory structure by using the following command:

```
[root@node-11 CCRtemp]# ls –al
```

The command gives output similar to the following:

```
total 40
drwxr-xr-x  4 root root 4096 Feb 26 17:10 .
dr-xr-x---. 4 root root 4096 Feb 26 17:07 ..
-rw-------  1 root root 4096 Feb 26 17:09 ccr.paxos.1.node-11
```

```
-rw-------  1 root root 4096 Feb 26 17:10 ccr.paxos.1.node-12
-rw-------  1 root root 4096 Feb 26 17:10 ccr.paxos.1.node-13
-rw-------  1 root root 4096 Feb 26 17:09 ccr.paxos.2.node-11
-rw-------  1 root root 4096 Feb 26 17:09 ccr.paxos.2.node-12
-rw-------  1 root root 4096 Feb 26 17:10 ccr.paxos.2.node-13
drwxr-xr-x  2 root root 4096 Feb 26 17:07 committed
drwxr-xr-x  2 root root 4096 Feb 26 17:07 committedTemp
```

6. Switch to the committedTemp subdirectory, and copy the files from the first quorum node into this temporary directory using the following command:

   ```
   [root@node-11 committedTemp]# scp root@node-11:/var/mmfs/ccr/committed/*
   ```

   The command gives output similar to the following:

   ```
   ccr.disks.2.1.ffffffff.010001
   ccr.nodes.1.1.e7e9c9f0.010000
   clusterEvents.8.11.963fe8ed.010231
   clusterEvents.8.12.963fe8ed.010232.bad.22517.4083746624.2018-02-26_16:57:02.857+0100
   clusterEvents.8.12.963fe8ed.010232.bad.24040.4169168640.2018-02-26_16:57:39.226+0100
   clusterEvents.8.12.963fe8ed.010232.bad.26086.4229216064.2018-02-26_16:59:25.250+0100
   clusterEvents.8.12.963fe8ed.010232.bad.27281.1088599808.2018-02-26_16:59:59.681+0100
   genKeyData.5.1.a043b58e.010004
   genKeyDataNew.6.1.a043b58e.010005
   genKeyDataNew.6.2.94f88a51.01010f
   mmLockFileDB.4.1.ffffffff.010003
   mmsdrfs.7.10.e29fc7cd.010226
   mmsdrfs.7.11.bf35437.01022a
   mmsysmon.json.3.1.cee097c7.010002
   ```

   Note: You can see the directory structure by using the following command:

   ```
   [root@node-11 committedTemp]# ls –al
   ```

   The command gives output similar to the following:

   ```
   total 48
   drwxr-xr-x 2 root root 4096 Feb 26 17:12 .
   drwxr-xr-x 4 root root 4096 Feb 26 17:10 ..
   -rw-r--r-- 1 root root    0 Feb 26 17:12 ccr.disks.2.1.ffffffff.010001
   -rw-r--r-- 1 root root  114 Feb 26 17:12 ccr.nodes.1.1.e7e9c9f0.010000
   -rw-r--r-- 1 root root  323 Feb 26 17:12 clusterEvents.8.11.963fe8ed.010231
   -rw-r--r-- 1 root root    0 Feb 26 17:12 clusterEvents.8.12.963fe8ed.010232.bad.22517.4083746624.2018-02-26_16:57:02.857+0100
   -rw-r--r-- 1 root root    0 Feb 26 17:12 clusterEvents.8.12.963fe8ed.010232.bad.24040.4169168640.2018-02-26_16:57:39.226+0100
   -rw-r--r-- 1 root root    0 Feb 26 17:12 clusterEvents.8.12.963fe8ed.010232.bad.26086.4229216064.2018-02-26_16:59:25.250+0100
   -rw-r--r-- 1 root root    0 Feb 26 17:12 clusterEvents.8.12.963fe8ed.010232.bad.27281.1088599808.2018-02-26_16:59:59.681+0100
   -rw------- 1 root root 3531 Feb 26 17:12 genKeyData.5.1.a043b58e.010004
   -rw------- 1 root root 3531 Feb 26 17:12 genKeyDataNew.6.1.a043b58e.010005
   -rw------- 1 root root 3531 Feb 26 17:12 genKeyDataNew.6.2.94f88a51.01010f
   -rw-r--r-- 1 root root    0 Feb 26 17:12 mmLockFileDB.4.1.ffffffff.010003
   -rw-r--r-- 1 root root 4793 Feb 26 17:12 mmsdrfs.7.10.e29fc7cd.010226
   -rw-r--r-- 1 root root 5395 Feb 26 17:12 mmsdrfs.7.11.bf35437.01022a
   -rw-r--r-- 1 root root   38 Feb 26 17:12 mmsysmon.json.3.1.cee097c7.010002
   ```

7. Verify the CRC of the files copied from the first quorum node during the previous step using the following command:

   ```
   [root@node-11 committedTemp]# cksum * | awk '{ printf "%x %s\n", $1, $3 }'
   ```

   The command gives output similar to the following:

   ```
   ffffffff ccr.disks.2.1.ffffffff.010001
   e7e9c9f0 ccr.nodes.1.1.e7e9c9f0.010000
   963fe8ed clusterEvents.8.11.963fe8ed.010231
   ffffffff clusterEvents.8.12.963fe8ed.010232.bad.22517.4083746624.2018-02-26_16:57:02.857+0100
   ffffffff clusterEvents.8.12.963fe8ed.010232.bad.24040.4169168640.2018-02-26_16:57:39.226+0100
   ffffffff clusterEvents.8.12.963fe8ed.010232.bad.26086.4229216064.2018-02-26_16:59:25.250+0100
   ffffffff clusterEvents.8.12.963fe8ed.010232.bad.27281.1088599808.2018-02-26_16:59:59.681+0100
   a043b58e genKeyData.5.1.a043b58e.010004
   a043b58e genKeyDataNew.6.1.a043b58e.010005
   94f88a51 genKeyDataNew.6.2.94f88a51.01010f
   ffffffff mmLockFileDB.4.1.ffffffff.010003
   e29fc7cd mmsdrfs.7.10.e29fc7cd.010226
   bf35437 mmsdrfs.7.11.bf35437.01022a
   cee097c7 mmsysmon.json.3.1.cee097c7.010002
   ```

The faulty files can be identified by a CRC mismatch.

8. Delete the files with mismatching or faulty CRC using the following command:

```
[root@node-11 committedTemp]# rm clusterEvents.8.12.963fe8ed.010232.bad.2*
```

The command gives output similar to the following:

```
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.22517.4083746624.2018-02-26_16:57:02.857+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.24040.4169168640.2018-02-26_16:57:39.226+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.26086.4229216064.2018-02-26_16:59:25.250+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.27281.1088599808.2018-02-26_16:59:59.681+0100'? y
```

9. Move the remaining files into the committed subdirectory using the following command:

```
[root@node-11 committedTemp]# mv -i * ../committed
```

10. Copy the committed files from the next quorum node into the `committedTemp` directory using the following command:

```
[root@node-11 committedTemp]# scp root@node-12:/var/mmfs/ccr/committed/* .
```

The command gives output similar to the following:

```
ccr.disks.2.1.ffffffff.010001
ccr.nodes.1.1.e7e9c9f0.010000
clusterEvents.8.11.963fe8ed.010231
clusterEvents.8.12.963fe8ed.010232.bad.18737.3245463360.2018-02-26_16:57:07.695+0100
clusterEvents.8.12.963fe8ed.010232.bad.19994.3932075776.2018-02-26_16:57:45.020+0100
clusterEvents.8.12.963fe8ed.010232.bad.21275.3060160320.2018-02-26_16:59:33.687+0100
clusterEvents.8.12.963fe8ed.010232.bad.22112.354830080.2018-02-26_16:59:59.467+0100
genKeyData.5.1.a043b58e.010004
genKeyDataNew.6.1.a043b58e.010005
genKeyDataNew.6.2.94f88a51.01010f
mmLockFileDB.4.1.ffffffff.010003
mmsdrfs.7.10.e29fc7cd.010226
mmsdrfs.7.11.bf35437.01022a
mmsysmon.json.3.1.cee097c7.010002
```

11. Verify the CRC of the files using the following command:.

```
[root@node-11 committedTemp]# cksum * | awk '{ printf "%x %s\n", $1, $3 }'
```

The command gives output similar to the following:

```
ffffffff ccr.disks.2.1.ffffffff.010001
e7e9c9f0 ccr.nodes.1.1.e7e9c9f0.010000
963fe8ed clusterEvents.8.11.963fe8ed.010231
ffffffff clusterEvents.8.12.963fe8ed.010232.bad.18737.3245463360.2018-02-26_16:57:07.695+0100
ffffffff clusterEvents.8.12.963fe8ed.010232.bad.19994.3932075776.2018-02-26_16:57:45.020+0100
ffffffff clusterEvents.8.12.963fe8ed.010232.bad.21275.3060160320.2018-02-26_16:59:33.687+0100
ffffffff clusterEvents.8.12.963fe8ed.010232.bad.22112.354830080.2018-02-26_16:59:59.467+0100
a043b58e genKeyData.5.1.a043b58e.010004
a043b58e genKeyDataNew.6.1.a043b58e.010005
94f88a51 genKeyDataNew.6.2.94f88a51.01010f
ffffffff mmLockFileDB.4.1.ffffffff.010003
e29fc7cd mmsdrfs.7.10.e29fc7cd.010226
bf35437 mmsdrfs.7.11.bf35437.01022a
cee097c7 mmsysmon.json.3.1.cee097c7.010002
```

12. Delete the files with mismatching CRC value using the following command:

```
[root@node-11 committedTemp]# rm clusterEvents.8.12.963fe8ed.010232.bad.*
```

The command gives output similar to the following:

```
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.18737.3245463360.2018-02-26_16:57:07.695+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.19994.3932075776.2018-02-26_16:57:45.020+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.21275.3060160320.2018-02-26_16:59:33.687+0100'? y
rm: remove regular empty file 'clusterEvents.8.12.963fe8ed.010232.bad.22112.354830080.2018-02-26_16:59:59.467+0100'? y

100%   38     0.0KB/s   00:00
```

13. Copy the remaining files again into the committed subdirectory, using the cp-command and the -i option .

Note: : Prompt `n` for each file which already exists in the committed subdirectory. This ensures that only the files which do not already exist are copied to the committed subdirectory.

```
root@node-11 committedTemp]# cp -i * ../committed
```

The command gives output similar to the following:

```
cp: overwrite '../committed/ccr.disks.2.1.ffffffff.010001'? n
cp: overwrite '../committed/ccr.nodes.1.1.e7e9c9f0.010000'? n
cp: overwrite '../committed/clusterEvents.8.11.963fe8ed.010231'? n
cp: overwrite '../committed/genKeyData.5.1.a043b58e.010004'? n
cp: overwrite '../committed/genKeyDataNew.6.1.a043b58e.010005'? n
cp: overwrite '../committed/genKeyDataNew.6.2.94f88a51.01010f'? n
cp: overwrite '../committed/mmLockFileDB.4.1.ffffffff.010003'? n
cp: overwrite '../committed/mmsdrfs.7.10.e29fc7cd.010226'? n
cp: overwrite '../committed/mmsdrfs.7.11.bf35437.01022a'? n
cp: overwrite '../committed/mmsysmon.json.3.1.cee097c7.010002'? n
```

14. Remove the files in the `committedTemp` subdirectory using the following command:

```
[root@node-11 committedTemp]# rm *
```

The command gives output similar to the following:

```
rm: remove regular empty file 'ccr.disks.2.1.ffffffff.010001'? y
rm: remove regular file 'ccr.nodes.1.1.e7e9c9f0.010000'? y
rm: remove regular file 'clusterEvents.8.11.963fe8ed.010231'? y
rm: remove regular file 'genKeyData.5.1.a043b58e.010004'? y
rm: remove regular file 'genKeyDataNew.6.1.a043b58e.010005'? y
rm: remove regular file 'genKeyDataNew.6.2.94f88a51.01010f'? y
rm: remove regular empty file 'mmLockFileDB.4.1.ffffffff.010003'? y
rm: remove regular file 'mmsdrfs.7.10.e29fc7cd.010226'? y
rm: remove regular file 'mmsdrfs.7.11.bf35437.01022a'? y
rm: remove regular file 'mmsysmon.json.3.1.cee097c7.010002'? y
```

Note: This step is taken to prepare the `committedTemp` subdirectory for the files from the next quorum node, if any.

15. Repeat steps 10 to 14 for all the remaining and available quorum nodes . The `/root/CCRtemp/committed` directory now contains all the intact files from all the quorum nodes, and it can be used to patch the CCR Paxos state.

16. Change back to the parent directory of the current subdirectory and get the most recent Paxos state based on the Paxos state files in this directory by using the **mmccr readpaxos** command:

```
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.1.node-11 | grep seq
```

The command gives output similar to the following:

```
dblk: seq 53, mbal (0.0), bal (0.0), inp ((n0,e0),0):(none):-1:None, leaderChallengeVersion 0
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.2.node-11 | grep seq
dblk: seq 52, mbal (1.1), bal (1.1), inp ((n0,e0),0):lu:3:[1,23333], leaderChallengeVersion 0
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.1.node-12 | grep seq
dblk: seq 53, mbal (0.0), bal (0.0), inp ((n0,e0),0):(none):-1:None, leaderChallengeVersion 0
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.2.node-12 | grep seq
dblk: seq 52, mbal (1.1), bal (1.1), inp ((n0,e0),0):lu:3:[1,23333], leaderChallengeVersion 0
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.1.node-13 | grep seq
dblk: seq 53, mbal (0.0), bal (0.0), inp ((n0,e0),0):(none):-1:None, leaderChallengeVersion 0
[root@node-11 CCRtemp]# mmccr readpaxos ccr.paxos.2.node-13 | grep seq
dblk: seq 52, mbal (1.1), bal (1.1), inp ((n0,e0),0):lu:3:[1,23333], leaderChallengeVersion 0
```

The CCR has two Paxos state files in its `/var/mmfs/ccr` directory, `ccr.paxos.1` and `ccr.paxos.2`. CCR writes alternately to these two files. Maintaining dual copies allows the CCR to always have a copy intact in case the write to one the file fails for some reason and makes this file corrupt. The most recent file is the file with the higher sequence number in it. Therefore, the CCR Paxos state file with the highest sequence number is the most recent one. Ensure that you use the path to the most recent Paxos state file while using the **mmccr readpaxos** command.

In the example above the CCR Paxos state file `ccr.paxos.1.node-11` is the most recent one. The `ccr.paxos.1.node-11` file has the sequence number 53. In case of a multi node cluster, not all quorum nodes must have the same set of sequence numbers, based on how many updates the CCR has seen until the **readpaxos** command is invoked.

The `ccr.paxos.1.node-11` file acts as the input file for the following patching step. The **mmccr patchpaxos** command must be invoked in the current CCR temp directory. The first parameter of the

mmccr patchpaxos command is the path to the most recent CCR Paxos state file. The second parameter of the **mmccr patchpaxos** command is the path to the collected intact CCR files gathered during the previous steps. The third parameter of the **mmccr patchpaxos** command is the path to the Paxos state file which will be created when this **mmccr patchpaxos** command is run:

```
[root@node-11 CCRtemp]# mmccr patchpaxos ./ccr.paxos.1.node-11 ./committed/
./myPatched_ccr.paxos.1
```

The command gives output similar to the following:

```
Committed state found in ./ccr.paxos.1.node-11:
config: minNodes: 1 version 0
  nodes: [(N1,S0,V0,L1), (N2,S1,V0,L1), (N3,S2,V0,L1)]
  disks: []
leader: id 1 version 3
updates: horizon -1
  {(n1,e0): 5, (n1,e1): 33, (n1,e2): 50}
values: 1, max deleted version 9
  mmRunningCommand = version 3 ""
files: 8, max deleted version 0
  1 = version 1 uid ((n1,e0),0) crc E7E9C9F0
  2 = version 1 uid ((n1,e0),1) crc FFFFFFFF
  3 = version 1 uid ((n1,e0),2) crc CEE097C7
  4 = version 1 uid ((n1,e0),3) crc FFFFFFFF
  5 = version 1 uid ((n1,e0),4) crc A043B58E
  6 = version 2 uid ((n1,e1),15) crc 94F88A51
  7 = version 11 uid ((n1,e2),42) crc 0BF35437
  8 = version 12 uid ((n1,e2),50) crc 963FE8ED

Comparing to content of './committed/':

  match file: name: 'ccr.nodes' suffix: '1.1.e7e9c9f0.010000' id: 1 version: 1 crc: e7e9c9f0 uid:
((n1,e0),0) and file list entry: 1.1.e7e9c9f0.010000
  match file: name: 'ccr.disks' suffix: '2.1.ffffffff.010001' id: 2 version: 1 crc: ffffffff uid:
((n1,e0),1) and file list entry: 2.1.ffffffff.010001
  match file: name: 'mmsysmon.json' suffix: '3.1.cee097c7.010002' id: 3 version: 1 crc: cee097c7 uid:
((n1,e0),2) and file list entry: 3.1.cee097c7.010002
  match file: name: 'mmLockFileDB' suffix: '4.1.ffffffff.010003' id: 4 version: 1 crc: ffffffff uid: ((n1,e0),3) and file list en
  match file: name: 'genKeyData' suffix: '5.1.a043b58e.010004' id: 5 version: 1 crc: a043b58e uid:
((n1,e0),4) and file list entry: 5.1.a043b58e.010004
  match file: name: 'genKeyDataNew' suffix: '6.2.94f88a51.01010f' id: 6 version: 2 crc: 94f88a51 uid:
((n1,e1),15) and file list entry: 6.2.94f88a51.01010f
  match file: name: 'mmsdrfs' suffix: '7.11.bf35437.01022a' id: 7 version: 11 crc: bf35437 uid:
((n1,e2),42) and file list entry: 7.11.bf35437.01022a
  older:   name: 'clusterEvents' suffix: '8.11.963fe8ed.010231' id: 8 version: 11 crc: 963fe8ed uid: ((n1,e2),49)
  *** reverting committed file list version 12 uid ((n1,e2),50)

Found 7 matching, 0 deleted, 0 added, 0 updated, 1 reverted, 0 reset


Verifying update history

Writing 1 changes to ./myPatched_ccr.paxos.1
config: minNodes: 1 version 0
  nodes: [(N1,S0,V0,L1), (N2,S1,V0,L1), (N3,S2,V0,L1)]
  disks: []
leader: id 1 version 3
updates: horizon -1
  {(n1,e0): 5, (n1,e1): 33, (n1,e2): 50}
values: 1, max deleted version 9
  mmRunningCommand = version 3 ""
files: 8, max deleted version 0
  1 = version 1 uid ((n1,e0),0) crc E7E9C9F0
  2 = version 1 uid ((n1,e0),1) crc FFFFFFFF
  3 = version 1 uid ((n1,e0),2) crc CEE097C7
  4 = version 1 uid ((n1,e0),3) crc FFFFFFFF
  5 = version 1 uid ((n1,e0),4) crc A043B58E
  6 = version 2 uid ((n1,e1),15) crc 94F88A51
  7 = version 11 uid ((n1,e2),42) crc 0BF35437
  8 = version 11 uid ((n1,e2),49) crc 963FE8ED
```

17. Copy the patched CCR Paxos state file and the files in the committed directory back to the appropriate directories on every quorum node using the following command:

```
[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-11:/var/mmfs/ccr/ccr.paxos.1
```

The command gives output similar to the following:

```
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-11:/var/mmfs/ccr/ccr.paxos.2
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-12:/var/mmfs/ccr/ccr.paxos.1
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-12:/var/mmfs/ccr/ccr.paxos.2
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-13:/var/mmfs/ccr/ccr.paxos.1
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp myPatched_ccr.paxos.1 root@node-13:/var/mmfs/ccr/ccr.paxos.2
myPatched_ccr.paxos.1

[root@node-11 CCRtemp]# scp ./committed/* root@node-11:/var/mmfs/ccr/committed/
ccr.disks.2.1.ffffffff.010001
ccr.nodes.1.1.e7e9c9f0.010000
clusterEvents.8.11.963fe8ed.010231
genKeyData.5.1.a043b58e.010004
genKeyDataNew.6.1.a043b58e.010005
genKeyDataNew.6.2.94f88a51.01010f
mmLockFileDB.4.1.ffffffff.010003
mmsdrfs.7.10.e29fc7cd.010226
mmsdrfs.7.11.bf35437.01022a
mmsysmon.json.3.1.cee097c7.010002

[root@node-11 CCRtemp]# scp ./committed/* root@node-12:/var/mmfs/ccr/committed/
ccr.disks.2.1.ffffffff.010001
ccr.nodes.1.1.e7e9c9f0.010000
clusterEvents.8.11.963fe8ed.010231
genKeyData.5.1.a043b58e.010004
genKeyDataNew.6.1.a043b58e.010005
genKeyDataNew.6.2.94f88a51.01010f
mmLockFileDB.4.1.ffffffff.010003
mmsdrfs.7.10.e29fc7cd.010226
mmsdrfs.7.11.bf35437.01022a
mmsysmon.json.3.1.cee097c7.010002

[root@node-11 CCRtemp]# scp ./committed/* root@node-13:/var/mmfs/ccr/committed/
ccr.disks.2.1.ffffffff.010001
ccr.nodes.1.1.e7e9c9f0.010000
clusterEvents.8.11.963fe8ed.010231
genKeyData.5.1.a043b58e.010004
genKeyDataNew.6.1.a043b58e.010005
genKeyDataNew.6.2.94f88a51.01010f
mmLockFileDB.4.1.ffffffff.010003
mmsdrfs.7.10.e29fc7cd.010226
mmsdrfs.7.11.bf35437.01022a
mmsysmon.json.3.1.cee097c7.010002                                                    100%   38     0.0KB/s   00:00
```

18. Start the `mmsdrserv` daemon and the monitor script which was stopped previously:

    ```
    [root@node-11 ~]# mmdsh -N quorumnodes "mmcommon startCcrMonitor"
    ```

19. Verify that the `mmsdrserv` daemon and its monitor script have restarted using the following command:

    ```
    [root@node-11 ~]# mmdsh -N quorumnodes "ps -C mmfsd,mmccrmonitor,mmsdrserv"
    ```

    The command gives output similar to the following:

    ```
    node-11.localnet.com:    PID TTY          TIME CMD
    node-11.localnet.com:   3518 ?        00:00:00 mmccrmonitor
    node-11.localnet.com:   3734 ?        00:00:00 mmsdrserv
    node-11.localnet.com:   3816 ?        00:00:00 mmccrmonitor
    node-12.localnet.com:    PID TTY          TIME CMD
    node-12.localnet.com:  30356 ?        00:00:00 mmccrmonitor
    node-12.localnet.com:  30572 ?        00:00:00 mmsdrserv
    node-12.localnet.com:  30648 ?        00:00:00 mmccrmonitor
    ```

```
node-13.localnet.com:    PID TTY          TIME CMD
node-13.localnet.com:    738 ?        00:00:00 mmccrmonitor
node-13.localnet.com:    958 ?        00:00:00 mmsdrserv
node-13.localnet.com:   1040 ?        00:00:00 mmccrmonitor
```

The **mmccr check** command will succeed now. The **mmccr check** gives output similar to the following:

```
[root@node-11 ~]# mmdsh -N quorumnodes "mmccr check -Y -e" | grep "mmdsh\|FC_COMMITTED_DIR"
node-12.localnet.com:  mmccr::0:1:::2:FC_COMMITTED_DIR:0::0:8:OK:
node-11.localnet.com:  mmccr::0:1:::1:FC_COMMITTED_DIR:0::0:8:OK:
node-13.localnet.com:  mmccr::0:1:::3:FC_COMMITTED_DIR:0::0:8:OK:
```

The mm-commands are active now, however the cluster is still down:

```
[root@node-11 ~]# mmgetstate -a

 Node number  Node name        GPFS state
------------------------------------------
       1      node-11          down
       2      node-12          down
       3      node-13          down
       4      node-14          down
       5      node-15          down

[root@node-11 ~]# mmlscluster

GPFS cluster information
========================
  GPFS cluster name:         gpfs-cluster-1.localnet.com
  GPFS cluster id:           317908494312547875
  GPFS UID domain:           localnet.com
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           CCR

GPFS cluster configuration servers:
-----------------------------------
  Primary server:    node-11.localnet.com (not in use)
  Secondary server:  (none)

 Node  Daemon node name      IP address   Admin node name       Designation
--------------------------------------------------------------------------------
   1   node-11.localnet.com  10.0.100.11  node-11.localnet.com  quorum
   2   node-12.localnet.com  10.0.100.12  node-12.localnet.com  quorum
   3   node-13.localnet.com  10.0.100.13  node-13.localnet.com  quorum
   4   node-14.localnet.com  10.0.100.14  node-14.localnet.com
   5   node-15.localnet.com  10.0.100.15  node-15.localnet.com
```

20. Start GPFS on all nodes and get the clusters up again using the following command:

```
[root@node-11 ~]# mmstartup -a
```

The command gives output similar to the following:

```
Mon Feb 26 18:04:05 CET 2018: mmstartup: Starting GPFS ...

[root@node-11 ~]# mmgetstate -a

 Node number  Node name        GPFS state
------------------------------------------
       1      node-11          active
       2      node-12          active
       3      node-13          active
       4      node-14          active
       5      node-15          active
```

Note:

The master copy of the GPFS configuration file can be corrupted. The CCR patch command rolls back to the latest available intact version of a corrupted file. This means that for the `mmsdrfs` file, you lose the

configuration changes made between the corrupted and the previous intact version. In such cases, it might be necessary to reboot the quorum node to cleanup the cached memory and all drivers in case GPFS shows different errors in the administration log during startup.

Chapter 31. Support for troubleshooting

This topic describes the support that is available for troubleshooting any issues that you might encounter while using IBM Spectrum Scale .

Contacting IBM support center

Specific information about a problem such as: symptoms, traces, error logs, GPFS logs, and file system status is vital to IBM in order to resolve a GPFS problem.

Obtain this information as quickly as you can after a problem is detected, so that error logs will not wrap and system parameters that are always changing, will be captured as close to the point of failure as possible. When a serious problem is detected, collect this information and then call IBM. For more information, see:

- "Information to be collected before contacting the IBM Support Center"
- "How to contact the IBM Support Center" on page 471.

Information to be collected before contacting the IBM Support Center

For effective communication with the IBM Support Center to help with problem diagnosis, you need to collect certain information.

Information to be collected for all problems related to GPFS

Regardless of the problem encountered with GPFS, the following data should be available when you contact the IBM Support Center:

1. A description of the problem.
2. Output of the failing application, command, and so forth.
3. A tar file generated by the **gpfs.snap** command that contains data from the nodes in the cluster. In large clusters, the **gpfs.snap** command can collect data from certain nodes (for example, the affected nodes, NSD servers, or manager nodes) using the **-N** option.

   If the **gpfs.snap** command cannot be run, collect these items:

   a. Any error log entries relating to the event:
      - On an AIX node, issue this command:

        ```
        errpt -a
        ```

      - On a Linux node, create a tar file of all the entries in the **/var/log/messages** file from all nodes in the cluster or the nodes that experienced the failure. For example, issue the following command to create a tar file that includes all nodes in the cluster:

        ```
        mmdsh -v -N all "cat /var/log/messages" > all.messages
        ```

      - On a Windows node, use the **Export List...** dialog in the Event Viewer to save the event log to a file.

   b. A master GPFS log file that is merged and chronologically sorted for the date of the failure (see "Creating a master GPFS log file" on page 198).

   c. If the cluster was configured to store dumps, collect any internal GPFS dumps written to that directory relating to the time of the failure. The default directory is **/tmp/mmfs**.

   d. On a failing Linux node, gather the installed software packages and the versions of each package by issuing this command:

      ```
      rpm -qa
      ```

e. On a failing AIX node, gather the name, most recent level, state, and description of all installed software packages by issuing this command:

   ```
   lslpp -l
   ```

f. File system attributes for all of the failing file systems, issue:

   ```
   mmlsfs Device
   ```

g. The current configuration and state of the disks for all of the failing file systems, issue:

   ```
   mmlsdisk Device
   ```

h. A copy of file **/var/mmfs/gen/mmsdrfs** from the primary cluster configuration server.

4. For Linux on Z, collect the data of the operating system as described in the *Linux on z Systems*® *Troubleshooting Guide* (www.ibm.com/support/knowledgecenter/linuxonibm/liaaf/lnz_r_sv.html).
5. If you are experiencing one of the following problems, see the appropriate section before contacting the IBM Support Center:
   - For delay and deadlock issues, see "Additional information to collect for delays and deadlocks."
   - For file system corruption or MMFS_FSSTRUCT errors, see "Additional information to collect for file system corruption or MMFS_FSSTRUCT errors."
   - For GPFS daemon crashes, see "Additional information to collect for GPFS daemon crashes."

Additional information to collect for delays and deadlocks

When a delay or deadlock situation is suspected, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, ensure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in "Information to be collected for all problems related to GPFS" on page 469.
2. The deadlock debug data collected automatically.
3. If the cluster size is relatively small and the **maxFilesToCache** setting is not high (less than 10,000), issue the following command:

   ```
   gpfs.snap --deadlock
   ```

   If the cluster size is large or the **maxFilesToCache** setting is high (greater than 1M), issue the following command:

   ```
   gpfs.snap --deadlock --quick
   ```

Additional information to collect for file system corruption or MMFS_FSSTRUCT errors

When file system corruption or **MMFS_FSSTRUCT** errors are encountered, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, ensure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in "Information to be collected for all problems related to GPFS" on page 469.
2. Unmount the file system everywhere, then run **mmfsck -n** in offline mode and redirect it to an output file.

The IBM Support Center will determine when and if you should run the **mmfsck -y** command.

Additional information to collect for GPFS daemon crashes

When the GPFS daemon is repeatedly crashing, the IBM Support Center will need additional information to assist with problem diagnosis. If you have not done so already, ensure you have the following information available before contacting the IBM Support Center:

1. Everything that is listed in "Information to be collected for all problems related to GPFS" on page 469.

2. Ensure the **/tmp/mmfs** directory exists on all nodes. If this directory does not exist, the GPFS daemon will not generate internal dumps.
3. Set the traces on this cluster and *all* clusters that mount any file system from this cluster:
   ```
   mmtracectl --set --trace=def --trace-recycle=global
   ```
4. Start the trace facility by issuing:
   ```
   mmtracectl --start
   ```
5. Recreate the problem if possible or wait for the assert to be triggered again.
6. Once the assert is encountered on the node, turn off the trace facility by issuing:
   ```
   mmtracectl --off
   ```
   If traces were started on multiple clusters, **mmtracectl --off** should be issued immediately on all clusters.
7. Collect **gpfs.snap** output:
   ```
   gpfs.snap
   ```

How to contact the IBM Support Center

The IBM Support Center is available for various types of IBM hardware and software problems that GPFS customers may encounter.

These problems include the following:

- IBM hardware failure
- Node halt or crash not related to a hardware failure
- Node hang or response problems
- Failure in other software supplied by IBM

If you have an IBM Software Maintenance service contract

If you have an IBM Software Maintenance service contract, contact the IBM Support Center, as follows:

Your location	Method of contacting the IBM Support Center
In the United States	Call **1-800-IBM-SERV** for support.
Outside the United States	Contact your local IBM Support Center or see the Directory of worldwide contacts (www.ibm.com/planetwide).

When you contact the IBM Support Center, the following will occur:

1. You will be asked for the information you collected in "Information to be collected before contacting the IBM Support Center" on page 469.
2. You will be given a time period during which an IBM representative will return your call. Be sure that the person you identified as your contact can be reached at the phone number you provided in the PMR.
3. An online Problem Management Record (PMR) will be created to track the problem you are reporting, and you will be advised to record the PMR number for future reference.
4. You may be requested to send data related to the problem you are reporting, using the PMR number to identify it.
5. Should you need to make subsequent calls to discuss the problem, you will also use the PMR number to identify the problem.

If you do not have an IBM Software Maintenance service contract

If you do not have an IBM Software Maintenance service contract, contact your IBM sales representative to find out how to proceed. Be prepared to provide the information you collected in "Information to be collected before contacting the IBM Support Center" on page 469.

For failures in non-IBM software, follow the problem-reporting procedures provided with that product.

Call home notifications to IBM Support

The call home feature automatically notifies IBM Support if certain types of events occur in the system. Using this information, IBM Support can contact the system administrator in case of any issues. Configuring call home reduces the response time for IBM Support to address the issues.

The details are collected from individual nodes that are marked as call home child nodes in the cluster. The details from each child node are collected by the call home node. You need to create a call home group by grouping call home child nodes. One of the nodes in the group is configured as the call home node, and it performs data collection and upload.

The data gathering and upload can be configured individually on each group. Use the groups to reflect logical units in the cluster. For example, it is easier to manage when you create a group for all CES nodes and another group for all non-CES nodes.

You can also use the **Settings** > **Call Home** page in the IBM Spectrum Scale management GUI to configure the call home feature. For more information on configuring call home using GUI, see "Configuring call home using GUI" on page 171.

For more information on how to configure and manage the call home feature, see Chapter 8, "Monitoring the IBM Spectrum Scale system by using call home," on page 163.

Chapter 32. References

The IBM Spectrum Scale system displays messages if it encounters any issues when you configure the system. The message severity tags helps to assess the severity of the issue.

Events

The recorded events are stored in local database on each node. The user can get a list of recorded events by using the **mmhealth node eventlog** command.

The recorded events can also be displayed through GUI.

The following sections list the RAS events that are applicable to various components of the IBM Spectrum Scale system:

AFM events

The following table lists the events that are created for the *AFM* component.

Table 64. Events for the AFM component

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_fileset_found	INFO_ADD_ENTITY	INFO	The afm fileset {0} was found.	An AFM fileset was detected.	An AFM fileset was detected. This is detected through the appearance of the fileset in the **mmdiag --afm** output.	N/A
afm_fileset_vanished	INFO_DELETE_ENTITY	INFO	The afm fileset {0} has vanished.	An AFM fileset is not in use anymore.	An AFM fileset is not in use anymore. This is detected through the absence of the fileset in the 'mmdiag --afm' output.	N/A
afm_cache_up	STATE_CHANGE	INFO	The AFM cache fileset {0} is active.	The AFM cache is up and ready for operations.	The AFM cache shows 'Active' or 'Dirty' as status in **mmdiag --afm**. This is expected and shows, that the cache is healthy.	N/A
afm_cache_disconnected	STATE_CHANGE	WARNING	Fileset {0} is disconnected.	The AFM cache fileset is not connected to its home server.	Shows that the connectivity between the MDS (Metadata Server of the fileset) and the mapped home server is lost.	The user action is based on the source of the disconnect. Check the settings on both sites - home and cache. Correct the connectivity issues. The state should change automatically back to active after solving the issues.

Table 64. Events for the AFM component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_cache_dropped	STATE_CHANGE	ERROR	Fileset {0} is in Dropped state.	The AFM cache fileset state moves to Dropped state.	An AFM cache fileset state moves to dropped due to different reasons like recovery failures, failback failures, etc.	There are many different reasons why the cache might go into the Dropped state. Some depend on previous cache states or what the user did before. Those different reasons and their steps to fix the issue can be found in "Monitoring fileset states for AFM DR" on page 136.
afm_cache_expired	INFO	ERROR	Fileset {0} in {1}-mode is now in Expired state.	Cache contents are no longer accessible due to time expiration.	Cache contents are no longer accessible due to time expiration.	N/A
afm_failback_complete	STATE_CHANGE	WARNING	The AFM cache fileset {0} in {1}-mode is in FailbackCompleted state.	The independent-writer failback is finished.	The independent-writer failback is finished, and needs further user actions.	The administrator must run the **mmafmctl failback --stop** to move the IW cache to Active state.
afm_failback_running	STATE_CHANGE	WARNING	The AFM cache fileset {0} in {1}-mode is in FailbackInProgress state.	A failback process on the independent-writer cache is in progress.	A failback process has been initiated on the independent-writer cache and is in progress.	No user action is needed at this point. After completion the state will automatically change into the FailbackCompleted state.
afm_failover_running	STATE_CHANGE	WARNING	The AFM cache fileset {0} is in FailoverInProgress state.	The AFM cache fileset is in the middle of a failover process.	The AFM cache fileset is in the middle of a failover process.	No user action is needed at this point. The cache state is moved automatically to **Active** when the failover is completed.
afm_flush_only	STATE_CHANGE	WARNING	The AFM cache fileset {0} is in FlushOnly state.	Indicates that operations are queued but have not started to flush to the home server.	Indicates that the operation of queuing is finished but flushing to the home server did not start yet.	This state will automatically change and needs no user action.
afm_cache_inactive	STATE_CHANGE	WARNING	The AFM cache fileset {0} is in Inactive state	Initial operations are not triggered by the user on this fileset yet.	The AFM fileset is in 'Inactive' state until initial operations on the fileset are triggered by the user.	Trigger first operations e.g with the **mmafmctl prefetch** command.
afm_failback_needed	STATE_CHANGE	ERROR	The AFM cache fileset {0} in {1}-mode is in NeedFailback state.	A previous failback operation could not be completed and needs to be rerun again.	This state is reached when an previously initialized failback was interrupted and was not completed.	Failback automatically gets triggered on the fileset. The administrator can manually rerun a failback with the **mmafmctl failback** command.

Table 64. Events for the AFM component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_resync_needed	STATE_CHANGE	WARNING	The AFM cache fileset {0} in {1}-mode is in NeedsResync state.	The AFM cache fileset detects some accidental corruption of data on the home server.	The AFM cache fileset detects some accidental corruption of data on the home server.	Use the **mmafmctl resync** command to trigger a resync. The fileset moves automatically to the Active state afterwards.
afm_queue_only	STATE_CHANGE	INFO	The AFM cache fileset {0} in {1}-mode is in QueueOnly state.	The AFM cache fileset is in the process of queueing changes. These changes are not flushed yet to home.	The AFM cache fileset is in the process of queueing changes.	N/A
afm_cache_recovery	STATE_CHANGE	WARNING	The AFM cache fileset {0} in {1}-mode is in Recovery state.	In this state the AFM cache fileset recovers from a previous failure and identifies changes that need to be synchronized to its home server.	A previous failure triggered a cache recovery.	This state will be automatically changed back to Active when the recovery is finished.
afm_cache_unmounted	STATE_CHANGE	ERROR	The AFM cache fileset {0} is in Unmounted state.	The AFM cache fileset is in an Unmounted state because of issues on the home site.	The AFM cache fileset will be in this state if the home server's NFS-mount is not accessible, if the home server's exports are not exported properly or if the home server's export does not exist.	Resolve issues on the home server's site. Later this state will change automatically.
afm_recovery_running	STATE_CHANGE	WARNING	AFM fileset {0} is triggered for recovery start.	A recovery was started on this AFM fileset.	A recovery process was started on this AFM cache fileset.	N/A
afm_recovery_finished	STATE_CHANGE	INFO	A recovery process ended for the AFM cache fileset {0}.	A recovery process has ended on this AFM fileset.	A recovery process has ended on this AFM cache fileset.	N/A
afm_fileset_expired	INFO	WARNING	The contents of the AFM cache fileset {0} are expired.	The AFM cache fileset contents are expired.	The contents of a fileset expire either as a result of the fileset being disconnected for the expiration timeout value, or when the fileset is marked as expired using the AFM administration commands. This event is triggered through an AFM callback.	N/A

Table 64. Events for the AFM component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_fileset_unexpired	INFO	WARNING	The contents of the AFM cache fileset {0} are unexpired.	The AFM cache fileset contents are unexpired.	The contents of these filesets are unexpired, and now available for operations. This event is triggered when the home gets reconnected and cache contents become available, or the administrator runs the **mmafmctl unexpire** command on the cache fileset. This event is triggered through an AFM callback.	N/A
afm_queue_dropped	STATE_CHANGE	ERROR	The AFM cache fileset {0} encountered an error synchronizing with its remote cluster.	The AFM cache fileset encountered an error synchronizing with its remote cluster. It cannot synchronize with the remote cluster until AFM recovery is executed.	This event occurs when a queue is dropped on the gateway node.	Initiate® I/O to trigger recovery on this fileset.
afm_recovery_failed	STATE_CHANGE	ERROR	AFM recovery on fileset {0} failed with error {1}.	AFM recovery failed.	AFM recovery failed.	Recovery will be retried on next access after the recovery retry interval (OR). Manually resolve known problems and recover the fileset.
afm_rpo_miss	INFO	INFO	AFM RPO miss on fileset {0}	The primary fileset is triggering RPO snapshot at a given time interval.	The AFM RPO (Recovery Point Objective) MISS event can occur if a RPO snapshot is missed due to network delay or failure of its creation on the secondary site.	No user action is required. Failed RPOs are re-queued on the primary gateway and retried at the secondary site.

Table 64. Events for the AFM component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_prim_init_fail	STATE_CHANGE	ERROR	The AFM cache fileset {0} is in PrimInitFail state.	The AFM cache fileset is in **PrimInitFail** state. No data will be moved from the primary to the secondary fileset.	This rare state appears if the initial creation of psnap0 on the primary cache fileset failed.	1. Check if the fileset is available, and exported to be used as primary. 2. The gateway node should be able to access this mount. 3. The primary id should be setup on the secondary gateway. 4. It might also help to use the **mmafmctl converToPrimary** command on the primary fileset again.
afm_prim_init_running	STATE_CHANGE	WARNING	The AFM primary cache fileset {0} is in **PrimInitProg** state.	The AFM cache fileset is synchronizing psnap0 with its secondary AFM cache fileset.	This AFM cache fileset is a primary fileset and synchronizing the content of psnap0 to the secondary AFM cache fileset.	This state will change back to **Active** automatically when the synchronization is finished.
afm_cache_suspended	STATE_CHANGE	WARNING	AFM fileset {0} was suspended.	The AFM cache fileset is suspended.	The AFM cache fileset is in **Suspended** state.	Run the **mmafmctl resume** command to resume operations on the fileset.
afm_cache_stopped	STATE_CHANGE	WARNING	The AFM fileset {0} was stopped.	The AFM cache fileset is stopped.	The AFM cache fileset is in **Stopped** state.	Run the **mmafmctl restart** command to continue operations on the fileset.
afm_sensors_active	TIP	HEALTHY	The AFM perfmon sensors are active.	The AFM perfmon sensors are active. This event's monitor is only running once an hour.	The AFM perfmon sensors' period attribute is greater than 0.	N/A
afm_sensors_inactive	TIP	TIP	The following AFM perfmon sensors are inactive: {0}.	The AFM perfmon sensors are inactive. This event's monitor is only running once an hour.	The AFM perfmon sensors' period attribute is 0.	Set the period attribute of the AFM sensors greater than 0. Use the command `mmperfmon config update SensorName.period=N`, where **SensorName** is one of the AFM sensors' name, and *N* is a natural number greater 0. You can also hide this event by using the **mmhealth event hide afm_sensors_inactive** command.
afm_fileset_created	INFO	INFO	AFM fileset {0} was created.	An AFM fileset was created.	An AFM fileset was created.	N/A

Table 64. Events for the AFM component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
afm_fileset_deleted	INFO	INFO	AFM fileset {0} was deleted.	An AFM fileset was deleted.	An AFM fileset was deleted.	N/A
afm_fileset_linked	INFO	INFO	AFM fileset {0} was linked.	An AFM fileset was linked.	An AFM fileset was linked.	N/A
afm_fileset_unlinked	INFO	INFO	AFM fileset {0} was unlinked.	An AFM fileset was unlinked.	An AFM fileset was unlinked.	N/A
afm_fileset_changed	INFO	INFO	AFM fileset {0} was changed	An AFM fileset was changed.	An AFM fileset was changed.	N/A
afm_sensors_ not_configured	TIP	TIP	The AFM perfmon sensor {0} is not configured.	The AFM perfmon sensor does not exist in **mmperfmon config show**	The AFM perfmon sensor is not configured in the sensors configuration file.	Include the sensors into the perfmon configuration through the **mmperfmon config update --config-file InputFile** command. An example for the configuration file can be found in the *mmperfmon command* section in the *IBM Spectrum Scale: Command and Programming Reference*.

Authentication events

The following table lists the events that are created for the *AUTH* component.

Table 65. Events for the AUTH component

Event	Event Type	Severity	Message	Description	Cause	User Action
ads_down	STATE_CHANGE	ERROR	The external Active Directory (AD) server is unresponsive.	The external AD server is unresponsive.	The local node is unable to connect to any AD server.	Local node is unable to connect to any AD server. Verify the network connection and check whether the AD servers are operational.
ads_failed	STATE_CHANGE	ERROR	The local winbindd service is unresponsive.	The local winbindd service is unresponsive.	The local winbindd service does not respond to ping requests. This is a mandatory prerequisite for Active Directory service.	Try to restart winbindd service and if not successful, perform winbindd troubleshooting procedures.
ads_up	STATE_CHANGE	INFO	The external Active Directory (AD) server is up.	The external AD server is up.	The external AD server is operational.	N/A
ads_warn	INFO	WARNING	External Active Directory (AD) server monitoring service returned unknown result	External AD server monitoring service returned unknown result.	An internal error occurred while monitoring the external AD server.	An internal error occurred while monitoring the external AD server. Perform the troubleshooting procedures.

Table 65. Events for the AUTH component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ldap_down	STATE_CHANGE	ERROR	The external LDAP server {0} is unresponsive.	The external LDAP server <LDAP server> is unresponsive.	The local node is unable to connect to the LDAP server.	Local node is unable to connect to the LDAP server. Verify the network connection and check whether the LDAP server is operational.
ldap_up	STATE_CHANGE	INFO	External LDAP server {0} is up.	The external LDAP server is up.	The external LDAP server is operational.	N/A
nis_down	STATE_CHANGE	ERROR	External Network Information Server (NIS) {0} is unresponsive.	External NIS server <NIS server> is unresponsive.	The local node is unable to connect to any NIS server.	Local node is unable to connect to any NIS server. Verify network connection and check whether the NIS servers are operational.
nis_failed	STATE_CHANGE	ERROR	The ypbind daemon is unresponsive.	The ypbind daemon is unresponsive.	The local ypbind daemon does not respond.	Local ypbind daemon does not respond. Try to restart the ypbind daemon. If not successful, perform ypbind troubleshooting procedures.
nis_up	STATE_CHANGE	INFO	External Network Information Server (NIS) {0} is up	External NIS server is operational.		N/A
nis_warn	INFO	WARNING	External Network Information Server (NIS) monitoring returned unknown result.	The external NIS server monitoring returned unknown result.	An internal error occurred while monitoring external NIS server.	Check the health state of the authentication service. Check if the sysmonitor is running. Perform the sysmonitor troubleshooting procedures to understand why the status cannot be collected.
sssd_down	STATE_CHANGE	ERROR	SSSD process is not functioning.	The SSSD process is not functioning.	The SSSD authentication service is not running.	Verify the authentication configuration. Verify the connection with the authentication server. Try to restart the sssd service manually using the **systemctl restart sssd** command. If the restart is unsuccessful, perform the SSSD troubleshooting procedures.
sssd_restart	INFO	INFO	SSSD process is not functioning. Trying to start it.	Attempt to start the SSSD authentication process.	The SSSD process is not functioning.	N/A
sssd_up	STATE_CHANGE	INFO	SSSD process is now functioning.	The SSSD process is now functioning properly.	The SSSD authentication process is running.	N/A

Table 65. Events for the AUTH component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
sssd_warn	INFO	WARNING	SSSD service monitoring returned unknown result.	The SSSD authentication service monitoring returned unknown result.	An internal error occurred in the SSSD service monitoring.	Check the health state of the authentication service. Check if the sysmonitor is running. Perform the sysmonitor troubleshooting procedures to understand why the status cannot be collected.
wnbd_down	STATE_CHANGE	ERROR	Winbindd service is not functioning.	The winbindd authentication service is not functioning.	The winbindd authentication service is not functioning.	Verify the authentication configuration. Verify the connection with Active Directory server. Try to restart the winbindd service manually using the **systemctl restart gpfs-winbind** command. If the restart is unsuccessful, perform the winbindd troubleshooting procedures.
wnbd_restart	INFO	INFO	Winbindd service is not functioning. Trying to start it.	Attempt to start the winbindd service.	The winbindd process was not functioning.	N/A
wnbd_up	STATE_CHANGE	INFO	Winbindd process is now functioning.	The winbindd authentication service is operational.		N/A
wnbd_warn	INFO	WARNING	Winbindd process monitoring returned unknown result.	The winbindd authentication process monitoring returned unknown result.	An internal error occurred while monitoring the winbindd authentication process.	Check the health state of the authentication service. Check if the sysmonitor is running. Perform the sysmonitor troubleshooting procedures to understand why the status cannot be collected.
yp_down	STATE_CHANGE	ERROR	Ypbind process is not functioning.	The ypbind process is not functioning.	The ypbind authentication service is not functioning.	Verify the authentication configuration. Verify the connection with authentication server. Try to restart ypbind service manually using the **systemctl restart ypbind** command. If the restart is unsuccessful, perform the ypbind troubleshooting procedures.
yp_restart	INFO	INFO	Ypbind process is not functioning. Trying to start it.	Attempt to start the ypbind process.	The ypbind process is not functioning.	N/A
yp_up	STATE_CHANGE	INFO	Ypbind process is now functioning.	The ypbind service is operational.		N/A

Table 65. Events for the AUTH component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
yp_warn	INFO	WARNING	Ypbind process monitoring returned unknown result	The ypbind process monitoring returned unknown result.	An internal error occurred while monitoring the ypbind service.	Check the health state of the authentication service. Check if the sysmonitor is running. Perform the sysmonitor troubleshooting procedures to understand why the status cannot be collected.

Block events

The following table lists the events that are created for the *Block* component.

Table 66. Events for the Block component

Event	Event Type	Severity	Message	Description	Cause	User Action
block_disable	INFO_EXTERNAL	INFO	Block service was disabled.	The block service was disabled on this node. Disabling a service means that all configuration files are also removed. This is different from stopping service that is running.	The block service was disabled.	N/A
block_enable	INFO_EXTERNAL	INFO	Block service was enabled.	The block service was enabled on this node. Enabling a protocol service means that all the required configuration files are also automatically installed with the current valid configuration settings.	The block service was enabled.	N/A
start_block_service	INFO_EXTERNAL	INFO	Block service was started.	The block service was started.	The block service was started.	N/A
stop_block_service	INFO_EXTERNAL	INFO	Block service was stopped.	The block service was stopped.	The block service was stopped.	N/A
scst_down	STATE_CHANGE	ERROR	iscsi-scstd process is not running.	The iscsi-scstd process is not running.	The iscsi-scstd process is not running.	Stop and start the block service. This will attempt to start the iscsi-scstd process also. The monitor attempts this restart several times. In case of a permanent failure, try the **systemctl restart scst** command to restart it manually.

Table 66. Events for the Block component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
scst_up	STATE_CHANGE	INFO	iscsi-scstd process is running.	The scsi-scstd process is running.	The scsi-scstd process is running.	N/A
scst_warn	INFO	WARNING	iscsi-scstd process monitoring returned unknown result.	The iscsi-scstd process monitoring returned an unknown result.	The iscsi-scstd process monitoring returned an unknown result.	Check the health state of the block service and restart, if necessary.

CES network events

The following table lists the events that are created for the *CES Network* component.

Table 67. Events for the CES Network component

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_bond_down	STATE_CHANGE	ERROR	All slaves of the CES-network bond {0} are down.	All slaves of the CES-network bond are down.	All slaves of this network bond are down.	Check the bonding configuration, network configuration, and cabling of all slaves of the bond.
ces_bond_degraded	STATE_CHANGE	INFO	Some slaves of the CES-network bond {0} are down.	Some of the CES-network bond parts are malfunctioning.	Some slaves of the bond are not functioning properly.	Check bonding configuration, network configuration, and cabling of the malfunctioning slaves of the bond.
ces_bond_up	STATE_CHANGE	INFO	All slaves of the CES bond {0} are working as expected.	This CES bond is functioning properly.	All slaves of this network bond are functioning properly.	N/A
ces_disable_node_network	INFO	INFO	Network was disabled.	Network is disabled.	Informational message. Clean up after a 'mmchnode --ces-disable' command. Disabling CES service on the node disables the network configuration.	N/A
ces_enable_node_network	INFO	INFO	Network was enabled.	The network configuration is enabled when CES service is enabled by using the **mmchnode --ces-enable** command.	Enabling CES service on the node also enables the network services.	N/A

Table 67. Events for the CES Network component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_many_tx_errors	STATE_CHANGE	ERROR	CES NIC {0} reported many TX errors since the last monitoring cycle.	The CES-related NIC reported many TX errors since the last monitoring cycle.	The `/proc/net/dev` lists much more TX errors for this adapter since the last monitoring cycle.	Check cable contacts or try a different cable. Refer the /proc/net/dev folder to find out TX errors reported for this adapter since the last monitoring cycle.
ces_network_connectivity_up	STATE_CHANGE	INFO	CES NIC {0} can connect to the gateway.	A CES-related NIC can connect to the gateway.	The gateway responds to the sent connections-checking packets.	N/A
ces_network_connectivity_down	STATE_CHANGE	ERROR	CES NIC {0} can not connect to the gateway	This CES-related NIC can not connect to the gateway	The gateway does not respond to the sent connections-checking packets.	Check the network configuration of the network adapter, the path to the gateway, and the gateway itself.
ces_network_down	STATE_CHANGE	ERROR	CES NIC {0} is down.	This CES-related network adapter is down.	This network adapter is disabled.	Enable the network adapter and if the problem persists, verify the system logs for more details.
ces_network_found	INFO	INFO	A new CES-related NIC {0} is detected.	A new CES-related network adapter is detected.	The output of the **`ip a`** command lists a new NIC.	N/A
ces_network_ips_down	STATE_CHANGE	WARNING	No CES IPs were assigned to this node.	No CES IPs were assigned to any network adapter of this node.	No network adapters have the CES-relevant IPs, which makes the node unavailable for the CES clients.	If CES has a FAILED status, analyze the reason for this failure. If the CES pool for this node does not have enough IPs, extend the pool.
ces_network_ips_up	STATE_CHANGE	INFO	CES-relevant IPs served by NICs are detected.	CES-relevant IPs are served by network adapters. This makes the node available for the CES clients.	At least one CES-relevant IP is assigned to a network adapter.	N/A

Table 67. Events for the CES Network component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_network_ips_not_assignable	STATE_CHANGE	ERROR	No NICs are set up for CES.	No network adapters are properly configured for CES.	There are no network adapters with a static IP, matching any of the IPs from the CES pool.	Setup the static IPs and netmasks of the CES NICs in the network interface configuration scripts, or add the new matching CES IPs to the pool. The static IPs must not be aliased.
ces_network_ips_not_defined	STATE_CHANGE	WARNING	No CES IP addresses have been defined	No CES IP addresses have been defined. Use the **mmces address add** command to add CES IP addresses.	No CES IP defined but at least one CES IP is needed	Use the **mmces address add** command to add CES IP addresses. Check the group membership of IP addresses and nodes.
ces_network_affine_ips_not_defined	STATE_CHANGE	WARNING	No CES IP addresses can be applied on this node. Check group membership of node and IP addresses	No CES IP addresses found which could be hosted on this node. IPs should be in the global pool or in a group where this node is a member.	No valid CES IP found which could be hosted on this node	Use the **mmces address add** command to add CES IP addresses either to the global pool or to a group where this node is a member.
ces_network_link_down	STATE_CHANGE	ERROR	Physical link of the CES NIC {0} is down.	The physical link of this CES-related network adapter is down.	The flag LOWER_UP is not set for this NIC in the output of the **ip a** command.	Check the cabling of this network adapter.
ces_network_link_up	STATE_CHANGE	INFO	Physical link of the CES NIC {0} is up.	The physical link of this CES-related network adapter is up.	The flag LOWER_UP is set for this NIC in the output of the **ip a** command.	N/A
ces_network_up	STATE_CHANGE	INFO	CES NIC {0} is up.	This CES-related network adapter is up.	This network adapter is enabled.	N/A
ces_network_vanished	INFO	INFO	CES NIC {0} could not be detected.	One of CES-related network adapters could not be detected.	One of the previously monitored NICs is not listed in the output of the **ip a** command.	N/A

Table 67. Events for the CES Network component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_no_tx_errors	STATE_CHANGE	INFO	CES NIC {0} had no or an insignificant number of TX errors.	A CES-related NIC had no or an insignificant number of TX errors.	The `/proc/net/dev` folder lists no or an insignificant number of TX errors for this adapter since the last monitoring cycle.	N/A
ces_startup_network	INFO	INFO	CES network service was started.	Information that the CES network has started.	CES network IPs are started.	N/A
handle_network_problem_info	INFO	INFO	Handle network problem - Problem: {0}, Argument: {1}	Information about network related reconfigurations. This can be enable or disable IPs and assign or unassign IPs.	A change in the network configuration.	N/A
move_cesip_from	INFO	INFO	Address {0} is moved from this node to node {1}.	CES IP address is moved from the current node to another node.	Rebalancing of CES IP addresses.	N/A
move_cesips_info	INFO	INFO	A move request for IP addresses is performed.	In case of node failures, CES IP addresses can be moved from one node to one or more other nodes. This message is logged on a node that is observing the affected node; not necessarily on any affected node itself.	A CES IP movement was detected.	N/A
move_cesip_to	INFO	INFO	Address {0} is moved from node {1} to this node.	A CES IP address is moved from another node to the current node.	Rebalancing of CES IP addresses.	N/A

CESIP events

The following table lists the events that are created for the *CESIP* component.

Table 68. Events for the CESIP component

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_ips_assigned	STATE_CHANGE	INFO	All {0} expected CES IPs are assigned.	All declared CES IPs are assigned.	There are no unassigned CES IP addresses.	N/A
ces_ips_unassigned	STATE_CHANGE	WARNING	{0} of {1} declared CES IPs are unassigned.	Not all of the declared CES IPs are assigned	There are unassigned CES IP addresses.	Check configuration of network interfaces. Run the **`mmces address list`** command for details.

Table 68. Events for the CESIP component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ces_ips_all_unassigned	STATE_CHANGE	WARNING	All {0} declared CES IPs are unassigned.	All of the declared CES IPs are unassigned.	All declared CES IP addresses are unassigned.	Check configuration of network interfaces. Run the **mmces address list** command for details.
ces_ips_warn	INFO	WARNING	The Spectrum Scale CES IP assignment monitor could not be executed. This could be a timeout issue.	Check of the CES IP assignment state returned an unknown result. This might be a temporary issue, like a timeout during the check procedure.	The CES IP assignment state could not be determined due to a problem.	Find potential issues for this kind of failure in the `/var/adm/ras/mmsysmonitor.log` file.

Cluster state events

The following table lists the events that are created for the *Cluster state* component.

Table 69. Events for the cluster state component

Event	Event Type	Severity	Message	Description	Cause	User Action
cluster_state_manager_reset	INFO	INFO	Clear memory of cluster state manager for this node.	A reset request for the monitor state manager was received.	A reset request for the monitor state manager was received.	N/A
cluster_state_manager_resend	STATE_CHANGE	INFO	The CSM requests resending all information.	The CSM requests resending all information.	The CSM is missing information about this node	N/A
heartbeat	STATE_CHANGE	INFO	Node {0} sent a heartbeat.	The node is alive.	The cluster node sent a heartbeat to the CSM.	N/A
heartbeat_missing	STATE_CHANGE	WARNING	CES is missing a heartbeat from the node {0}.	CES is missing a heartbeat from the node.	The cluster node did not sent a heartbeat to the CSM.	Check network connectivity of the node. Check if sysmonitor is running there.
node_suspended	STATE_CHANGE	INFO	Node {0} is suspended.	The node is suspended.	The cluster node is now suspended.	Run the **mmces node resume** to stop the node from being suspended.
node_resumed	STATE_CHANGE	INFO	Node {0} is not suspended anymore.	The node is resumed after being suspended.	The cluster node was resumed after being suspended.	N/A
service_added	INFO	INFO	On the node {0} the {1} monitor was started.	A new monitor was started by Sysmonitor.	A new monitor was started.	N/A
service_removed	INFO	INFO	On the node {0} the {1} monitor was removed.	A monitor was removed by Sysmonitor.	A monitor was removed.	N/A.

Table 69. Events for the cluster state component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
service_running	STATE_CHANGE	INFO	The service {0} is running on node {1}.	The service is not stopped or disabled anymore.	The service is not stopped or disabled anymore.	N/A
service_stopped	STATE_CHANGE	INFO	The service {0} is stopped on node {1}.	The service is stopped.	The service was stopped.	Run 'mmces service start <service>' to start the service.
service_disabled	STATE_CHANGE	INFO	The service {0} is disabled.	The service is disabled.	The service was disabled.	Run the **mmces service enable <service>** command to enable the service.
eventlog_cleared	INFO	INFO	On the node {0} the eventlog was cleared.	The user cleared the eventlog with the **mmhealth node eventlog --clearDB**. This also clears the events of the **mmces events list** command.	The user cleared the eventlog.	N/A
service_reset	STATE_CHANGE	INFO	The service {0} on node {1} was reconfigured, and its events were cleared.	All current service events were cleared.	The service was reconfigured.	N/A

Transparent Cloud Tiering events

The following table lists the events that are created for the *Transparent Cloud Tiering* component.

Table 70. Events for the Transparent Cloud Tiering component

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_account_active	STATE_CHANGE	INFO	Cloud provider account that is configured with Transparent cloud tiering service is active.	Cloud provider account that is configured with Transparent cloud tiering service is active.		N/A
tct_account_bad_req	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of request error.	Transparent cloud tiering is failed to connect to the cloud provider because of request error.	Bad request.	Check trace messages and error logs for further details.
tct_account_certinvalid path	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because it was unable to find valid certification path.	Transparent cloud tiering is failed to connect to the cloud provider because it was unable to find valid certification path.	Unable to find valid certificate path.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_account_connect error	STATE_CHANGE	ERROR	An error occurred while attempting to connect a socket to the cloud provider URL.	The connection was refused remotely by cloud provider.	No process is accessing the cloud provider.	Check whether the cloud provider host name and port numbers are valid.
tct_account_configerror	STATE_CHANGE	ERROR	Transparent cloud tiering refused to connect to the cloud provider.	Transparent cloud tiering refused to connect to the cloud provider.	Some of the cloud provider-dependent services are down.	Check whether the cloud provider-dependent services are up and running.
tct_account_configured	STATE_CHANGE	WARNING	Cloud provider account is configured with Transparent cloud tiering but the service is down.	Cloud provider account is configured with Transparent cloud tiering but the service is down.	Transparent cloud tiering the service is down.	Run the command **mmcloudgateway service start** command to resume the cloud gateway service.
tct_account_containe creatererror	STATE_CHANGE	ERROR	The cloud provider container creation is failed.	The cloud provider container creation is failed.	The cloud provider account might not be authorized to create container.	Check trace messages and error logs for further details. Also, check that the account create-related issues in the *Transparent Cloud Tiering issues* section of the *IBM Spectrum Scale Problem Determination Guide*.
tct_account_dbcorrupt	STATE_CHANGE	ERROR	The database of Transparent cloud tiering service is corrupted.	The database of Transparent cloud tiering service is corrupted.	Database is corrupted.	Check trace messages and error logs for further details. Use the **mmcloudgateway files rebuildDB** command to repair it.
tct_account_direrror	STATE_CHANGE	ERROR	Transparent cloud tiering failed because one of its internal directories is not found.	Transparent cloud tiering failed because one of its internal directories is not found.	Transparent cloud tiering service internal directory is missing.	Check trace messages and error logs for further details.
tct_account_invalidurl	STATE_CHANGE	ERROR	Cloud provider account URL is not valid.	The reason could be because of HTTP 404 Not Found error.	The reason could be because of HTTP 404 Not Found error.	Check whether the cloud provider URL is valid.
tct_account_ invalidcredentials	STATE_CHANGE	ERROR	The cloud provider account credentials are invalid.	The Transparent cloud tiering service failed to connect to the cloud provider because the authentication failed.	Cloud provider account credentials either changed or are expired.	Run the **mmcloudgateway account update** command to change the cloud provider account password.
tct_account_ malformedurl	STATE_CHANGE	ERROR	Cloud provider account URL is malformed	Cloud provider account URL is malformed.	Malformed cloud provider URL.	Check whether the cloud provider URL is valid.
tct_account_ manyretries	INFO	WARNING	Transparent cloud tiering service is having too many retries internally.	Transparent cloud tiering service is having too many retries internally.	The Transparent cloud tiering service might be having connectivity issues with the cloud provider.	Check trace messages and error logs for further details.
tct_account_noroute	STATE_CHANGE	ERROR	The response from cloud provider is invalid.	The response from cloud provider is invalid.	The cloud provider URL return response code -1.	Check whether the cloud provider URL is accessible.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_account_not configured	STATE_CHANGE	WARNING	Transparent cloud tiering is not configured with cloud provider account.	The Transparent cloud tiering is not configured with cloud provider account.	The Transparent cloud tiering is installed but account is not configured or deleted.	Run the **mmcloudgateway account create** command to create the cloud provider account.
tct_account_precond error	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of precondition failed error.	Transparent cloud tiering is failed to connect to the cloud provider because of precondition failed error.	Cloud provider URL returned HTTP 412 Precondition Failed.	Check trace messages and error logs for further details.
tct_account_rkm_down	STATE_CHANGE	ERROR	The remote key manager configured for Transparent cloud tiering is not accessible.	The remote key manager that is configured for Transparent cloud tiering is not accessible.	The Transparent cloud tiering is failed to connect to IBM Security Key Lifecycle Manager.	Check trace messages and error logs for further details.
tct_account_lkm_down	STATE_CHANGE	ERROR	The local key manager configured for Transparent cloud tiering is either not found or corrupted.	The local key manager configured for Transparent cloud tiering is either not found or corrupted.	Local key manager not found or corrupted.	Check trace messages and error logs for further details.
tct_account_servererror	STATE_CHANGE	ERROR	Transparent cloud tiering service is failed to connect to the cloud provider because of cloud provider service unavailability error.	Transparent cloud tiering service is failed to connect to the cloud provider because of cloud provider server error or container size has reached max storage limit.	Cloud provider returned HTTP 503 Server Error.	Check trace messages and error logs for further details.
tct_account_socket timeout	STATE_CHANGE	ERROR	Timeout has occurred on a socket while connecting to the cloud provider.	Timeout has occurred on a socket while connecting to the cloud provider.	Network connection problem.	Check trace messages and the error log for further details. Check whether the network connection is valid.
tct_account_sslbadcert	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of bad certificate.	Transparent cloud tiering is failed to connect to the cloud provider because of bad certificate.	Bad SSL certificate.	Check trace messages and error logs for further details.
tct_account_sslcerterror	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of the untrusted server certificate chain.	Transparent cloud tiering is failed to connect to the cloud provider because of untrusted server certificate chain.	Untrusted server certificate chain error.	Check trace messages and error logs for further details.
tct_account_sslerror	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of error the SSL subsystem.	Transparent cloud tiering is failed to connect to the cloud provider because of error the SSL subsystem.	Error in SSL subsystem.	Check trace messages and error logs for further details.
tct_account_ssl handshakeerror	STATE_CHANGE	ERROR	The cloud account status is failed due to unknown SSL handshake error.	The cloud account status is failed due to unknown SSL handshake error.	Transparent cloud tiering and cloud provider could not negotiate the desired level of security.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_account_ssl handshakefailed	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because they could not negotiate the desired level of security.	Transparent cloud tiering is failed to connect to the cloud provider because they could not negotiate the desired level of security.	Transparent cloud tiering and cloud provider server could not negotiate the desired level of security.	Check trace messages and error logs for further details.
tct_account_ssl invalidalgo	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of invalid SSL algorithm parameters.	Transparent cloud tiering is failed to connect to the cloud provider because of invalid or inappropriate SSL algorithm parameters.	Invalid or inappropriate SSL algorithm parameters.	Check trace messages and error logs for further details.
tct_account_ssl invalidpadding	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of invalid SSL padding.	Transparent cloud tiering is failed to connect to the cloud provider because of invalid SSL padding.	Invalid SSL padding.	Check trace messages and error logs for further details.
tct_account_ssl nottrustedcert	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of not trusted server certificate.	Transparent cloud tiering is failed to connect to the cloud provider because of not trusted server certificate.	Cloud provider server SSL certificate is not trusted.	Check trace messages and error logs for further details.
tct_account_ssl unrecognizedmsg	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of unrecognized SSL message.	Transparent cloud tiering is failed to connect to the cloud provider because of unrecognized SSL message.	Unrecognized SSL message.	Check trace messages and error logs for further details.
tct_account_sslnocert	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of no available certificate.	Transparent cloud tiering is failed to connect to the cloud provider because of no available certificate.	No available certificate.	Check trace messages and error logs for further details.
tct_account_ssl scoketclosed	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because remote host closed connection during handshake.	Transparent cloud tiering is failed to connect to the cloud provider because remote host closed connection during handshake.	Remote host closed connection during handshake.	Check trace messages and error logs for further details.
tct_account_sslkeyerror	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect cloud provider because of bad SSL key.	Transparent cloud tiering is failed to connect cloud provider because of bad SSL key or misconfiguration.	Bad SSL key or misconfiguration.	Check trace messages and error logs for further details.
tct_account_ sslpeererror	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because its identity has not been verified.	Transparent cloud tiering is failed to connect to the cloud provider because its identity is not verified.	Cloud provider identity is not verified.	Check trace messages and error logs for further details.
tct_account_ssl protocolerror	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect cloud provider because of error in the operation of the SSL protocol.	Transparent cloud tiering is failed to connect cloud provider because of error in the operation of the SSL protocol.	SSL protocol implementation error.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_account_ssl unknowncert	STATE_CHANGE	ERROR	Transparent cloud tiering is failed to connect to the cloud provider because of unknown certificate.	Transparent cloud tiering is failed to connect to the cloud provider because of unknown certificate.	Unknown SSL certificate.	Check trace messages and error logs for further details.
tct_account_time skewerror	STATE_CHANGE	ERROR	The time observed on the Transparent cloud tiering service node is not in sync with the time on target cloud provider.	The time observed on the Transparent cloud tiering service node is not in sync with the time on target cloud provider.	Current time stamp of Transparent cloud tiering service is not in sync with target cloud provider.	Change Transparent cloud tiering service node time stamp to be in sync with NTP server and rerun the operation.
tct_account_ unknownerror	STATE_CHANGE	ERROR	The cloud provider account is not accessible due to unknown error.	The cloud provider account is not accessible due to unknown error.	Unknown runtime exception.	Check trace messages and error logs for further details.
tct_account_ unreachable	STATE_CHANGE	ERROR	Cloud provider account URL is not reachable.	The cloud provider's URL is unreachable because either it is down or network issues.	The cloud provider URL is not reachable.	Check trace messages and the error log for further details. Check the DNS settings.
tct_fs_configured	STATE_CHANGE	INFO	The Transparent cloud tiering is configured with file system.	The Transparent cloud tiering is configured with file system.		N/A
tct_fs_notconfigured	STATE_CHANGE	WARNING	The Transparent cloud tiering is not configured with file system.	The Transparent cloud tiering is not configured with file system.	The Transparent cloud tiering is installed but file system is not configured or deleted.	Run the command **mmcloudgateway filesystem create** to configure the file system.
tct_fs_running_out _space	INFO	WARNING	Available disk space is {0}.	Filesystem where TCT got installed is running out of space.	Filesystem where TCT got installed is running out of space.	Free up disk space on the file system where TCT is installed
tct_service_down	STATE_CHANGE	ERROR	Transparent cloud tiering service is down.	The Transparent cloud tiering service is down and could not be started.	The **mmcloudgateway service status** command returns 'Stopped' as the status of the Transparent cloud tiering service.	Run the command **mmcloudgateway service start** to start the cloud gateway service.
tct_service_suspended	STATE_CHANGE	WARNING	Transparent cloud tiering service is suspended.	The Transparent cloud tiering service is suspended manually.	The **mmcloudgateway service status** command returns 'Suspended' as the status of the Transparent cloud tiering service.	Run the **mmcloudgateway service start** command to resume the Transparent cloud tiering service.
tct_service_up	STATE_CHANGE	INFO	Transparent cloud tiering service is up and running.	The Transparent cloud tiering service is up and running.		N/A
tct_service_warn	INFO	WARNING	Transparent cloud tiering monitoring returned unknown result.	The Transparent cloud tiering check returned unknown result.		Perform troubleshooting procedures.
tct_service_restart	INFO	WARNING	The Transparent cloud tiering service failed. Trying to recover.	Attempt to restart the Transparent cloud tiering process.	A problem with the Transparent cloud tiering process is detected.	N/A
tct_service_not configured	STATE_CHANGE	WARNING	Transparent cloud tiering is not configured.	The Transparent cloud tiering service was either not configured or never started.	TheTransparent cloud tiering service was either not configured or never started.	Set up the Transparent cloud tiering and start its service.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_csap_online	STATE_CHANGE	INFO	Cloud storage access point configured with Transparent Cloud Tiering service is active. CSAP/Container pair set: {id}.	Cloud storage access point configured with Transparent Cloud Tiering service is active.		N/A
tct_csap_unreachable	STATE_CHANGE	ERROR	Cloud storage access point URL is not reachable. CSAP/Container pair set: {id}.	The cloud storage access point URL is unreachable due to either it is down or network issues.	The cloud storage access point URL is not reachable	Check trace messages and the error log for further details. Check the DNS settings.
tct_csap_invalidurl	STATE_CHANGE	ERROR	Cloud storage access point URL is not valid. CSAP/Container pair set: {id}.	The reason could be because of HTTP 404 Not Found error	The reason could be because of HTTP 404 Not Found error	Check the cloud provider URL is valid
tct_csap_malformedurl	STATE_CHANGE	ERROR	Cloud storage access point URL is malformed. CSAP/Container pair set: {id}.	The cloud storage access point URL is malformed.	Malformed cloud provider URL.	Check the cloud provider URL is valid.
tct_csap_noroute	STATE_CHANGE	ERROR	The response from cloud storage access point is invalid. CSAP/Container pair set: {id}.	The response from cloud storage access point is invalid	The cloud storage access point URL return response code -1.	Check the cloud storage access point URL is accessible.
tct_csap_connecterror	STATE_CHANGE	ERROR	An error occurred while attempting to connect a socket to cloud storage access point URL. CSAP/Container pair set: {id}.	The connection was refused remotely by cloud storage access point.	No process is listening on cloud storage access point address.	Check cloud storage access point hostname and port numbers are valid.
tct_csap_sockettimeout	STATE_CHANGE	ERROR	Timeout has occurred on a socket while connecting cloud storage access point URL. CSAP/Container pair set: {id}.	Timeout has occurred on a socket while connecting cloud storage access point URL.	Network connection problem.	Check the trace messages and the error log for further details. Check if the network connection is valid.
tct_csap_configerror	STATE_CHANGE	ERROR	Transparent Cloud Tiering refused to connect to cloud storage access point. CSAP/Container pair set: {id}.	Transparent Cloud Tiering refused to connect to cloud storage access point.	Some of the cloud provider dependent services are down.	Check cloud provider dependent services are up and running.
tct_csap_ invalidcredentials	STATE_CHANGE	ERROR	The cloud storage access point account {0} credentials are invalid. CSAP/Container pair set: {id}.	The Transparent Cloud Tiering service failed to connect cloud storage access point because the authentication is failed.	Cloud storage access point account credentials either changed or expired.	Run the command 'mmcloudgateway account update' to change the cloud provider account password.
tct_network_ interface_down	STATE_CHANGE	ERROR	The network of Transparent Cloud Tiering node is down. CSAP/Container pair set: {id}.	The network of Transparent Cloud Tiering node is down.	Network connection problem.	Check trace messages and error logs for further details. Check network connection is valid.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_csap_ sslhandshakeerror	STATE_CHANGE	ERROR	The cloud storage access point status is failed due to unknown SSL handshake error. CSAP/Container pair set: {id}.	The cloud storage access point status is failed due to unknown SSL handshake error.	TCT and cloud storage access point could not negotiate the desired level of security.	Check trace messages and error logs for further details.
tct_csap_sslcerterror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of untrusted server certificate chain. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of untrusted server certificate chain.	Untrusted server certificate chain error.	Check trace messages and error logs for further details.
tct_csap_ sslscoketclosed	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because remote host closed connection during handshake. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because remote host closed connection during handshake	Remote host closed connection during handshake.	Check trace messages and error logs for further details.
tct_csap_sslbadcert	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of a bad certificate. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of a bad certificate.	Bad SSL certificate.	Check the trace messages and error logs for further details.
tct_csap_ certinvalidpath	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because it could not find a valid certification path. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because it could not find a valid certification path.	Unable to find a valid certificate path.	Check the trace messages and error logs for further details.
tct_csap_ sslhandshakefailed	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because they could not negotiate the desired level of security. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because they could not negotiate the desired level of security.	TCT and cloud storage access point could not negotiate the desired level of security.	Check the trace messages and error logs for further details.
tct_csap_ sslunknowncert	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of unknown certificate. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of unknown certificate.	Unknown SSL certificate.	Check the trace messages and error logs for further details
tct_csap_sslkeyerror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of bad SSL key. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of bad SSL key or misconfiguration	Bad SSL key or misconfiguration.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_csap_sslpeererror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because its identity has not been verified. CSAP/Container pair set: {id}".	Transparent Cloud Tiering failed to connect cloud storage access point because its identity has not been verified.	Cloud provider identity has not been verified.	Check trace messages and error logs for further details.
tct_csap_ sslprotocolerror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of error in the operation of the SSL protocol. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of error in the operation of the SSL protocol.	SSL protocol implementation error.	Check trace messages and error logs for further details.
tct_csap_sslerror	STATE_.CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of error the SSL subsystem. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of error the SSL subsystem.	Error in SSL subsystem.	Check trace messages and error logs for further details.
tct_csap_sslnocert	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of no available certificate. CSAP/Container pair set: {id}	Transparent Cloud Tiering failed to connect cloud storage access point because of no available certificate.	No available certificate.	Check trace messages and error logs for further details.
tct_csap_ sslnottrustedcert	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of not trusted server certificate. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of not trusted server certificate.	Cloud storage access point server SSL certificate is not trusted.	Check trace messages and error logs for further details.
tct_csap_sslinvalidalgo	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of invalid SSL algorithm parameters. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of invalid or inappropriate SSL algorithm parameters.	Invalid or inappropriate SSL algorithm parameters.	Check trace messages and error logs for further details.
tct_csap_ sslinvalidpadding	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of invalid SSL padding. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of invalid SSL padding.	Invalid SSL padding.	Check trace messages and error logs for further details.
tct_csap_ sslunrecognizedmsgng	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of unrecognized SSL message. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of unrecognized SSL message. CSAP/Container pair set: {id}.	Unrecognized SSL message.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_csap_bad_req	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of request error. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of request error.	Bad request.	Check trace messages and error logs for further details.
tct_csap_preconderror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of precondition failed error. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of precondition failed error.	Cloud storage access point URL returned HTTP 412 Precondition Failed.	Check trace messages and error logs for further details.
tct_csap_unknownerror	STATE_CHANGE	ERROR	The cloud storage access point account is not accessible due to uknown error. CSAP/Container pair set: {id}.	The cloud storage access point account is not accessible due to uknown error.	Unknown Runtime exception.	Check trace messages and error logs for further details.
tct_container_ createererror	STATE_CHANGE	ERROR	The cloud provider container creation is failed. CSAP/Container pair set: {id}.	The cloud provider container creation is failed.	The cloud provider account may not be authorized to create container.	Check trace messages and error logs for further details.
tct_container_ alreadyexists	STATE_CHANGE	ERROR	The cloud provider container creation is failed as it already exists. CSAP/Container pair set: {id}.	The cloud provider container creation is failed as it already exists.	The cloud provider container already exists.	Check trace messages and error logs for further details.
tct_container_ limitexceeded	STATE_CHANGE	ERROR	The cloud provider container creation is failed as it exceeded the max limit. CSAP/Container pair set: {id}.	The cloud provider container creation is failed as it exceeded the max limit.	The cloud provider containers exceeded the max limit.	Check trace messages and error logs for further details.
tct_container_notexists	STATE_CHANGE	ERROR	The cloud provider container does not exist. CSAP/Container pair set: {id}.	The cloud provider container does not exist.	The cloud provider container does not exist.	Check cloud provider if the container exists.
tct_csap_timeskewerror	STATE_CHANGE	ERROR	The time observed on the Transparent Cloud Tiering service node is not in sync with the time on target cloud storage access point. CSAP/Container pair set: {id}.	The time observed on the Transparent Cloud Tiering service node is not in sync with the time on target cloud storage access point.	Transparent Cloud Tiering service node current timestamp is not in sync with target cloud storage access point.	Change Transparent Cloud Tiering service node timestamp to be in sync with NTP server and re-run the operation.
tct_csap_servererror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed to connect cloud storage access point because of cloud storage access point service unavailability error. CSAP/Container pair set:{id}.	Transparent Cloud Tiering failed to connect cloud storage access point because of cloud storage access point server error or container size has reached max storage limit.	Cloud storage access point returned HTTP 503 Server Error.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_internal_direrror	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed because one of its internal directory is not found. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed because one of its internal directory is not found.	Transparent Cloud Tiering service internal directory is missing.	Check trace messages and error logs for further details.
tct_resourcefile_ notfound	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed because resource address file is not found. CSAP/Container pair set: {id}.	Transparent Cloud Tiering failed because resource address file is not found.	Transparent Cloud Tiering failed because resource address file is not found.	Check trace messages and error logs for further details.
tct_csap_dbcorrupt	STATE_CHANGE	ERROR	The database of Transparent Cloud Tiering service is corrupted. CSAP/Container pair set: {id}.	The database of Transparent Cloud Tiering service is corrupted.	Database is corrupted.	Run the command 'mmcloudgateway files rebuildDB' to rebuild the database.
tct_csap_rkm_down	STATE_CHANGE	ERROR	The remote key manager configured for Transparent Cloud Tiering is not accessible. CSAP/Container pair set: {id}.	The remote key manager configured for Transparent Cloud Tiering is not accessible.	The Transparent Cloud Tiering is failed to connect to IBM Security Key Lifecycle Manager.	Check trace messages and error logs for further details.
tct_csap_lkm_down	STATE_CHANGE	ERROR	The local key manager configured for Transparent Cloud Tiering is either not found or corrupted. CSAP/Container pair set: {id}.	The local key manager configured for Transparent Cloud Tiering is either not found or corrupted.	Local key manager not found or corrupted.	Check trace messages and error logs for further details.
tct_csap_forbidden	STATE_CHANGE	ERROR	Cloud storage access point failed with authorization error. CSAP/Container pair set: {id}.	The reason could be because of HTTP 403 Forbidden.	The reason could be because of HTTP 403 Forbidden.	Check the authorization configurations on the cloud provider.
tct_csap_access_denied	STATE_CHANGE	ERROR	Cloud storage access point failed with authorization error. CSAP/Container pair set: {id}.	Access denied due to authorization error.	Access denied due to authorization error.	Check the authorization configurations on the cloud provider.
tct_fs_corrupted	STATE_CHANGE	ERROR	The filesystem {0} of Transparent Cloud Tiering service is corrupted. CSAP/Container pair set: {id}.	The filesystem of Transparent Cloud Tiering service is corrupted..	Filesystem is corrupted	Check trace messages and error logs for further details.
tct_dir_corrupted	STATE_CHANGE	ERROR	The directory of Transparent Cloud Tiering service is corrupted. CSAP/Container pair set: {id}.	The directory of Transparent Cloud Tiering service is corrupted.	Directory is corrupted.	Check trace messages and error logs for further details.
tct_km_error	STATE_CHANGE	ERROR	The key manager configured for Transparent Cloud Tiering is either not found or corrupted. CSAP/Container pair set: {id}.	The key manager configured for Transparent Cloud Tiering is either not found or corrupted.	Key manager not found or corrupted.	Check trace messages and error logs for further details.

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_rootdir_notfound	STATE_CHANGE	ERROR	Transparent Cloud Tiering failed because its container pair root directory not found. Container pair set: {id}.	Transparent Cloud Tiering failed because its container pair root directory not found.	Transparent Cloud Tiering failed because its container pair root directory not found.	Check trace messages and error logs for further details.
tct_csap_ toomanyretries	INFO	WARNING	Transparent Cloud Tiering service is having too many retries internally. CSAP/Container pair set: {id}.	Transparent Cloud Tiering service is having too many retries internally.	Probable reason could be Transparent Cloud Tiering service has connectivity issues with cloud provider.	Check trace messages and error logs for further details.
tct_csap_found	INFO_ADD_ENTITY	INFO	CSAP/container pair {0} was found.	A new CSAP/container pair was found	A new CSAP/container pair, which is relevant for the Spectrum Scale monitoring, is listed by the **mmcloudgateway service list** command.	N/A
tct_cs_found	INFO_ADD_ENTITY	INFO	Cloud services {0} was found.	A new cloud service was found.	A new cloud service is listed by the **mmcloudgateway service status** command.	N/A
tct_cs_vanished	INFO_DELETE_ENTITY	INFO	Cloud services is not available.	One of Cloud services can not be detected anymore.	One of the previously monitored Cloud services is not listed by the **mmcloudgateway service status** command anymore, possibly because TCT service is in suspended state.	N/A
tct_cs_enabled	STATE_CHANGE	INFO	Cloud services {id} is enabled.	Cloud services is enabled for cloud operations.	Cloud services has been enabled by administrator.	N/A
tct_cs_disabled	STATE_CHANGE	WARNING	Cloud services {id} is disabled.	Cloud services is disabled.	Cloud services has been disabled by administrator.	N/A
tct_account_network_ down	STATE_CHANGE	ERROR	The network of Transparent Cloud Tiering node is down.	The network of Transparent Cloud Tiering node is down.	Network connection problem.	Check the trace messages and the error logs for further details. Check if the network connection is valid.
tct_csap_removed	INFO_DELETE_ENTITY	INFO	CSAP/container pair {0} is not available.	A CSAP/container pairs can not be detected anymore.	One of the previously monitored CSAP/container pairs is not listed by the **mmcloudgateway service list** command anymore, possibly because TCT service is in suspended state.	N/A
tct_csap_base_found	INFO_ADD_ENTITY	INFO	CSAP {0} was found.	A new CSAP was found	A new CSAP is listed by the **mmcloudgateway service list** command.	N/A

Table 70. Events for the Transparent Cloud Tiering component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
tct_csap_base_removed	INFO_DELETE_ENTITY	INFO	CSAP {0} was deleted or converted to a CSAP/container pair.	A CSAP was deleted.	One of the previously monitored CSAP is not listed by **mmcloudgateway service list** command anymore.	N/A

Disk events

The following table lists the events that are created for the *Disk* component.

Table 71. Events for the Disk component

Event	Event Type	Severity	Message	Description	Cause	User Action
disk_down	STATE_CHANGE	WARNING	Disk {0} is reported as not up.	A disk is reported as down.	This can indicate a hardware issue.	If the down state is unexpected, then refer to the *Disk issues* section in the *IBM Spectrum Scale Troubleshooting Guide*.
disk_up	STATE_CHANGE	INFO	Disk {0} is up.	Disk is up.	A disk was detected in up state.	N/A
disk_found	INFO	INFO	The disk {0} was found.	A disk was detected.	A disk was detected.	N/A
disk_vanished	INFO	INFO	The disk {0} has vanished.	A declared disk was not detected.	A disk is not in use for an IBM Spectrum Scale filesystem. This could be a valid situation.	N/A
disc_recovering	STATE_CHANGE	WARNING	Disk {0} is reported as recovering	A disk is in recovering state	A disk is in recovering state	If the recovering state is unexpected, then refer to the section Disk issues in the Troubleshooting guide
disc_unrecovered	STATE_CHANGE	WARNING	Disk {0} is reported as unrecovered	A disk is in unrecovered state	A disk is in unrecovered state. The metadata scan might have failed.	If the unrecovered state is unexpected, then refer to the section Disk issues in the Troubleshooting guide

File system events

The following table lists the events that are created for the *File System* component.

Table 72. Events for the file system component

Event	Event Type	Severity	Message	Description	Cause	User Action
filesystem_found	INFO	INFO	The file system {0} is detected.	A file system listed in the IBM Spectrum Scale configuration was detected.	N/A	N/A
filesystem_vanished	INFO	INFO	The file system {0} is not detected.	A file system listed in the IBM Spectrum Scale configuration was not detected.	A file system, which is listed as a mounted file system in the IBM Spectrum Scale configuration, is not detected. This could be valid situation that demands troubleshooting.	Issue the **mmlsmount all_ local** command to verify whether all the expected file systems are mounted.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fs_forced_unmount	STATE_CHANGE	ERROR	The file system {0} was {1} forced to unmount.	A file system was forced to unmount by IBM Spectrum Scale.	A situation like a kernel panic might have initiated the unmount process.	Check error messages and logs for further details. Also, see the *File system forced unmount* and *File system issues* topics in the IBM Spectrum Scale documentation.
fserrallocblock	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Corrupted alloc segment detected while attempting to alloc disk block.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbadaclref	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	File references invalid ACL.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrbaddirblock	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid directory block.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbaddiskaddrindex	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Bad disk index in disk address.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbaddiskaddrsector	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Bad sector number in disk address or start sector plus length is exceeding the size of the disk.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrbaddittoaddr	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid ditto address.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbadinodeorgen	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Deleted inode has a directory entry or the generation number do not match to the directory.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbadinodestatus	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Inode status is changed to *Bad*. The expected status is: *Deleted*.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrbadptrreplications	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid computed pointer replication factors.	Invalid computed pointer replication factors.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbadreplicationcounts	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid current or maximum data or metadata replication counts.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrbadxattrblock	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid extended attribute block.	A file system corruption is detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrcheckheaderfailed	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	CheckHeader returned an error.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrclonetree	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid cloned file tree structure.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrdeallocblock	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Corrupted alloc segment detected while attempting to dealloc the disk block.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrdotdotnotfound	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Unable to locate an entry.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrgennummismatch	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	The generation number entry in '..' does not match with the actual generation number of the parent directory.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinconsistentfilesetrootdir	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Inconsistent fileset or root directory. That is, fileset is in use, root dir '..' points to itself.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrinconsistentfilesetsnapshot	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Inconsistent fileset or snapshot records. That is, fileset snapList points to a SnapItem that does not exist.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinconsistentinode	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Size data in inode are inconsistent.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrindirectblock	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid indirect block header information in the inode.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrindirectionlevel	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Invalid indirection level in inode.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinodecorrupted	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	Infinite loop in the lfs layer because of a corrupted inode or directory entry.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinodenummismatch	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Msg={2}	The inode number that is found in the '..' entry does not match with the actual inode number of the parent directory.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrinvalid	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Unknown error={2}.	Unrecognized FSSTRUCT error received.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinvalidfilesetmetadata record	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Unknown error={2}.	Invalid fileset metadata record.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrinvalidsnapshotstates	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Unknown error={2}.	Invalid snapshot states. That is, more than one snapshot in an inode space is being emptied (SnapBeingDeleted One).	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fserrsnapinodemodified	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Unknown error={2}.	Inode was modified without saving old content to shadow inode file.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fserrvalidate	STATE_CHANGE	ERROR	The following error occurred for the file system {0}: ErrNo={1}, Unknown error={2}.	A file system corruption detected. Validation routine failed on a disk read.	A file system corruption detected.	Check error message and the *mmfs.log.latest* log for further details. For more information, see the *Checking and repairing a file system* and *Managing file systems.* topics in the IBM Spectrum Scale documentation. If the file system is severely damaged, the best course of action is available in the *Additional information to collect for file system corruption or MMFS_ FSSTRUCT errors* topic.
fsstruct_error	STATE_CHANGE	WARNING	The following structure error is detected in the file system {0}: Err={1} msg={2}.	A file system structure error is detected. This issue might cause different events.	A file system issue was detected.	When an fsstruct error is show in mmhealth, the customer is asked to run a filesystem check. Once the problem is solved the user needs to clear the fsstruct error from mmhealth manually by running the following command: `mmsysmonc event filesystem fsstruct_fixed <filesystem_name>`
fsstruct_fixed	STATE_CHANGE	INFO	The structure error reported for the file system {0} is marked as fixed.	A file system structure error is marked as fixed.	A file system issue was resolved.	N/A
fs_unmount_info	INFO	INFO	The file system {0} is unmounted {1}.	A file system is unmounted.	A file system is unmounted.	N/A

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
fs_remount_mount	STATE_CHANGE _EXTERNAL	INFO	The file system {0} is mounted.	A file system is mounted.	A new or previously unmounted file system is mounted.	N/A
mounted_fs_check	STATE_CHANGE	INFO	The file system {0} is mounted.	The file system is mounted.	A file system is mounted and no mount state mismatch information detected.	N/A
stale_mount	STATE_CHANGE	ERROR	Found stale mounts for the file system {0}.	A mount state information mismatch was detected between the details reported by the **mmlsmount** command and the information that is stored in the /proc/mounts.	A file system might not be fully mounted or unmounted.	Issue the **mmlsmount all_ local** command to verify that all expected file systems are mounted.
unmounted_fs_ok	STATE_CHANGE	INFO	The file system {0} is probably needed, but not declared as automount.	An internally mounted or a declared but not mounted file system was detected.	A declared file system is not mounted.	N/A
unmounted_fs_check	STATE_CHANGE	WARNING	The filesystem {0} is probably needed, but not declared as automount.	An internally mounted or a declared but not mounted file system was detected.	A file system might not be fully mounted or unmounted.	Issue the **mmlsmount all_ local** command to verify that all expected file systems are mounted.
pool_normal	STATE_CHANGE	INFO	The pool {id[1]} of file system {id[0]} reached a normal level.	The pool reached a normal level.	The pool reached a normal level.	N/A
pool_high_error	STATE_CHANGE	ERROR	The pool {id[1]} of file system {id[0]} reached a nearly exhausted level.	The pool reached a nearly exhausted level.	The pool reached a nearly exhausted level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.
pool_high_warn	STATE_CHANGE	WARNING	The pool {id[1]} of file system {id[0]} reached a warning level.	"The pool reached a warning level.	The pool reached a warning level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.
pool_no_data	INFO	INFO	The state of pool {id[1]} in file system {id[0]} is unknown.	Could not determine fill state of the pool.	Could not determine fill state of the pool.	
pool-metadata_normal	STATE_CHANGE	INFO	The pool {id[1]} of file system {id[0]} reached a normal metadata level.	The pool reached a normal level.	The pool reached a normal level.	N/A
pool-metadata_high_error	STATE_CHANGE	ERROR	The pool {id[1]} of file system {id[0]} reached a nearly exhausted metadata level.	The pool reached a nearly exhausted level.	The pool reached a nearly exhausted level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.
pool-metadata_high_warn	STATE_CHANGE	WARNING	The pool {id[1]} of file system {id[0]} reached a warning level for metadata.	The pool reached a warning level.	The pool reached a warning level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
pool-metadata_removed	STATE_CHANGE	INFO	No usage data for pool {id[1]} in file system {id[0]}.	No pool usage data in performance monitoring.	No pool usage data in performance monitoring.	N/A
pool-metadata_no_data	STATE_CHANGE	INFO	No usage data for pool {id[1]} in file system {id[0]}.	No pool usage data in performance monitoring.	No pool usage data in performance monitoring.	N/A
pool-data_normal	STATE_CHANGE	INFO	The pool {id[1]} of file system {id[0]} reached a normal data level.	The pool reached a normal level.	The pool reached a normal level.	N/A
pool-data_high_error	STATE_CHANGE	ERROR	The pool {id[1]} of file system {id[0]} reached a nearly exhausted data level.	The pool reached a nearly exhausted level.	The pool reached a nearly exhausted level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.
pool-data_high_warn	STATE_CHANGE	WARNING	The pool {id[1]} of file system {id[0]} reached a warning level for metadata.	The pool reached a warning level.	The pool reached a warning level.	Add more capacity to pool or move data to different pool or delete data and/or snapshots.
pool-data_removed	STATE_CHANGE	INFO	No usage data for pool {id[1]} in file system {id[0]}.	No pool usage data in performance monitoring.	No pool usage data in performance monitoring.	N/A
pool-data_no_data	STATE_CHANGE	INFO	No usage data for pool {id[1]} in file system {id[0]}.	No pool usage data in performance monitoring.	No pool usage data in performance monitoring.	N/A
inode_normal	STATE_CHANGE	INFO	The inode usage of fileset {id[1]} in file system {id[0]} reached a normal level.	The inode usage in the fileset reached a normal level.	The inode usage in the fileset reached a normal level.	N/A
inode_high_error	STATE_CHANGE	ERROR	The inode usage of fileset {id[1]} in file system {id[0]} reached a nearly exhausted level.	The inode usage in the fileset reached a nearly exhausted level.	The inode usage in the fileset reached a nearly exhausted level.	Use the **mmchfileset** command to increase the inode space. You can also use the **Monitoring** > **Events** page in the IBM Spectrum Scale GUI to fix this event with the help of a directed maintenance procedure (DMP). Select the event from the list of events and select **Actions** > **Run Fix Procedure** to launch the DMP.
inode_high_warn	STATE_CHANGE	WARNING	The inode usage of fileset {id[1]} in file system {id[0]} reached a warning level.	The inode usage of fileset {id[1]} in file system {id[0]} reached a warning level.	The inode usage in the fileset reached warning level.	"Delete data."
inode_removed	STATE_CHANGE	INFO	No inode usage data for fileset {id[1]} in file system {id[0]}.	No inode usage data in performance monitoring.	No inode usage data in performance monitoring.	N/A
inode_no_data	STATE_CHANGE	INFO	No inode usage data for fileset {id[1]} in file system {id[0]}.	No inode usage data in performance monitoring.	No inode usage data in performance monitoring.	N/A

Table 72. Events for the file system component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
disk_failed_cb	INFO_EXTERNAL	INFO	Disk {0} is reported as failed. FS={1}, event={2}. Affected NSD servers are notified about the disk_down state.	A disk is reported as failed. This event also appears on the manual user actions like the **mmdeldisk** command. It shows up only on filesystem manager nodes, and triggers a **disk_down** event on all NSD nodes which serve the failed disk.	A callback reported a failing disk .	If the failure state is unexpected, then refer to the Chapter 19, "Disk issues," on page 349 section, and perform the appropriate troubleshooting procedures.

GPFS events

The following table lists the events that are created for the *GPFS* component.

Table 73. Events for the GPFS component

Event	Event Type	Severity	Message	Description	Cause	User Action
ccr_client_init_ok	STATE_CHANGE	INFO	GPFS CCR client initialization is ok {0}.	GPFS CCR client initialization is ok.	N/A	N/A
ccr_client_init_fail	STATE_CHANGE	ERROR	GPFS CCR client initialization failed Item={0}, ErrMsg={1}, Failed={2}.	GPFS CCR client initialization failed. See message for details.	The item specified in the message is either not available or corrupt.	Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** for more details.
ccr_client_init_warn	STATE_CHANGE	WARNING	GPFS CCR client initialization failed Item={0}, ErrMsg={1}, Failed={2}.	GPFS CCR client initialization failed. See message for details.	The item specified in the message is either not available or corrupt.	Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** for more details.
ccr_auth_keys_ok	STATE_CHANGE	INFO	The security file used by GPFS CCR is ok {0}.	The security file used by GPFS CCR is ok.	N/A	N/A
ccr_auth_keys_fail	STATE_CHANGE	ERROR	The security file used by GPFS CCR is corrupt Item={0}, ErrMsg={1}, Failed={2}	The security file used by GPFS CCR is corrupt. See message for details.	Either the security file is missing or corrupt.	Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** for more details.
ccr_paxos_cached_ok	STATE_CHANGE	INFO	The stored GPFS CCR state is ok {0}	The stored GPFS CCR state is ok.		N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ccr_paxos_cached_fail	STATE_CHANGE	ERROR	The stored GPFS CCR state is corrupt Item={0}, ErrMsg={1}, Failed={2}	The stored GPFS CCR state is corrupt. See message for details.	Either the stored GPFS CCR state file is corrupt or empty.	Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** for more details.
ccr_paxos_12_fail	STATE_CHANGE	ERROR	The stored GPFS CCR state is corrupt Item={0}, ErrMsg={1}, Failed={2}	The stored GPFS CCR state is corrupt. See message for details.	The stored GPFS CCR state is corrupt. See message for details.	Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** for more details.
ccr_paxos_12_ok	STATE_CHANGE	INFO	The stored GPFS CCR state is ok {0}	The stored GPFS CCR state is ok.	N/A	N/A
ccr_paxos_12_warn	STATE_CHANGE	WARNING	The stored GPFS CCR state is corrupt Item={0}, ErrMsg={1}, Failed={2}	The stored GPFS CCR state is corrupt. See message for details.	One stored GPFS state file is missing or corrupt.	No user action necessary, GPFS will repair this automatically.
ccr_local_server_ok	STATE_CHANGE	INFO	The local GPFS CCR server is reachable {0}	The local GPFS CCR server is reachable.	N/A	N/A
ccr_local_server_warn	STATE_CHANGE	WARNING	The local GPFS CCR server is not reachable Item={0}, ErrMsg={1}, Failed={2}	The local GPFS CCR server is not reachable. See message for details.	Either the local network or firewall is not configured properly or the local GPFS daemon is not responding.	Check the network and firewall configuration with regards to the used GPFS communication port (default: 1191). Restart GPFS on this node.
ccr_ip_lookup_ok	STATE_CHANGE	INFO	The IP address lookup for the GPFS CCR component is ok {0}	The IP address lookup for the GPFS CCR component is ok.	N/A	N/A
ccr_ip_lookup_warn	STATE_CHANGE	WARNING	The IP address lookup for the GPFS CCR component takes too long. Item={0}, ErrMsg={1}, Failed={2}	The IP address lookup for the GPFS CCR component takes too long, resulting in slow administration commands. See message for details.	Either the local network or the DNS is misconfigured.	Check the local network and DNS configuration.
ccr_quorum_nodes_fail	STATE_CHANGE	ERROR	A majority of the quorum nodes are not reachable over the management network Item={0},ErrMsg={1}, Failed={2}	A majority of the quorum nodes are not reachable over the management network. GPFS declares quorum loss. See message for details.	Due to the misconfiguration of network or firewall, the quorum nodes cannot communicate with each other.	Check the network and firmware (default port 1191 must not be blocked) configuration of the quorum nodes that are not reachable.
ccr_quorum_nodes_ok	STATE_CHANGE	INFO	All quorum nodes are reachable {0}	All quorum nodes are reachable.	N/A	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
ccr_quorum_nodes_warn	STATE_CHANGE	WARNING	Clustered Configuration Repository issue with Item={0}, ErrMsg={1}, Failed={2}	At least one quorum node is not reachable. See message for details.	The quorum node is not reachable due to the network or firewall misconfiguration.	Check the network and firmware (default port 1191 must not be blocked) configuration of the quorum node that is not reachable.
ccr_comm_dir_fail	STATE_CHANGE	ERROR	The files committed to the GPFS CCR are not complete or corrupt Item={0},ErrMsg={1}, Failed={2}	The files committed to the GPFS CCR are not complete or corrupt. See message for details.	The local disk might be full.	Check the local disk space and remove the unnecessary files. Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** command for more details.
ccr_comm_dir_ok	STATE_CHANGE	INFO	The files committed to the GPFS CCR are complete and intact {0}	The files committed to the GPFS CCR are complete and intact..	N/A	N/A
ccr_comm_dir_warn	STATE_CHANGE	WARNING	The files committed to the GPFS CCR are not complete or corrupt Item={0},ErrMsg={1}, Failed={2}	The files committed to the GPFS CCR are not complete or corrupt. See message for details.	The local disk might be full.	Check the local disk space and remove the unnecessary files. Recover this degraded node from a still intact node by using the **mmsdrrestore -p <NODE>** command with **<NODE>** specifying the intact node. See the man page of the **mmsdrrestore** command for more details.
ccr_tiebreaker_dsk_fail	STATE_CHANGE	ERROR	Access to tiebreaker disks failed Item={0},ErrMsg={1}, Failed={2}	Access to all tiebreaker disks failed. See message for details.	Corrupted disk.	Check whether the tiebreaker disks are available.
ccr_tiebreaker_dsk_ok	STATE_CHANGE	INFO	All tiebreaker disks used by the GPFS CCR are accessible {0}	All tiebreaker disks used by the GPFS CCR are accessible.	N/A	N/A
ccr_tiebreaker_dsk_warn	STATE_CHANGE	WARNING	At least one tiebreaker disk is not accessible Item={0}, ErrMsg={1}, Failed={2}	At least one tiebreaker disk is not accessible. See message for details.	Corrupted disk.	Check whether the tiebreaker disks are accessible.
nodeleave_info	INFO	INFO	The CES node {0} left the cluster.	Shows the name of the node that leaves the cluster. This event might be logged on a different node; not necessarily on the leaving node.	A CES node left the cluster. The name of the leaving node is provided.	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
nodestatechange_info	INFO	INFO	Message: A CES node state change: Node {0} {1} {2} flag	Shows the modified node state. For example, the node turned to suspended mode, network down.	A node state change was detected. Details are shown in the message.	N/A
quorumloss	INFO	WARNING	The cluster detected a quorum loss.	The number of required quorum nodes does not match the minimum requirements. This can be an expected situation.	The cluster is in inconsistent or split-brain state. Reasons could be network or hardware issues, or quorum nodes are removed from the cluster. The event might not be logged on the same node that causes the quorum loss.	Recover from the underlying issue. Make sure the cluster nodes are up and running.
gpfs_down	STATE_CHANGE	ERROR	The IBM Spectrum Scale service is not running on this node. Normal operation cannot be done.	The IBM Spectrum Scale service is not running. This can be an expected state when the IBM Spectrum Scale service is shut down.	The IBM Spectrum Scale service is not running.	Check the state of the IBM Spectrum Scale file system daemon, and check for the root cause in the /var/adm/ras/ mmfs.log.latest log.
gpfs_up	STATE_CHANGE	INFO	The IBM Spectrum Scale service is running.	The IBM Spectrum Scale service is running.	The IBM Spectrum Scale service is running.	N/A
gpfs_warn	INFO	WARNING	IBM Spectrum Scale process monitoring returned unknown result. This could be a temporary issue.	Check of the IBM Spectrum Scale file system daemon returned unknown result. This could be a temporary issue, like a timeout during the check procedure.	The IBM Spectrum Scale file system daemon state could not be determined due to a problem.	Find potential issues for this kind of failure in the /var/adm/ras/ mmsysmonitor.log file.
info_on_duplicate_events	INFO	INFO	The event {0}{id} was repeated {1} times	Multiple messages of the same type were deduplicated to avoid log flooding.	Multiple events of the same type processed.	N/A
shared_root_bad	STATE_CHANGE	ERROR	Shared root is unavailable.	The CES shared root file system is bad or not available. This file system is required to run the cluster because it stores the cluster-wide information. This problem triggers a failover.	The CES framework detects the CES shared root file system to be unavailable on the node.	Check if the CES shared root file system and other expected IBM Spectrum Scale file systems are mounted properly.
shared_root_ok	STATE_CHANGE	INFO	Shared root is available.	The CES shared root file system is available. This file system is required to run the cluster because it stores cluster-wide information.	The CES framework detects the CES shared root file system to be OK.	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
quorum_down	STATE_CHANGE	ERROR	A quorum loss is detected.	The monitor service has detected a quorum loss. Reasons could be network or hardware issues, or quorum nodes are removed from the cluster. The event might not be logged on the node that causes the quorum loss.	The local node does not have quorum. It might be in an inconsistent or split-brain state.	Check whether the cluster quorum nodes are running and can be reached over the network. Check local firewall settings.
quorum_up	STATE_CHANGE	INFO	Quorum is detected.	The monitor detected a valid quorum.		N/A
quorum_warn	INFO	WARNING	The IBM Spectrum Scale quorum monitor could not be executed. This could be a timeout issue	The quorum state monitoring service returned an unknown result. This might be a temporary issue, like a timeout during the monitoring procedure.	The quorum state could not be determined due to a problem.	Find potential issues for this kind of failure in the /var/adm/ras/ mmsysmonitor.log file.
deadlock_detected	INFO	WARNING	The cluster detected a IBM Spectrum Scale file system deadlock	The cluster detected a deadlock in the IBM Spectrum Scale file system.	High file system activity might cause this issue.	The problem might be temporary or permanent. Check the /var/adm/ras/ mmfs.log.latest log files for more detailed information.
gpfsport_access_up	STATE_CHANGE	INFO	Access to IBM Spectrum Scale ip {0} port {1} ok	The TCP access check of the local IBM Spectrum Scale file system daemon port is successful.	The IBM Spectrum Scale file system service access check is successful.	N/A
gpfsport_down	STATE_CHANGE	ERROR	IBM Spectrum Scale port {0} is not active	The expected local IBM Spectrum Scale file system daemon port is not detected.	The IBM Spectrum Scale file system daemon is not running.	Check whether the IBM Spectrum Scale service is running.
gpfsport_access_down	STATE_CHANGE	ERROR	No access to IBM Spectrum Scale ip {0} port {1}. Check firewall settings	The access check of the local IBM Spectrum Scale file system daemon port is failed.	The port is probably blocked by a firewall rule.	Check whether the IBM Spectrum Scale file system daemon is running and check the firewall for blocking rules on this port.
gpfsport_up	STATE_CHANGE	INFO	IBM Spectrum Scale port {0} is active	The expected local IBM Spectrum Scale file system daemon port is detected.	The expected local IBM Spectrum Scale file system daemon port is detected.	N/A
gpfsport_warn	INFO	WARNING	IBM Spectrum Scale monitoring ip {0} port {1} returned unknown result	The IBM Spectrum Scale file system daemon port returned an unknown result.	The IBM Spectrum Scale file system daemon port could not be determined due to a problem.	Find potential issues for this kind of failure in the `/var/adm/ras/ mmsysmonitor.log` file.
gpfsport_access_warn	INFO	WARNING	IBM Spectrum Scale access check ip {0} port {1} failed. Check for valid IBM Spectrum Scale-IP	The access check of the IBM Spectrum Scale file system daemon port returned an unknown result.	The IBM Spectrum Scale file system daemon port access could not be determined due to a problem.	Find potential issues for this kind of failure in the `/var/adm/ras/ mmsysmonitor.log` file.
longwaiters_found	STATE_CHANGE	ERROR	Detected IBM Spectrum Scale long-waiters.	Longwaiter threads found in the IBM Spectrum Scale file system.	High load might cause this issue.	Check log files. This could be also a temporary issue.

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
no_longwaiters_found	STATE_CHANGE	INFO	No IBM Spectrum Scalelong-waiters	No longwaiter threads found in the IBM Spectrum Scale file system.	No longwaiter threads found in the IBM Spectrum Scale file system.	N/A
longwaiters_warn	INFO	WARNING	IBM Spectrum Scale long-waiters monitoring returned unknown result.	The long waiters check returned an unknown result.	The IBM Spectrum Scale file system long waiters check could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
quorumreached_detected	INFO	INFO	Quorum is achieved.	The cluster has achieved quorum.	The cluster has achieved quorum.	N/A
monitor_started	INFO	INFO	The IBM Spectrum Scale monitoring service has been started	The IBM Spectrum Scale monitoring service has been started, and is actively monitoring the system components.	N/A	Use the **mmhealth** command to query the monitoring status.
event_hidden	INFO_EXTERNAL	INFO	The event {0} was hidden.	An event used in the system health framework was hidden. It can still be seen with the **--verbose flag** in **mmhealth node show ComponentName**, if it is active. However, it will not affect its component's state anymore.	The **mmhealth event hide** command was used.	Use the **mmhealth event list hidden** command to see all hidden events. Use the **mmhealth event unhide**command to unhide the event again.
event_unhidden	INFO_EXTERNAL	INFO	The event {0} was unhidden.	An event was unhidden. This means, that the event will affect its component's state now if it is active. Furthermore it will be shown in the event table of 'mmhealth node show ComponentName' without --verbose flag.	The 'mmhealth event unhide' command was used.	If this is an active TIP event, fix it or hide it with **mmhealth event hide** command.

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gpfs_pagepool_small	INFO_EXTERNAL	INFO	The GPFS pagepool is smaller than or equal to 1G.	The size of the pagepool is essential to achieve optimal performance. With a larger pagepool, IBM Spectrum Scale can cache/prefetch more data which makes I/O operations more efficient. This event is raised because the pagepool is configured less than or equal to 1 GB.	The size of the pagepool is essential to achieve optimal performance. With a larger pagepool, IBM Spectrum Scale can cache/prefetch more data which makes IO operations more efficient. This event is raised because the pagepool is configured less than or equal to 1G.	Review the *Cache usage recommendations'* topic in the *General system configuration and tuning considerations* section ' for the pagepool size in the Knowledge Center. Although the pagepool should be higher than 1 GB, there are situations in which the administrator decides against a pagepool greater 1 GB. In this case or in case that the current setting fits what is recommended in the Knowledge Center, hide the event, either through the GUI or by using the **mmhealth event hide** command. The pagepool can be changed with the **mmchconfig** command. The **gpfs_pagepool_small** event will automatically disappear as soon as the new pagepool value larger than 1 GB is active. You must either reatart the system, or run the **mmchconfig -i flag** command. Consider that the actively used configuration is monitored. You can list the actively used configuration with the **mmdiag --config** command. The **mmlsconfig** command can include changes which are not activated yet.
gpfs_pagepool_ok	TIP	INFO	The GPFS pagepool is higher than 1 GB.	The GPFS pagepool is higher than 1G. Please consider, that the actively used config is monitored. You can see the actively used configuration with the **mmdiag --config** command.	The GPFS pagepool is higher than 1 GB.	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gpfs_maxfilestocache_small	TIP	TIP	The GPFS maxfilestocache is smaller than or equal to 100,000.	The size of *maxFilesToCache* is essential to achieve optimal performance, especially on protocol nodes. With a larger *maxFilesToCache* size, IBM Spectrum Scale can handle more concurrently open files, and is able to cache more recently used files, which makes I/O operations more efficient. This event is raised because the *maxFilesToCache* value is configured less than or equal to 100,000 on a protocol node.	The size of *maxFilesToCache* is essential to achieve optimal performance, especially on protocol nodes. With a larger *maxFilesToCache* size, IBM Spectrum Scale can handle more concurrently open files, and is able to cache more recently used files, which makes I/O operations more efficient. This event is raised because the *maxFilesToCache* value is configured less than or equal to 100,000 on a protocol node.	Review the*Cache usage recommendations'* topic in the *General system configuration and tuning considerations* section for the *maxFilesToCache* size in the Knowledge Center. Although the *maxFilesToCache* size should be higher than 100,000, there are situations in which the administrator decides against a *maxFilesToCache* size greater 100,000. In this case or in case that the current setting fits what is recommended in the Knowledge Center, hide the event either through the GUI or using the **mmhealth event hide** command. The *maxFilesToCache* can be changed with the **mmchconfig** command. The *gpfs_maxfilestocache_small* event will automatically disappear as soon as the new *maxFilesToCache* event with a value larger than 100,000 is active. You need to restart the gpfs daemon for this to take affect. Consider that the actively used configuration is monitored. You can list the actively used configuration with the **mmdiag --config**command. The **mmlsconfig** can include changes which are not activated yet.
gpfs_maxfilestocache_ok	TIP	INFO	The GPFS *maxFilesToCache* value is higher than 100,000.	The GPFS *maxFilesToCache* value is higher than 100,000. Please consider, that the actively used config is monitored. You can see the actively used configuration with the **mmdiag --config** command.	The GPFS *maxFilesToCache* is higher than 100,000.	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gpfs_maxstatcache_high	TIP	TIP	The GPFS *maxStatCache* value is higher than 0 on a Linux system.	The size of *maxStatCache* is useful to improve the performance of both the system and the IBM Spectrum Scale stat() calls for applications with a working set that does not fit in the regular file cache. Nevertheless the stat cache is not effective on a Linux platform. Therefore, it is recommended to set the *maxStatCache* attribute to 0 on a Linux platform. This event is raised because the *maxStatCache* value is configured higher than 0 on a Linux system.	The size of *maxStatCache* is useful to improve the performance of both the system and the IBM Spectrum Scale stat() calls for applications with a working set that does not fit in the regular file cache. Nevertheless the stat cache is not effective on a Linux platform. Therefore, it is recommended to set the *maxStatCache* attribute to 0 on a Linux platform. This event is raised because the *maxStatCache* value is configured higher than 0 on a Linux system.	Review the*Cache usage recommendations'* topic in the *General system configuration and tuning considerations* section for the *maxStatCache* size in the Knowledge Center. Although the *maxStatCache* size should be 0 on a Linux system, there are situations in which the administrator decides against a *maxStatCache* size of 0. In this case or in case that the current setting fits what is recommended in the Knowledge Center, hide the event either through the GUI or using the **mmhealth event hide** command. The *maxStatCache* can be changed with the **mmchconfig** command. The *gpfs_maxstatcache_high* event will automatically disappear as soon as the new *maxStatCache* value of 0 is active. You need to restart the gpfs daemon for this to take affect. Consider that the actively used configuration is monitored. You can list the actively used configuration with the **mmdiag --config**command. The **mmlsconfig** can include changes which are not activated yet.
gpfs_maxstatcache_ok	TIP	INFO	The GPFS maxFilesToCache is 0 on a linux system.	The GPFS *maxFilesToCache* is 0 on a Linux system. Consider that the actively used configuration is monitored. You can list the actively used configuration with the **mmdiag --config**command.	The GPFS *maxFilesToCache* is 0 on a Linux system.	N/A

Table 73. Events for the GPFS component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
callhome_not_enabled	TIP	TIP	Callhome is not installed, configured or enabled.	Callhome is a functionality that uploads cluster configuration and log files onto the IBM ECuREP servers. The uploaded data provides information that not only helps developers to improve the product, but also helps the support to resolve the PMR cases.	The cause can be one of the following: • The call home packages are not installed, • The call home is not configured, • There are no call home groups. • No call home group was enabled.	Install and configure callhome.
callhome_enabled	TIP	INFO	Call home is installed, configured and enabled.	By enabling the call home functionality you are providing useful information to the developers. This information will help the developers improve the product.	The call home packages are installed. The call home functionality is configured and enabled.	N/A
callhome_not_monitored	TIP	INFO	Callhome status is not monitored on the current node.	Callhome status is not monitored on the current node, but was, when it was the cluster manager.	Previously this node was a cluster manager, and call home monitoring was running on it.	N/A
local_fs_normal	STATE_CHANGE	INFO	The local file system with the mount point {0} reached a normal level.	The fill state of the file system to the dataStructureDump path (**mmdiag --config**) or /tmp/mmfs if not defined, and /var/mmfsis checked.	The fill level of the local file systems is ok.	N/A
local_fs_filled	STATE_CHANGE	WARNING	The local file system with the mount point {0} reached a warning level.	The fill state of the file system to the dataStructureDump path (**mmdiag --config**) or /tmp/mmfs if not defined, and /var/mmfsis checked.	The local file systems reached a warning level of under 1000 MB.	Delete some data on the local disk.
local_fs_full	STATE_CHANGE	ERROR	The local file system with the mount point {0} reached a nearly exhausted level.	The fill state of the file system to the dataStructureDump path (**mmdiag --config**) or /tmp/mmfs if not defined, and /var/mmfsis checked.	The local file systems reached a warning level of under 100 MB.	Delete some data on the local disk.
local_fs_unknown	INFO	WARNING	The fill level of the local file systems is unknown because of a non-expected output of the df command.	The fill state of the file system to the dataStructureDump path (**mmdiag --config**) or /tmp/mmfs if not defined, and /var/mmfsis checked.	Could not determine fill state of the local filesystems.	Does the df command exists on the node? Are there time issues with the df command, so that it can run into a time out?
local_fs_path_not_found	STATE_CHANGE	INFO	The configured dataStructureDump path {0} does not exists. Skipping monitoring.	The configured dataStructureDump path does not exists yet, therefore the disk capacity monitoring will be skipped.	The path of the dataStructureDump does not exists.	N/A

GUI events

The following table lists the events that are created for the *GUI* component.

Table 74. Events for the GUI component

Event	Event Type	Severity	Message	Description	Cause	User Action
gui_down	STATE_CHANGE	ERROR	The status of the GUI service must be {0} but it is {1} now.	The GUI service is down.	The GUI service is not running on this node, although it has the node class GUI_MGMT_SERVER_NODE.	Restart the GUI service or change the node class for this node.
gui_up	STATE_CHANGE	INFO	The status of the GUI service is {0} as expected.	The GUI service is running	The GUI service is running as expected.	N/A
gui_warn	INFO	INFO	The GUI service returned an unknown result.	The GUI service returned an unknown result.	The **service** or **systemctl** command returned unknown results about the GUI service.	Use either the **service** or **systemctl** command to check whether the GUI service is in the expected status. If there is no gpfsgui service although the node has the node class GUI_MGMT_SERVER_NODE, see the GUI documentation. Otherwise, monitor whether this warning appears more often.
gui_reachable_node	STATE_CHANGE	INFO	The GUI can reach the node {0}.	The GUI checks the reachability of all nodes.	The specified node can be reached by the GUI node.	None.
gui_unreachable_node	STATE_CHANGE	ERROR	The GUI can not reach the node {0}.	The GUI checks the reachability of all nodes.	The specified node can not be reached by the GUI node.	Check your firewall or network setup and if the specified node is up and running.
gui_cluster_up	STATE_CHANGE	INFO	The GUI detected that the cluster is up and running.	The GUI checks the cluster state.	The GUI calculated that a sufficient amount of quorum nodes is up and running.	None.
gui_cluster_down	STATE_CHANGE	ERROR	The GUI detected that the cluster is down.	The GUI checks the cluster state.	The GUI calculated that an insufficient amount of quorum nodes is up and running.	Check why the cluster lost quorum.
gui_cluster_state_unknown	STATE_CHANGE	WARNING	The GUI can not determine the cluster state.	The GUI checks the cluster state.	The GUI can not determine if a sufficient amount of quorum nodes is up and running.	None.
time_in_sync	STATE_CHANGE	INFO	The time on node {0} is in sync with the clusters median.	The GUI checks the time on all nodes.	The time on the specified node is in sync with the cluster median.	None.
time_not_in_sync	STATE_CHANGE	NODE	The time on node {0} is not in sync with the clusters median.	The GUI checks the time on all nodes.	The time on the specified node is not in sync with the cluster median.	Synchronize the time on the specified node.
time_sync_unknown	STATE_CHANGE	WARNING	The time on node {0} could not be determined.	The GUI checks the time on all nodes.	The time on the specified node could not be determined.	Check if the node is reachable from the GUI.

Table 74. Events for the GUI component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gui_pmcollector_connection_failed	STATE_CHANGE	ERROR	The GUI can not connect to the pmcollector running on {0} using port {1}.	The GUI checks the connection to the pmcollector.	The GUI can not connect to the pmcollector.	Check if the pmcollector service is running, and verify the firewall/network settings.
gui_pmcollector_connection_ok	STATE_CHANGE	INFO	The GUI can connect to the pmcollector running on {0} using port {1}.	The GUI checks the connection to the pmcollector.	The GUI can connect to the pmcollector.	None.
host_disk_normal	STATE_CHANGE	INFO	The local file systems on node {0} reached a normal level.	The GUI checks the fill level of the local file systems.	The fill level of the local file systems is ok.	None.
host_disk_filled	STATE_CHANGE	WARNING	A local file system on node {0} reached a warning level. {1}	The GUI checks the fill level of the local file systems.	The local file systems reached a warning level.	Delete data on the local disk.
host_disk_full	STATE_CHANGE	ERROR	A local file system on node {0} reached a nearly exhausted level. {1}	The GUI checks the fill level of the local filesystems.	The local file systems reached a nearly exhausted level.	Delete data on the local disk.
host_disk_unknown	STATE_CHANGE	WARNING	The fill level of local file systems on node {0} is unknown.	The GUI checks the fill level of the local filesystems.	Could not determine fill state of the local filesystems.	None.
sudo_ok	STATE_CHANGE	INFO	Sudo wrappers were enabled on the cluster and the GUI configuration for the cluster '{0}' is correct.	No problems regarding the current configuration of the GUI and the cluster were found.		N/A
sudo_admin_not_configured	STATE_CHANGE	ERROR	Sudo wrappers are enabled on the cluster '{0}', but the GUI is not configured to use Sudo Wrappers.	Sudo wrappers are enabled on the cluster, but the value for GPFS_ADMIN in /usr/lpp/mmfs/gui/conf/gpfsgui.properties was either not set or is still set to root. The value of GPFS_ADMIN should be set to the user name for which sudo wrappers were configured on the cluster.		Make sure that sudo wrappers were correctly configured for a user that is available on the GUI node and all other nodes of the cluster. This user name should be set as the value of the GPFS_ADMIN option in /usr/lpp/mmfs/gui/conf/gpfsgui.properties. After that restart the GUI using 'systemctl restart gpfsgui'.

Table 74. Events for the GUI component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
sudo_admin_not_exist	STATE_CHANGE	ERROR	Sudo wrappers are enabled on the cluster '{0}', but there is a misconfiguration regarding the user '{1}' that was set as GPFS_ADMIN in the GUI properties file.	Sudo wrappers are enabled on the cluster, but the user name that was set as GPFS_ADMIN in the GUI properties file at /usr/lpp/mmfs/gui/conf/gpfsgui.properties does not exist on the GUI node.		Make sure that sudo wrappers were correctly configured for a user that is available on the GUI node and all other nodes of the cluster. This user name should be set as the value of the GPFS_ADMIN option in /usr/lpp/mmfs/gui/conf/gpfsgui.properties. After that restart the GUI using 'systemctl restart gpfsgui'.
sudo_connect_error	STATE_CHANGE	ERROR	Sudo wrappers are enabled on the cluster '{0}', but the GUI cannot connect to other nodes with the user name '{1}' that was defined as GPFS_ADMIN in the GUI properties file.	When sudo wrappers are configured and enabled on a cluster, the GUI does not execute commands as root, but as the user for which sudo wrappers were configured. This user should be set as GPFS_ADMIN in the GUI properties file at /usr/lpp/mmfs/gui/conf/gpfsgui.properties		Make sure that sudo wrappers were correctly configured for a user that is available on the GUI node and all other nodes of the cluster. This user name should be set as the value of the GPFS_ADMIN option in /usr/lpp/mmfs/gui/conf/gpfsgui.properties. After that restart the GUI using 'systemctl restart gpfsgui'.
sudo_admin_set_but_disabled	STATE_CHANGE	WARNING	Sudo wrappers are not enabled on the cluster '{0}', but GPFS_ADMIN was set to a non-root user.	Sudo wrappers are not enabled on the cluster, but the value for GPFS_ADMIN in /usr/lpp/mmfs/gui/conf/gpfsgui.properties was set to a non-root user. The value of GPFS_ADMIN should be set to 'root' when sudo wrappers are not enabled on the cluster.</explanation>		Set GPFS_ADMIN in /usr/lpp/mmfs/gui/conf/gpfsgui.properties to 'root'. After that restart the GUI using 'systemctl restart gpfsgui'.
gui_config_cluster_id_ok	STATE_CHANGE	INFO	The cluster ID of the current cluster '{0}' and the cluster ID in the database do match.	No problems regarding the current configuration of the GUI and the cluster were found.		N/A

Table 74. Events for the GUI component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gui_config_cluster_id_mismatch	STATE_CHANGE	ERROR	The cluster ID of the current cluster '{0}' and the cluster ID in the database do not match ('{1}'). It seems that the cluster was recreated.	When a cluster is deleted and created again, the cluster ID changes, but the GUI's database still references the old cluster ID.		Clear the GUI's database of the old cluster information by dropping all tables: psql postgres postgres -c 'drop schema fscc cascade'. Then restart the GUI (systemctl restart gpfsgui).
gui_config_command_audit_ok	STATE_CHANGE	INFO	Command Audit is turned on on cluster level.	Command Audit is turned on on cluster level. This way the GUI will refresh the data it displays automatically when Spectrum Scale commands are executed via the CLI on other nodes in the cluster.		N/A
gui_config_command_audit_off_cluster	STATE_CHANGE	WARNING	Command Audit is turned off on cluster level.	Command Audit is turned off on cluster level. This configuration will lead to lags in the refresh of data displayed in the GUI.	Command Audit is turned off on cluster level.	Change the cluster configuration option *commandAudit* to 'on' (mmchconfig commandAudit = on) or 'syslogonly' (mmchconfig commandAudit = syslogonly). This way the GUI will refresh the data it displays automatically when Spectrum Scale commands are executed via the CLI on other nodes in the cluster.

Table 74. Events for the GUI component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gui_config_command_audit_off_nodes	STATE_CHANGE	WARNING	Command Audit is turned off on the following nodes: {1}	Command Audit is turned off on some nodes. This configuration will lead to lags in the refresh of data displayed in the GUI.	Command Audit is turned off on some nodes.	Change the cluster configuration option 'commandAudit' to 'on' (mmchconfig commandAudit = on -N [node name]) or 'syslogonly' (mmchconfig commandAudit = syslogonly -N [node name]) for the affected nodes. This way the GUI will refresh the data it displays automatically when Spectrum Scale commands are executed via the CLI on other nodes in the cluster.
gui_config_sudoers_ok	STATE_CHANGE	INFO	The /etc/sudoers configuration is correct.	The /etc/sudoers configuration is correct.		N/A
gui_config_sudoers_error	STATE_CHANGE	ERROR	There is a problem with the /etc/sudoers configuration. The secure_path of the scalemgmt user is not correct. Current value: {0} / Expected value: {1}	There is a problem with the /etc/sudoers configuration.		Make sure that '#includedir /etc/sudoers.d' directive is set in /etc/sudoers so the sudoers configuration drop-in file for the scalemgmt user (which the GUI process uses) is loaded from `/etc/sudoers.d/scalemgmt_sudoers`. Also make sure that the #includedir directive is the last line in the `/etc/sudoers` configuration file
gui_pmsensors_connection_failed	STATE_CHANGE	ERROR	The performance monitoring sensor service 'pmsensors' on node {0} is not sending any data.	The GUI checks if data can be retrieved from the pmcollector service for this node.	The performance monitoring sensor service 'pmsensors' is not sending any data. The service might be down or the time of the node is more than 15 minutes away from the time on the node hosting the performance monitoring collector service 'pmcollector'.	Check with 'systemctl status pmsensors'. If pmsensors service is 'inactive', run 'systemctl start pmsensors'.
gui_pmsensors_connection_ok	STATE_CHANGE	INFO	The state of performance monitoring sensor service 'pmsensor' on node {0} is OK.	The GUI checks if data can be retrieved from the pmcollector service for this node.	The state of performance monitoring sensor service 'pmsensor' is OK and it is sending data.	None.

Table 74. Events for the GUI component (continued)

Event	Event Type	Severity	Message	Description	Cause	User Action
gui_snap_running	INFO	WARNING	Operations for rule {1} are still running at the start of the next management of rule {1}.	Operations for a rule are still running at the start of the next management of that rule	Operations for a rule are still running.	None.
gui_snap_rule_ops_exceeded	INFO	WARNING	The number of pending operations exceeds {1} operations for rule {2}.	The number of pending operations for a rule exceed a specified value.	The number of pending operations for a rule exceed a specified value.	None.
gui_snap_total_ops_exceeded	INFO	WARNING	The total number of pending operations exceeds {1} operations.	The total number of pending operations exceed a specified value.	The total number of pending operations exceed a specified value.	None.
gui_snap_time_limit_exceeded_fset	INFO	WARNING	A snapshot operation exceeds {1} minutes for rule {2} on file system {3}, file set {0}.	The snapshot operation resulting from the rule is exceeding the established time limit.	A snapshot operation exceeds a specified number of minutes.	None.
gui_snap_time_limit_exceeded_fs	INFO	WARNING	A snapshot operation exceeds {1} minutes for rule {2} on file system {0}.	The snapshot operation resulting from the rule is exceeding the established time limit.	A snapshot operation exceeds a specified number of minutes.	None.
gui_snap_create_failed_fset	INFO	ERROR	A snapshot creation invoked by rule {1} failed on file system {2}, file set {0}.	The snapshot was not created according to the specified rule.	A snapshot creation invoked by a rule fails.	Try to create the snapshot again manually.
gui_snap_create_failed_fs	INFO	ERROR	A snapshot creation invoked by rule {1} failed on file system {0}.	The snapshot was not created according to the specified rule.	A snapshot creation invoked by a rule fails.	Try to create the snapshot again manually.
gui_snap_delete_failed_fset	INFO	ERROR	A snapshot deletion invoked by rule {1} failed on file system {2}, file set {0}.	The snapshot was not deleted according to the specified rule.	A snapshot deletion invoked by a rule fails.	Try to manually delete the snapshot.
gui_snap_delete_failed_fs	INFO	ERROR	A snapshot deletion invoked by rule {1} failed on file system {0}.	The snapshot was not deleted according to the specified rule.	A snapshot deletion invoked by a rule fails.	Try to manually delete the snapshot.

Hadoop connector events

The following table lists the events that are created for the *Hadoop connector* component.

Table 75. Events for the Hadoop connector component

Event	Event type	Severity	Message	Description	Cause	User Action
hadoop_datanode_down	STATE_CHANGE	ERROR	Hadoop DataNode service is down.	The Hadoop DataNode service is down.	The Hadoop DataNode process is not running.	Start the Hadoop DataNode service.

Table 75. Events for the Hadoop connector component (continued)

Event	Event type	Severity	Message	Description	Cause	User Action
hadoop_datanode_up	STATE_CHANGE	INFO	Hadoop DataNode service is up.	The Hadoop DataNode service is running.	The Hadoop DataNode process is running.	N/A
hadoop_datanode_warn	INFO	WARNING	Hadoop DataNode monitoring returned unknown results.	The Hadoop DataNode service check returned unknown results.	The Hadoop DataNode service status check returned unknown results.	If this status persists after a few minutes, restart the DataNode service.
hadoop_namenode_down	STATE_CHANGE	ERROR	Hadoop NameNode service is down.	The Hadoop NameNode service is down.	The Hadoop NameNode process is not running.	Start the Hadoop NameNode service.
hadoop_namenode_up	STATE_CHANGE	INFO	Hadoop NameNode service is up.	The Hadoop NameNode service is running.	The Hadoop NameNode process is running.	N/A
hadoop_namenode_warn	INFO	WARNING	Hadoop NameNode monitoring returned unknown results.	The Hadoop NameNode service status check returned unknown results.	The Hadoop NameNode service status check returned unknown results.	If this status persists after a few minutes, restart the NameNode service.

Keystone events

The following table lists the events that are created for the *Keystone* component.

Table 76. Events for the Keystone component

Event	EventType	Severity	Message	Description	Cause	User action
ks_failed	STATE_CHANGE	ERROR	The status of the keystone (httpd) process must be {0} but it is {1} now.	The keystone (httpd) process is not in the expected state.	If the object authentication is **local**, **AD**, or**LDAP**, then the process is failed unexpectedly. If the object authentication is **none** or **userdefined**, then the process is expected to be stopped, but it was running.	Make sure that the process is in the expected state.
ks_ok	STATE_CHANGE	INFO	The status of the keystone (httpd) is {0} as expected.	The keystone (httpd) process is in the expected state.	If the object authentication is **local**, **AD**, or**LDAP**, process is running. If the object authentication is **none** or **userdefined**, then the process is stopped as expected.	N/A

Table 76. Events for the Keystone component (continued)

Event	EventType	Severity	Message	Description	Cause	User action
ks_restart	INFO	WARNING	The {0} service is failed. Trying to recover.	The {0} service failed. Trying to recover.	A service was not in the expected state.	None, recovery is automatic.
ks_url_exfail	STATE_CHANGE	WARNING	Keystone request failed using {0}.	A request to an external keystone URL failed.	A HTTP request to an external keystone server failed.	Check that httpd / keystone is running on the expected server, and is accessible with the defined ports.
ks_url_failed	STATE_CHANGE	ERROR	The {0} request to keystone is failed.	A keystone URL request failed.	An HTTP request to keystone failed.	Check that httpd / keystone is running on the expected server and is accessible with the defined ports.
ks_url_ok	STATE_CHANGE	INFO	The {0} request to keystone is successful.	A keystone URL request was successful.	A HTTP request to keystone returned successfully.	N/A
ks_url_warn	INFO	WARNING	Keystone request on {0} returned unknown result.	A keystone URL request returned an unknown result.	A simple HTTP request to keystone returned with an unexpected error.	Check that httpd / keystone is running on the expected server and is accessible with the defined ports.
ks_warn	INFO	WARNING	Keystone (httpd) process monitoring returned unknown result.	The keystone (httpd) monitoring returned an unknown result.	A status query for httpd returned an unexpected error.	Check service script and settings of httpd.
postgresql_failed	STATE_CHANGE	ERROR	The status of the postgresql-obj process must be {0} but it is {1} now.	The postgresql-obj process is in an unexpected mode.	The database backend for object authentication is supposed to run on a single node. Either the database is not running on the designated node or it is running on a different node.	Check that postgresql-obj is running on the expected server.
postgresql_ok	STATE_CHANGE	INFO	The status of the postgresql-obj process is {0} as expected.	The postgresql-obj process is in the expected mode.	The database backend for object authentication is supposed to run on the right node while being stopped on other nodes.	N/A
postgresql_warn	INFO	WARNING	The status of the postgresql-obj process monitoring returned unknown result.	The postgresql-obj process monitoring returned an unknown result.	A status query for postgresql-obj returned with an unexpected error.	Check postgres database engine.
ldap_reachable	STATE_CHANGE	INFO	External LDAP server {0} is up.	The external LDAP server is operational.	The external LDAP server is operational.	N/A
ldap_unreachable	STATE_CHANGE	ERROR	External LDAP server {0} is unresponsive.	The external LDAP server is unresponsive.	The local node is unable to connect to the LDAP server.	Verify network connection and check if that LDAP server is operational.

Message queue events

The following table lists the events that are created for the *Message Queue* component.

Table 77. Events for the Message Queue component

Event	Event Type	Severity	Message	Description	Cause	User Action
zookeeper_ok	STATE_CHANGE	INFO	The zookeeper process is as expected, state is {0}.	The zookeeper process is running.	The zookeeper process is running.	N/A
zookeeper_warn	INFO	WARNING	The zookeeper process monitoring returned an unknown result.	The zookeeper check returned an unknown result.	A status query for zookeeper returned with an unexpected error.	Check the service script and settings.
zookeeper_failed	STATE_CHANGE	ERROR	zookeeper process output should be {0} but is {1}.	The zookeeper process is not running.	The zookeeper process is not running.	Check the status of zookeeper process, and review its logs.
kafka_ok	STATE_CHANGE	INFO	The kafka process is as expected, state is {0}.	The kafka process is running.	The kafka process is running.	N/A
kafka_warn	INFO	WARNING	The kafka process monitoring returned an unknown result.	The kafka check returned an unknown result.	A status query for kafka returned with an unexpected error.	Check the service script and settings.
kafka_failed	STATE_CHANGE	ERROR	kafka process should be {0} but is {1}	The kafka process is not running.	The kafka process is not running.	Check status of kafka process and review its logs.
stop_msgqueue_service	INFO	INFO	MSGQUEUE service was stopped.	Information about a MSGQUEUE service stop.	The MSGQUEUE service was stopped.	N/A
start_msgqueue_service	INFO	INFO	MSGQUEUE service was started.	Information about a MSGQUEUE service start.	The MSGQUEUE service was started.	N/A

Network events

The following table lists the events that are created for the *Network* component.

Table 78. Events for the Network component

Event	EventType	Severity	Message	Description	Cause	User Action
bond_degraded	STATE_CHANGE	INFO	Some slaves of the network bond {0} is down.	Some of the bond parts are malfunctioning.	Some slaves of the bond are not functioning properly.	Check the bonding configuration, network configuration, and cabling of the malfunctioning slaves of the bond.
bond_down	STATE_CHANGE	ERROR	All slaves of the network bond {0} are down.	All slaves of a network bond are down.	All slaves of this network bond are down.	Check the bonding configuration, network configuration, and cabling of all slaves of the bond.

Table 78. Events for the Network component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
bond_up	STATE_CHANGE	INFO	All slaves of the network bond {0} are working as expected.	This bond is functioning properly.	All slaves of this network bond are functioning properly.	N/A
ces_disable_node network	INFO	INFO	Network is disabled.	The network configuration is disabled as the **mmchnode --ces-disable** command is issued by the user.	The network configuration is disabled as the **mmchnode --ces-disable** command is issued by the user.	N/A
ces_enable_node network	INFO	INFO	Network is enabled.	The network configuration is enabled as a result of issuing the **mmchnode --ces- enable** command.	The network configuration is enabled as a result of issuing the **mmchnode --ces-enable** command.	N/A
ces_startup_network	INFO	INFO	CES network service is started.	The CES network is started.	CES network IPs are started.	N/A
handle_network _problem_info	INFO	INFO	The following network problem is handled: Problem: {0}, Argument: {1}	Information about network-related reconfigurations. For example, enable or disable IPs and assign or unassign IPs.	A change in the network configuration.	N/A
ib_rdma_enabled	STATE_CHANGE	INFO	Infiniband in RDMA mode is enabled.	Infiniband in RDMA mode is enabled for IBM Spectrum Scale.	The user has enabled verbsRdma with mmchconfig.	N/A
ib_rdma_disabled	STATE_CHANGE	INFO	Infiniband in RDMA mode is disabled.	Infiniband in RDMA mode is not enabled for IBM Spectrum Scale.	The user has not enabled verbsRdma with mmchconfig.	N/A
ib_rdma_ports_undefined	STATE_CHANGE	ERROR	No NICs and ports are set up for IB RDMA.	No NICs and ports are set up for IB RDMA.	The user has not set verbsPorts with mmchconfig.	Set up the NICs and ports to use with the verbsPorts setting in mmchconfig.
ib_rdma_ports_wrong	STATE_CHANGE	ERROR	The verbsPorts is incorrectly set for IB RDMA.	The verbsPorts setting has wrong contents.	The user has wrongly set verbsPorts with mmchconfig.	Check the format of the verbsPorts setting in mmlsconfig.
ib_rdma_ports_ok	STATE_CHANGE	INFO	The verbsPorts is correctly set for IB RDMA.	The verbsPorts setting has a correct value.	The user has set verbsPorts correctly.	
ib_rdma_port_width_low	STATE_CHANGE	WARNING	IB RDMA NIC {id} uses a smaller port width than enabled.	The currently active link width is lower than the enabled maximum link width.	The currently active link width is lower than the enabled maximum link width.	Check the settings of the specified IB RDMA NIC (ibportstate).

Table 78. Events for the Network component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
ib_rdma_port_width_ok	STATE_CHANGE	INFO	IB RDMA NIC {id} uses maximum enabled port width.	The currently active link width equal to the enabled maximum link width.	The currently active link width equal to the enabled maximum link width.	N/A
ib_rdma_port_speed_low	STATE_CHANGE	WARNING	IB RDMA NIC {id} uses a smaller port speed than enabled.	The currently active link speed is lower than the enabled maximum link speed.	The currently active link speed is lower than the enabled maximum link speed.	Check the settings of the specified IB RDMA NIC (ibportstate).
ib_rdma_port_speed_ok	STATE_CHANGE	INFO	IB RDMA NIC {id} uses maximum enabled port speed.	The currently active link speed equal to the enabled maximum link speed.	The currently active link speed equal to the enabled maximum link speed.	N/A
ib_rdma_port_width_subopti	STATE_CHANGE	TIP	IB RDMA NIC {id} uses a smaller port width than supported.	The currently enabled link width is lower than the supported maximum link width.	The currently enabled link width is lower than the supported maximum link width.	Check the settings of the specified IB RDMA NIC (ibportstate).
ib_rdma_port_width_optimal	STATE_CHANGE	INFO	IB RDMA NIC {id} uses maximum supported port width.	The currently enabled link width is equal to the supported maximum link width.	The currently enabled link width is equal to the supported maximum link width.	N/A
ib_rdma_port_speed_subopti	STATE_CHANGE	TIP	IB RDMA NIC {id} uses a smaller port speed than supported.	The currently enabled link speed is lower than the supported maximum link speed.	The currently enabled link speed is lower than the supported maximum link speed.	Check the settings of the specified IB RDMA NIC (ibportstate).
ib_rdma_port_speed_optimal	STATE_CHANGE	INFO	IB RDMA NIC {id} uses maximum supported port speed.	The currently enabled link speed is equal to the supported maximum link speed.	The currently enabled link speed is equal to the supported maximum link speed.	N/A
ib_rdma_verbs_started	STATE_CHANGE	INFO	VERBS RDMA was started.	IBM Spectrum Scale started VERBS RDMA	The IB RDMA-related libraries, which IBM Spectrum Scale uses, are working properly.	
ib_rdma_verbs_failed	STATE_CHANGE	ERROR	VERBS RDMA was not started.	IBM Spectrum Scale could not start VERBS RDMA.	The IB RDMA related libraries are improperly installed or configured.	Check `/var/adm/ras/mmfs.log.latest` for the root cause hints. Check if all relevant IB libraries are installed and correctly configured.

Table 78. Events for the Network component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
ib_rdma_libs_wrong_path	STATE_CHANGE	ERROR	The library files could not be found.	At least one of the library files (`librdmacm` and `libibverbs`) could not be found with an expected path name.	Either the libraries are missing or their pathnames are wrongly set.	Check the verbsLibName, verbsRdmaCm LibName settings by issuing the **mmdiag --config** command.
ib_rdma_libs_found	STATE_CHANGE	INFO	All checked library files could be found.	All checked library files (`librdmacm` and `libibverbs`) could be found with expected path names.	The library files are in the expected directories and have expected names.	
ib_rdma_nic_found	INFO_ADD_ENTITY	INFO	IB RDMA NIC {id} was found.	A new IB RDMA NIC was found.	A new relevant IB RDMA NIC is listed by `ibstat`.	
ib_rdma_nic_vanished	INFO_DELETE_ENTITY	INFO	IB RDMA NIC {id} has vanished.	The specified IB RDMA NIC can not be detected anymore.	One of the previously monitored IB RDMA NICs is not listed by `ibstat` anymore.	
ib_rdma_nic_recognized	STATE_CHANGE	INFO	IB RDMA NIC {id} was recognized.	The specified IB RDMA NIC was correctly recognized for usage by IBM Spectrum Scale.	The specified IB RDMA NIC is reported in `mmfsadm dump verb`.	
ib_rdma_nic_unrecognized	STATE_CHANGE	ERROR	IB RDMA NIC {id} was not recognized.	The specified IB RDMA NIC was not correctly recognized for usage by IBM Spectrum Scale.	The specified IB RDMA NIC is not reported in `mmfsadm dump verb`.	Check the 'verbsPorts' setting by issuing the **mmdiag --config**. If no configuration issue is found, restart the GPFSdeamon on the current node on the local node by using **mmshutdown** and **mmstartup** commands.
ib_rdma_nic_up	STATE_CHANGE	INFO	NIC {0} can connect to the gateway.	The specified IB RDMA NIC is up.	The specified IB RDMA NIC is up according to ibstat.	
ib_rdma_nic_down	STATE_CHANGE	ERROR	NIC {id} can connect to the gateway.	The specified IB RDMA NIC is down.	The specified IB RDMA NIC is down according to ibstat.	Enable the specified IB RDMA NIC
ib_rdma_link_up	STATE_CHANGE	INFO	IB RDMA NIC {id} is up.	The physical link of the specified IB RDMA NIC is up.	Physical state of the specified IB RDMA NIC is 'LinkUp' according to ibstat.	

Table 78. Events for the Network component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
ib_rdma_link_down	STATE_CHANGE	ERROR	IB RDMA NIC {id} is down.	The physical link of the specified IB RDMA NIC is down.	Physical state of the specified IB RDMA NIC is not 'LinkUp' according to ibstat.	Check the cabling of the specified IB RDMA NIC.
many_tx_errors	STATE_CHANGE	ERROR	NIC {0} had many TX errors since the last monitoring cycle.	The network adapter had many TX errors since the last monitoring cycle.	The /proc/net/dev folder lists the TX errors that are reported for this adapter.	Check the network cabling and network infrastructure.
move_cesip_from	INFO	INFO	The IP address {0} is moved from this node to the node {1}.	A CES IP address is moved from the current node to another node.	Rebalancing of CES IP addresses.	N/A
move_cesip_to	INFO	INFO	The IP address {0} is moved from node {1} to this node.	A CES IP address is moved from another node to the current node.	Rebalancing of CES IP addresses.	N/A
move_cesips_infos	INFO	INFO	A CES IP movement is detected.	The CES IP addresses can be moved if a node failover from one node to one or more other nodes. This message is logged on a node monitoring this; not necessarily on any affected node.	A CES IP movement was detected.	N/A
network_connectivity_down	STATE_CHANGE	ERROR	The NIC {0} cannot connect to the gateway.	This network adapter cannot connect to the gateway.	The gateway does not respond to the sent connections-checking packets.	Check the network configuration of the network adapter, gateway configuration, and path to the gateway.
network_connectivity_up	STATE_CHANGE	INFO	The NIC {0} can connect to the gateway.	This network adapter can connect to the gateway.	The gateway responds to the sent connections-checking packets.	N/A
network_down	STATE_CHANGE	ERROR	Network is down.	This network adapter is down.	This network adapter is disabled.	Enable this network adapter.
network_found	INFO	INFO	The NIC {0} is detected.	A new network adapter is detected.	A new NIC, which is relevant for the IBM Spectrum Scale monitoring, is listed by the **ip a** command.	N/A
network_ips_down	STATE_CHANGE	ERROR	No relevant NICs detected.	No relevant network adapters detected.	No network adapters are assigned with the IPs that are the dedicated to the IBM Spectrum Scale system.	Find out, why the IBM Spectrum Scale-relevant IPs were not assigned to any NICs.

Table 78. Events for the Network component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
network_ips_up	STATE_CHANGE	INFO	Relevant IPs are assigned to the NICs that are detected in the system.	Relevant IPs are assigned to the network adapters.	At least one IBM Spectrum Scale-relevant IP is assigned to a network adapter.	N/A
network_ips_partially_down	STATE_CHANGE	ERROR	Some relevant IPs are not served by found NICs: {0}	Some relevant IPs are not served by network adapters	At least one Spectrum Scale-relevant IP is not assigned to a network adapter.	Find out, why the specified Spectrum Scale-relevant IPs were not assigned to any NICs
network_link_down	STATE_CHANGE	ERROR	Physical link of the NIC {0} is down.	The physical link of this adapter is down.	The flag LOWER_UP is not set for this NIC in the output of the **ip a** command.	Check the cabling of this network adapter.
network_link_up	STATE_CHANGE	INFO	Physical link of the NIC {0} is up.	The physical link of this adapter is up.	The flag LOWER_UP is set for this NIC in the output of the **ip a** command.	N/A
network_up	STATE_CHANGE	INFO	Network is up.	This network adapter is up.	This network adapter is enabled.	N/A
network_vanished	INFO	INFO	The NIC {0} could not be detected.	One of network adapters could not be detected.	One of the previously monitored NICs is not listed in the output of the **ip a** command.	N/A
no_tx_errors	STATE_CHANGE	INFO	The NIC {0} had no or an insignificant number of TX errors.	The NIC had no or an insignificant number of TX errors.	The /proc/net/dev folder lists no or insignificant number of TX errors for this adapter.	Check the network cabling and network infrastructure.

NFS events

The following table lists the events that are created for the *NFS* component.

Table 79. Events for the NFS component

Event	EventType	Severity	Message	Description	Cause	User Action
dbus_error	STATE_CHANGE	WARNING	DBus availability check failed.	Failed to query DBus, if the NFS service is registered.	The DBus was detected as down. This might cause several issues on the local node.	Stop the NFS service, restart the DBus, and start the NFS service again.

Table 79. Events for the NFS component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
disable_nfs_service	INFO	INFO	CES NFS service is disabled.	The NFS service is disabled on this node. Disabling a service also removes all configuration files. This is different from stopping a service.	The user has executed the **mmces service disable nfs** command.	N/A
enable_nfs_service	INFO	INFO	CES NFS service is enabled.	The NFS service is enabled on this node. Enabling a protocol service also automatically installs the required configuration files with the current valid configuration settings.	The user has executed the **mmces service enable nfs** command.	N/A
ganeshaexit	INFO	INFO	CES NFS is stopped.	An NFS server instance has terminated.	An NFS instance terminated or was killed.	Restart the NFS service when the root cause for this issue is solved.
ganeshagrace	INFO	INFO	CES NFS is set to grace mode.	The NFS server is set to grace mode for a limited time. This gives time to the previously connected clients to recover their file locks.	The grace period is always cluster wide. NFS export configurations might have changed, and one or more NFS servers were restarted.	N/A
nfs3_down	INFO	WARNING	NFS v3 NULL check is failed.	The NFS v3 NULL check failed when expected it to be functioning. This check verifies if the NFS server reacts to NFS v3 requests. The NFS v3 protocol must be enabled for this check. If this down state is detected, further checks are done to figure out if the NFS server is still working. If the NFS server seems not to be working, then a failover is triggered. If NFS v3 and NFS v4 protocols are configured, then only the v3 NULL test is performed.	The NFS server might hang or is under high load so that the request might not be processed.	Check the health state of the NFS server and restart, if necessary.

Table 79. Events for the NFS component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
nfs3_up	INFO	INFO	NFS v3 check is successful.	The NFS v3 NULL check works as expected.		
nfs4_down	INFO	WARNING	NFS v4 check is failed.	The NFS v4 NULL check failed. This check verifies if the NFS server reacts to NFS v4 requests. The NFS v4 protocol must be enabled for this check. If this down state is detected, further checks are done to figure out if the NFS server is still working. If the NFS server seems to be not working, then a failover is triggered.	The NFS server may hang or is under high load, so that the request could not be processed.	Check the health state of the NFS server and restart, if necessary.
nfs4_up	INFO	INFO	NFS v4 check is successful.	The NFS v4 NULL check was successful.		N/A
nfs_active	STATE_CHANGE	INFO	NFS service is now active.	The NFS service must be up and running, and in a healthy state to provide the configured file exports.	The NFS server is detected as active.	N/A
nfs_dbus_error	STATE_CHANGE	WARNING	NFS check via DBus failed.	The NFS service must be registered on DBus to be fully working. This is currently not the case.	The NFS service is registered on DBus, but there was a problem accessing it.	Check the health state of the NFS service and restart the NFS service. Check the log files for reported issues.
nfs_dbus_failed	STATE_CHANGE	WARNING	NFS check via DBus did not return expected message.	NFS service configuration settings (log configuration settings) are queried through DBus. The result is checked for expected keywords.	The NFS service is registered on DBus, but the check via DBus did not return the expected result.	Stop the NFS service and start it again. Check the log configuration of the NFS service.
nfs_dbus_ok	STATE_CHANGE	INFO	NFS check via DBus is successful.	The check if the NFS service is registered on DBus and working, was successful.	The NFS service is registered on DBus and working.	N/A
nfs_in_grace	STATE_CHANGE	WARNING	NFS is in grace mode.	The monitor detected that CES NFS is in grace mode. During this time the NFS state is shown as degraded.	The NFS service was started or restarted.	N/A

Table 79. Events for the NFS component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
nfs_not_active	STATE_CHANGE	ERROR	NFS service is not active.	A check showed that the CES NFS service, which is supposed to be running is not active.	Process might have hung.	Restart the CES NFS.
nfs_not_dbus	STATE_CHANGE	WARNING	NFS service not available as DBus service.	The NFS service is currently not registered on DBus. In this mode, the NFS service is not fully working. Exports cannot be added or removed, and not set in grace mode, which is important for data consistency.	The NFS service might have been started while the DBus was down.	Stop the NFS service, restart the DBus, and start the NFS service again.
nfs_sensors_active	TIP	INFO	The NFS perfmon sensor {0} is active.	The NFS perfmon sensors are active. This event's monitor is only running once an hour.	The NFS perfmon sensors' period attribute is greater than 0.	
nfs_sensors_ inactive	TIP	TIP	The following NFS perfmon sensor {0} is inactive.	The NFS perfmon sensors are inactive. This event's monitor is only running once an hour.	The NFS perfmon sensors' period attribute is 0.	Set the period attribute of the NFS sensors to a value greater than 0. For this use the command **mmperfmon config update SensorName.period =N**, where *SensorName* is the name of a specific NFS sensor, and *N* is a natural number greater 0. Please consider, that this TIP monitor is running only once per hour, and it might take up to one hour in the worst case to detect the changes in the configuration.
nfsd_down	STATE_CHANGE	ERROR	NFSD process is not running.	Checks for an NFS service process.	The NFS server process was not detected.	Check the health state of the NFS server and restart, if necessary. The process might hang or is in failed state.
nfsd_up	STATE_CHANGE	INFO	NFSD process is running.	The NFS server process was detected.		N/A
nfsd_warn	INFO	WARNING	NFSD process monitoring returned unknown result.	The NFS server process monitoring returned an unknown result.	The NFS server process state could not be determined due to a problem.	Check the health state of the NFS server and restart, if necessary. The process might hang or is in a defunct state. Make sure the kernel NFS server is not running.
nfsd_restart	INFO	WARNING	NFSD process restarted.	An expected NFS service process was not running and then restarted.	The NFS server process was not detected and restarted.	Check the health state of the NFS server and restart, if necessary. Check the issues which lead to the unexpected failure. Make sure kernel NFS server is not running.

Table 79. Events for the NFS component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
portmapper_down	STATE_CHANGE	ERROR	Portmapper port 111 is not active.	The portmapper is needed to provide the NFS services to clients.	The portmapper is not running on port 111.	Check if the portmapper service is running, and if any services are conflicting with the portmapper service on this system.
portmapper_up	STATE_CHANGE	INFO	Portmapper port is now active.	The portmapper is running on port 111.		N/A
portmapper_warn	INFO	WARNING	Portmapper port monitoring (111) returned unknown result.	The portmapper process monitoring returned an unknown result.	The portmapper status could not be determined due to a problem.	Restart the portmapper, if necessary.
postIpChange_info	INFO	INFO	IP addresses modified (post change).	The portmapper process monitoring returned an unknown result	CES IP addresses were moved or added to the node, and activated.	N/A
rquotad_down	INFO	INFO	The rpc.rquotad process is not running.	Currently not in use. Future.	N/A	N/A
rquotad_up	INFO	INFO	The rpc.rquotad process is running.	Currently not in use. Future.	N/A	N/A
start_nfs_service	INFO	INFO	CES NFS service is started.	Notification about a NFS service start.	The NFS service was started by issuing the **mmces service start nfs** command.	N/A
statd_down	STATE_CHANGE	ERROR	The rpc.statd process is not running.	The statd process is used by NFSv3 to handle file locks.	The statd process is not running.	Stop and start the NFS service. This also attempts to start the statd process.
statd_up	STATE_CHANGE	INFO	The rpc.statd process is running.	The statd process is used by NFS v3 to handle file locks.		N/A
stop_nfs_service	INFO	INFO	CES NFS service is stopped.	Notification about an NFS service stop.	The NFS service was stopped (e.g. by using the **mmces service stop nfs**).	N/A

Object events

The following table lists the events that are created for the *Object* component.

Table 80. Events for the object component

Event	EventType	Severity	Message	Description	Cause	User Action
account-auditor_failed	STATE_CHANGE	ERROR	The status of the account-auditor process must be {0} but it is {1} now.	The account-auditor process is not in the expected state.	The account-auditor process is expected to be running on the singleton node only.	Check the status of openstack-swift-account-auditor process and object singleton flag.
account-auditor_ok	STATE_CHANGE	INFO	The account-auditor process status is {0} as expected.	The account-auditor process is in the expected state.	The account-auditor process is expected to be running on the singleton node only.	N/A

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
account-auditor_warn	INFO	WARNING	The account-auditor process monitoring returned unknown result.	The account-auditor process monitoring service returned an unknown result.	A status query for openstack-swift-account-auditor process returned with an unexpected error.	Check service script and settings.
account-reaper_failed	STATE_CHANGE	ERROR	The status of the account-reaper process must be {0} but it is {1} now.	The account-reaper process is not running.	The account-reaper process is not running.	Check the status of openstack-swift-account-reaper process.
account-reaper_ok	STATE_CHANGE	INFO	The status of the account-reaper process is {0} as expected.	The account-reaper process is running.	The account-reaper process is running.	N/A
account-reaper_warn	INFO	WARNING	The account-reaper process monitoring service returned an unknown result.	The account-reaper process monitoring service returned an unknown result.	A status query for openstack-swift-account-reaper returned with an unexpected error.	Check service script and settings.
account-replicator_failed	STATE_CHANGE	ERROR	The status of the account-replicator process must be {0} but it is {1} now.	The account-replicator process is not running.	The account-replicator process is not running.	Check the status of openstack-swift-account-replicator process.
account-replicator_ok	STATE_CHANGE	INFO	The status of the account-replicator process is {0} as expected.	The account-replicator process is running.	The account-replicator process is running.	N/A
account-replicator_warn	INFO	WARNING	The account-replicator process monitoring service returned an unknown result.	The account-replicator check returned an unknown result.	A status query for openstack-swift-account-replicator returned with an unexpected error.	Check the service script and settings.
account-server_failed	STATE_CHANGE	ERROR	The status of the account-server process must be {0} but it is {1} now.	The account-server process is not running.	The account-server process is not running.	Check the status of openstack-swift-account process.
account-server_ok	STATE_CHANGE	INFO	The status of the account process is {0} as expected.	The account-server process is running.	The account-server process is running.	N/A
account-server_warn	INFO	WARNING	The account-server process monitoring service returned unknown result.	The account-server check returned unknown result.	A status query for openstack-swift-account returned with an unexpected error.	Check the service script and existing configuration.
container-auditor_failed	STATE_CHANGE	ERROR	The status of the container-auditor process must be {0} but it is {1} now.	The container-auditor process is not in the expected state.	The container-auditor process is expected to be running on the singleton node only.	Check the status of openstack-swift-container-auditor process and object singleton flag.
container-auditor_ok	STATE_CHANGE	INFO	The status of the container-auditor process is {0} as expected.	The container-auditor process is in the expected state.	The container-auditor process is running on the singleton node only as expected.	N/A

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
container-auditor_warn	INFO	WARNING	The container-auditor process monitoring service returned unknown result.	The container-auditor monitoring service returned an unknown result.	A status query for openstack-swift-container-auditor returned with an unexpected error.	Check service script and settings.
container-replicator_failed	STATE_CHANGE	ERROR	The status of the container-replicator process must be {0} but it is {1} now.	The container-replicator process is not running.	The container-replicator process is not running.	Check the status of openstack-swift-container-replicator process.
container-replicator_ok	STATE_CHANGE	INFO	The status of the container-replicator process is {0} as expected.	The container-replicator process is running.	The container-replicator process is running.	N/A
container-replicator_warn	INFO	WARNING	The status of the container-replicator process monitoring service returned unknown result.	The container-replicator check returned an unknown result.	A status query for openstack-swift-container-replicator returned with an unexpected error.	Check service script and settings.
container-server_failed	STATE_CHANGE	ERROR	The status of the container-server process must be {0} but it is {1} now.	The container-server process is not running.	The container-server process is not running.	Check the status of openstack-swift-container process.
container-server_ok	STATE_CHANGE	INFO	The status of the container-server is {0} as expected.	The container-server process is running.	The container-server process is running.	N/A
container-server_warn	INFO	WARNING	The container-server process monitoring service returned unknown result.	The container-server check returned an unknown result.	A status query for openstack-swift-container returned with an unexpected error.	Check the service script and settings.
container-updater_failed	STATE_CHANGE	ERROR	The status of the container-updater process must be {0} but it is {1} now.	The container-updater process is not in the expected state.	The container-updater process is expected to be running on the singleton node only.	Check the status of openstack-swift-container-updater process and object singleton flag.
container-updater_ok	STATE_CHANGE	INFO	The status of the container-updater process is {0} as expected.	The container-updater process is in the expected state.	The container-updater process is expected to be running on the singleton node only.	N/A
container-updater_warn	INFO	WARNING	The container-updater process monitoring service returned unknown result.	The container-updater check returned an unknown result.	A status query for openstack-swift-container-updater returned with an unexpected error.	Check the service script and settings.
disable_Address_database_node	INFO	INFO	An address database node is disabled.	Database flag is removed from this node.	A CES IP with a database flag linked to it is either removed from this node or moved to this node.	N/A
disable_Address_singleton_node	INFO	INFO	An address singleton node is disabled.	Singleton flag is removed from this node.	A CES IP with a singleton flag linked to it is either removed from this node or moved from/to this node.	N/A

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
enable_Address_database_node	INFO	INFO	An address database node is enabled.	The database flag is moved to this node.	A CES IP with a database flag linked to it is either removed from this node or moved from/to this node.	N/A
enable_Address_singleton_node	INFO	INFO	An address singleton node is enabled.	The singleton flag is moved to this node.	A CES IP with a singleton flag linked to it is either removed from this node or moved from/to this node.	N/A
ibmobjectizer_failed	STATE_CHANGE	ERROR	The status of the ibmobjectizer process must be {0} but it is {1} now.	The ibmobjectizer process is not in the expected state.	The ibmobjectizer process is expected to be running on the singleton node only.	Check the status of the ibmobjectizer process and object singleton flag.
ibmobjectizer_ok	STATE_CHANGE	INFO	The status of the ibmobjectizer process is {0} as expected.	The ibmobjectizer process is in the expected state.	The ibmobjectizer process is expected to be running on the singleton node only.	N/A
ibmobjectizer_warn	INFO	WARNING	The ibmobjectizer process monitoring service returned unknown result	The ibmobjectizer check returned an unknown result.	A status query for ibmobjectizer returned with an unexpected error.	Check the service script and settings.
memcached_failed	STATE_CHANGE	ERROR	The status of the memcached process must be {0} but it is {1} now.	The memcached process is not running.	The memcached process is not running.	Check the status of memcached process.
memcached_ok	STATE_CHANGE	INFO	The status of the memcached process is {0} as expected.	The memcached process is running.	The memcached process is running.	N/A
memcached_warn	INFO	WARNING	The memcached process monitoring service returned unknown result.	The memcached check returned an unknown result.	A status query for memcached returned with an unexpected error.	Check the service script and settings.
obj_restart	INFO	WARNING	The {0} service is failed. Trying to recover.	An object service was not in the expected state.	An object service might have stopped unexpectedly.	None, recovery is automatic.
object-expirer_failed	STATE_CHANGE	ERROR	The status of the object-expirer process must be {0} but it is {1} now.	The object-expirer process is not in the expected state.	The object-expirer process is expected to be running on the singleton node only.	Check the status of openstack-swift-object-expirer process and object singleton flag.
object-expirer_ok	STATE_CHANGE	INFO	The status of the object-expirer process is {0} as expected.	The object-expirer process is in the expected state.	The object-expirer process is expected to be running on the singleton node only.	N/A
object-expirer_warn	INFO	WARNING	The object-expirer process monitoring service returned unknown result.	The object-expirer check returned an unknown result.	A status query for openstack-swift-object-expirer returned with an unexpected error.	Check the service script and settings.
object-replicator_failed	STATE_CHANGE	ERROR	The status of the object-replicator process must be {0} but it is {1} now.	The object-replicator process is not running.	The object-replicator process is not running.	Check the status of openstack-swift-object-replicator process.

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
object-replicator_ok	STATE_CHANGE	INFO	The status of the object-replicator process is {0} as expected.	The object-replicator process is running.	The object-replicator process is running.	N/A
object-replicator_warn	INFO	WARNING	The object-replicator process monitoring service returned unknown result.	The object-replicator check returned an unknown result.	A status query for openstack-swift-object-replicator returned with an unexpected error.	Check the service script and settings.
object-server_failed	STATE_CHANGE	ERROR	The status of the object-server process must be {0} but it is {1} now.	The object-server process is not running.	The object-server process is not running.	Check the status of the openstack-swift-object process.
object-server_ok	STATE_CHANGE	INFO	The status of the object-server process is {0} as expected.	The object-server process is running.	The object-server process is running.	N/A
object-server_warn	INFO	WARNING	The object-server process monitoring service returned unknown result.	The object-server check returned an unknown result.	A status query for openstack-swift-object-server returned with an unexpected error.	Check the service script and settings.
object-updater_failed	STATE_CHANGE	ERROR	The status of the object-updater process must be {0} but it is {1} now.	The object-updater process is not in the expected state.	The object-updater process is expected to be running on the singleton node only.	Check the status of the openstack-swift-object-updater process and object singleton flag.
object-updater_ok	STATE_CHANGE	INFO	The status of the object-updater process is {0} as expected.	The object-updater process is in the expected state.	The object-updater process is expected to be running on the singleton node only.	N/A
object-updater_warn	INFO	WARNING	The object-updater process monitoring returned unknown result.	The object-updater check returned an unknown result.	A status query for openstack-swift-object- updater returned with an unexpected error.	Check the service script and settings.
openstack-object-sof_failed	STATE_CHANGE	ERROR	The status of the object-sof process must be {0} but is {1}.	The swift-on-file process is not in the expected state.	The swift-on-file process is expected to be running then the capability is enabled and stopped when disabled.	Check the status of the openstack-swift-object-sof process and capabilities flag in spectrum-scale -object.conf.
openstack-object-sof_ok	STATE_CHANGE	INFO	The status of the object-sof process is {0} as expected.	The swift-on-file process is in the expected state.	The swift-on-file process is expected to be running then the capability is enabled and stopped when disabled.	N/A
openstack-object-sof_warn	INFO	INFO	The object-sof process monitoring returned unknown result.	The openstack-swift-object-sof check returned an unknown result.	A status query for openstack-swift-object-sof returned with an unexpected error.	Check the service script and settings.
postIpChange_info_o	INFO	INFO	The following IP addresses are modified: {0}	CES IP addresses have been moved and activated.		N/A
proxy-server_failed	STATE_CHANGE	ERROR	The status of the proxy process must be {0} but it is {1} now.	The proxy-server process is not running.	The proxy-server process is not running.	Check the status of the openstack-swift-proxy process.

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
proxy-server_ok	STATE_CHANGE	INFO	The status of the proxy process is {0} as expected.	The proxy-server process is running.	The proxy-server process is running.	N/A
proxy-server_warn	INFO	WARNING	The proxy-server process monitoring returned unknown result.	The proxy-server process monitoring returned an unknown result.	A status query for openstack-swift-proxy-server returned with an unexpected error.	Check the service script and settings.
ring_checksum_failed	STATE_CHANGE	ERROR	Checksum of the ring file {0} does not match the one in CCR.	Files for object rings have been modified unexpectedly.	Checksum of file did not match the stored value.	Check the ring files.
ring_checksum_ok	STATE_CHANGE	INFO	Checksum of the ring file {0} is OK.	Files for object rings were successfully checked.	Checksum of file found unchanged.	N/A
ring_checksum_warn	INFO	WARNING	Issue while checking checksum of the ring file {0}.	Checksum generation process failed.	The ring_checksum check returned an unknown result.	Check the ring files and the md5sum executable.
proxy-httpd-server_failed	STATE_CHANGE	ERROR	Proxy process should be {0} but is {1}.	The proxy-server process is not running.	The proxy-server process is not running.	Check status of openstack-swift-proxy process.
proxy-httpd-server_ok	INFO	INFO	Proxy process as expected, state is {0}.	The proxy-server process is running.	The proxy-server process is running.	N/A
proxy_access_up	STATE_CHANGE	INFO	Access to proxy service ip {0} port {1} ok.	The access check of the proxy service port was successful.		N/A
proxy_access_down	STATE_CHANGE	ERROR	No access to proxy service ip {0} port {1}. Check firewall.	The access check of the proxy service port failed.	The port is probably blocked by a firewall rule.	Check if the proxy service is running, and the firewall rules.
proxy_access_warn	STATE_CHANGE	WARNING	Proxy service access check ip {0} port {1} failed. Check for validity.	The access check of the proxy service port returned an unknown result.	The proxy service port access could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
account_access_up	STATE_CHANGE	INFO	Access to account service ip {0} port {1} ok.	The access check of the account service port was successful.		N/A
account_access_down	STATE_CHANGE	ERROR	No access to account service ip {0} port {1}. Check firewall.	The access check of the account service port failed.	The port is probably blocked by a firewall rule	Check if the account service is running and the firewall rules
account_access_warn	INFO	WARNING	Account service access check ip {0} port {1} failed. Check for validity.	The access check of the account service port returned an unknown result.	The account service port access could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
container_access_up	STATE_CHANGE	INFO	Access to container service ip {0} port {1} ok.	The access check of the container service port was successful.		N/A

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
container_access_down	STATE_CHANGE	ERROR	No access to container service ip {0} port {1}. Check firewall.	The access check of the container service port failed.	The port is probably blocked by a firewall rule.	Check if the filesystem daemon is running, and the firewall rules.
container_access_warn	INFO	WARNING	Container service access check ip {0} port {1} failed. Check for validity.	The access check of the container service port returned an unknown result.	The container service port access could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
object_access_up	STATE_CHANGE	INFO	Access to object store ip {0} port {1} ok.	The access check of the object service port was successful.		N/A
object_access_down	STATE_CHANGE	ERROR	No access to object store ip {0} port {1}. Check firewall.	The access check of the object service port failed.	The port is probably blocked by a firewall rule.	Check if the object service is running, and the firewall rules.
object_access_warn	INFO	WARNING	Object store access check ip {0} port {1} failed.	The access check of the object service port returned an unknown result.	The object service port access could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
object_sof_access_up	STATE_CHANGE	INFO	Access to unified object store ip {0} port {1} ok.	The access check of the unified object service port was successful.		N/A
object_sof_access_down	STATE_CHANGE	ERROR	No access to unified object store ip {0} port {1}. Check firewall.	The access check of the unified object service port failed.	The port is probably blocked by a firewall rule.	Check if thunified object service is running, and the firewall rules.
object_sof_access_warn	INFO	WARNING	Unified object store access check ip {0} port {1} failed. Check for validity.	The access check of the unified object access service port returned an unknown result.	The unified object service port access could not be determined due to a problem.	Find potential issues for this kind of failure in the logs.
stop_obj_service	INFO_EXTERNAL	INFO	OBJ service was stopped.	Information about an OBJ service stop.	The OBJECT service was stopped (e.g. using the **mmces service stop obj** command).	N/A
start_obj_service	INFO_EXTERNAL	INFO	OBJ service was started.	information about a OBJ service start.	The OBJECT service was started (e.g. using the **mmces service start obj** command).	N/A
object_quarantined	INFO_EXTERNAL	WARNING	The object \"{0}\", container \"{1}\", account \"{2}\" has been quarantined. Path of quarantined object: \"{3}\.	The object which was being accessed is quarantined.	Mismatch in data or metadata.	

Table 80. Events for the object component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
openstack-swift-object-auditor_failed	STATE_CHANGE	ERROR	Object-auditor process should be {0} but is {1}.	The object-auditor process is not in the expected state.	The openstack-swift-object-auditor process is expected to be running on the singleton node only, and when the capability multi-region is enabled. It needs to be stopped in other cases	Check status of openstack-swift-object-auditor process and capabilities flag in `spectrum-scale-object.conf`
openstack-swift-object-auditor_ok	STATE_CHANGE	INFO	Object-auditor process as expected, state is {0}}.	The object-auditor process is in the expected state.	The openstack-swift-object-auditor process is expected to be running on the singleton node only, and when the capability multi-region is enabled. It needs to be stopped in other cases.	N/A
openstack-swift-object-auditor_warn	INFO	INFO	Object-auditor process monitoring returned unknown result.	The openstack-swift-object-auditor check returned an unknown result.	A status query for openstack-swift-object-auditor returned with an unexpected error.	Check service script and settings.

Performance events

The following table lists the events that are created for the *Performance* component.

Table 81. Events for the Performance component

Event	EventType	Severity	Message	Description	Cause	User Action
pmcollector_down	STATE_CHANGE	ERROR	The status of the pmcollector service must be {0} but it is {1} now.	The performance monitoring collector is down.	Performance monitoring is configured in this node but the pmcollector service is currently down.	Use the **systemctl start pmcollector** command to start the performance monitoring collector service.
pmsensors_down	STATE_CHANGE	ERROR	The status of the pmsensors service must be {0} but it is {1}now.	The performance monitor sensors are down.	Performance monitoring service is configured on this node but the performance sensors are currently down.	Use the **systemctl start pmsensors** command to start the performance monitoring sensor service or remove the node from the global performance monitoring configuration by using the **mmchnode** command.

Table 81. Events for the Performance component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
pmsensors_up	STATE_CHANGE	INFO	The status of the pmsensors service is {0} as expected.	The performance monitor sensors are running.	The performance monitoring sensor service is running as expected.	N/A
pmcollector_up	STATE_CHANGE	INFO	The status of the pmcollector service is {0} as expected.	The performance monitor collector is running.	The performance monitoring collector service is running as expected.	N/A
pmcollector_warn	INFO	INFO	The pmcollector process returned unknown result.	The monitoring service for performance monitor collector returned an unknown result.	The monitoring service for performance monitoring collector returned an unknown result.	Use the **service** or **systemctl** command to verify whether the performance monitoring collector service is in the expected status. If there is no pmcollector service running on the node and the performance monitoring service is configured on the node, check with the *Performance monitoring* section in the IBM Spectrum Scale documentation.

Table 81. Events for the Performance component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
pmsensors_warn	INFO	INFO	The pmsensors process returned unknown result.	The monitoring service for performance monitor sensors returned an unknown result.	The monitoring service for performance monitoring sensors returned an unknown result.	Use the **service** or **systemctl** command to verify whether the performance monitoring sensor is in the expected status. Perform the troubleshooting procedures if there is no pmcollector service running on the node and the performance monitoring service is configured on the node. For more information, see the *Performance monitoring* section in the IBM Spectrum Scale documentation.

SMB events

The following table lists the events that are created for the *SMB* component.

Table 82. Events for the SMB component

Event	EventType	Severity	Message	Description	Cause	User Action
ctdb_down	STATE_CHANGE	ERROR	CTDB process not running.	The CTDB process is not running.		Perform the troubleshooting procedures.
ctdb_recovered	STATE_CHANGE	INFO	CTDB Recovery finished.	CTDB completed database recovery.		N/A
ctdb_recovery	STATE_CHANGE	WARNING	CTDB recovery detected.	CTDB is performing a database recovery.		N/A
ctdb_state_down	STATE_CHANGE	ERROR	CTDB state is {0}.	The CTDB state is unhealthy.		Perform the troubleshooting procedures.
ctdb_state_up	STATE_CHANGE	INFO	CTDB state is healthy.	The CTDB state is healthy.		N/A
ctdb_up	STATE_CHANGE	INFO	CTDB process now running.	The CTDB process is running.		N/A
ctdb_warn	INFO	WARNING	CTDB monitoring returned unknown result.	The CTDB check returned unknown result.		Perform the troubleshooting procedures.

Table 82. Events for the SMB component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
smb_restart	INFO	WARNING	The SMB service is failed. Trying to recover.	Attempt to start the SMBD process.	The SMBD process was not running.	N/A
smbd_down	STATE_CHANGE	ERROR	SMBD process not running.	The SMBD process is not running.		Perform the troubleshooting procedures.
smbd_up	STATE_CHANGE	INFO	SMBD process now running.	The SMBD process is running.		N/A
smbd_warn	INFO	WARNING	The SMBD process monitoring returned unknown result.	The SMBD process monitoring returned an unknown result.		Perform the troubleshooting procedures.
smbport_down	STATE_CHANGE	ERROR	The SMB port {0} is not active.	SMBD is not listening on a TCP protocol port.		Perform the troubleshooting procedures.
smbport_up	STATE_CHANGE	INFO	The SMB port {0} is now active.	An SMB port was activated.		N/A
smbport_warn	INFO	WARNING	The SMB port monitoring {0} returned unknown result.	An internal error occurred while monitoring SMB TCP protocol ports.		Perform the troubleshooting procedures.
stop_smb_service	INFO_EXTERNAL	INFO	SMB service was stopped.	Information about an SMB service stop.	The SMB service was stopped . For example, using **mmces service stop smb**.	N/A
start_smb_service	INFO_EXTERNAL	INFO	SMB service was started.	Information about an SMB service start.	The SMB service was started . For example, using **mmces service start smb**.	N/A
smb_sensors_active	TIP	INFO	The SMB perfmon sensors are active.	The SMB perfmon sensors are active. This event's monitor is only running once an hour.	The SMB perfmon sensors' period attribute is greater than 0.	N/A

Table 82. Events for the SMB component (continued)

Event	EventType	Severity	Message	Description	Cause	User Action
smb_sensors_inactive	TIP	TIP	The following SMB perfmon sensors are inactive: {0}.	The SMB perfmon sensors are inactive. This event's monitor is only running once an hour.	The SMB perfmon sensors' period attribute is 0.	Set the period attribute of the SMB sensors to a value greater than 0. Use the following command: `mmperfmon config update SensorName.period=N` , where **SensorName** is one of the SMB sensors' name, and *N* is a natural number greater than 0. This TIP monitor is only running only once per hour, and might take up to one hour in worst case to detect the changes in the configuration.
ctdb_version_mismatch	ERROR_EXTERNAL	ERROR	CTDB could not start on CES node {0} as its version is {1} but version {2} is already running in the cluster.	CTDB cannot start up on a node as it has detected that on other CES nodes a CTDB cluster is running at a different version. This prevents the SMB service to get healthy.	CTDB cannot start up on a node as it detected a conflicting version running in the cluster. The name of the failing node, the starting CTDB version, and the conflicting version running in the cluster are provided.	Get all gpfs-smb packages to the same version
ctdb_version_match	INFO_EXTERNAL	INFO	CTDB has passed the version check on CES node {0}	The CTDB service has successfully passed the version check	The CTDB service has successfully passed the version check on a node and can join the running cluster. The node is given.	N/A

Threshold events

The following table lists the events that are created for the *threshold* component.

Table 83. Events for the threshold component

Event	EventType	Severity	Message	Description	Cause	User Action
reset_threshold	INFO	INFO	Requesting current threshold states.	Sysmon restart detected, requesting current threshold states.		N/A
thresholds_new_rule"	INFO_ADD_ENTITY	INFO	Rule {0} was added.	A threshold rule was added.		N/A
thresholds_del_rule	INFO_DELETE_ENTITY	INFO	Rule {0} was removed.	A threshold rule was removed.		N/A
thresholds_normal	STATE_CHANGE	INFO	The value of {1} defined in {2} for component {id} reached a normal level.	The thresholds value reached a normal level.		N/A
thresholds_error	STATE_CHANGE	ERROR	The value of {1} for the component(s) {id} exceeded threshold error level {0} defined in {2}.	The thresholds value reached an error level.		N/A
thresholds_warn	STATE_CHANGE	WARNING	The value of {1} for the component(s) {id} exceeded threshold warning level {0} defined in {2}.	The thresholds value reached a warning level.		N/A
thresholds_removed	STATE_CHANGE	INFO	The value of {1} for the component(s) {id} defined in {2} was removed.	The thresholds value could not be determined.	No component usage data in performance monitoring.	N/A
thresholds_no_data	STATE_CHANGE	INFO	The value of {1} for the component(s) {id} defined in {2} return no data.	The thresholds value could not be determined.	The thresholds value could not be determined.	N/A
thresholds_no_rules	STATE_CHANGE	INFO	No thresholds defined.	No thresholds defined.	No thresholds defined.	N/A

Transparent cloud tiering status description

This topic describes the various statuses and their description that is associated with the health of Cloud services running on each node in the cluster.

Table 84. Cloud services status description

Entity	Status	Description	Comments
TCT Account Status	Not configured	The Transparent cloud tiering installed, but the account is not configured or the account is deleted.	Run the **mmcloudgateway account create** command to create the cloud provider account.
	Active	The cloud provider account that is configured with Transparent cloud tiering service is active.	
	Configured	The cloud provider account is configured with Transparent cloud tiering, but the service is down.	Run the **mmcloudgateway service start** command to resume the cloud gateway service.
	unreachable	The cloud provider access point URL is unreachable due to either it being down or network issues.	Make sure that the cloud provider is online. Check if the network is reachable between the cloud provider and Transparent cloud tiering. Also, check the DNS settings. Check the trace messages and error log for further details.
	invalid csp endpoint URL	The reason might be because of an HTTP 404 Not Found error.	Check whether the configured cloud provider URL is valid.
	malformed_URL	The cloud provider access point URL is malformed.	Check if the cloud provider URL is valid. Check whether the URL has proper legal protocol such as http or https. Check if the URL has space or any special characters that cannot be parsed.
	no_route_to_csp	The response from the cloud storage access point is invalid.	Check the following: • DNS and firewall settings • Network is reachable between Transparent cloud tiering and the cloud provider
	connect_exception	The connection is refused remotely by the CSAP.	Check the following: • Network is reachable between Transparent cloud tiering and the cloud provider • Cloud provider URL is valid. • Cloud provider is listening for connections
	socket_timeout	Timeout occurred on a socket while it was connecting to the cloud provider.	Check if network is reachable between Transparent cloud tiering and the cloud provider.
	invalid_cloud_config	Transparent cloud tiering refuses to connect to the CSAP.	Check if the cloud object store is configured correctly. In case of Swift cloud provider, check whether both Keystone and Swift provider configuration are proper. Also, check if Swift is reachable over Keystone.
	credentials_invalid	The Transparent cloud tiering service fails to connect to CSAP because of failed authentication.	Check if the Access Key and Secret Key are valid. Check if the user name and password are correct.
	mcstore_node_network_down	The network of the Transparent cloud tiering node is down.	Check if network interface on the Transparent cloud tiering node is proper and is able to communicate with public and private networks.

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
TCT Account Status	ssl_handshake_exception	The CSAP fails due to an unknown SSL handshake error.	Do one or more of the following: • Check if the cloud provider supports secure communication and is properly configured with certificate chain. • Check whether provided cloud provider URL is secure (https) • Run this command on a Transparent cloud tiering node to check secure connection to cloud provider: **openssl s_client -connect <cloud_provider_ipaddress>:<secured_port>**
	SSL handshake certificate exception	Transparent cloud tiering failed to connect to the CSAP because of an untrusted server certificate chain or certificate is invalid.	Check whether server certificate is not expired, still valid, properly DER encoded. Make sure that a self-signed or internal CA signed certificate is added to Transparent cloud tiering trust store. Use `–server-cert-path` option to add a self-signed certificate. Check the server certificate validity by using the `openssl x509 -in <server_cert> -text -noout` command. Check in the **Validity** section and make sure that the certificate is not expired and still valid.
	ssl handshake sock closed exception	Transparent cloud tiering fails to connect to the CSAP because the remote host closed connection during the handshake.	Do the following: • Check network connection and make sure secure port is reachable. • Run this command on the Transparent cloud tiering node and check if there is secure connection to the cloud provider: `openssl s_client -connect <cloud_provider_ipaddress>:<secured_port>`
	ssl handshake bad certificate exception	Transparent cloud tiering fails to connect to the CSAP because the server certificate does not exist the trust store.	Ensure that a self-signed or internal CA-signed certificate is properly added to the Transparent cloud tiering trust store. Use the `–server-cert-path` option to add a self-signed certificate.
	ssl handshake invalid path certificate exception	Transparent cloud tiering fails to connect to the CSAP because it is unable to find a valid certification path.	Do the following: • Make sure that proper self-signed or internal CA-signed certificate is added to the Transparent cloud tiering trust store. Use the `–server-cert-path` option to add a self-signed certificate. • Make sure that the Client (Transparent cloud tiering) and Server (cloud provider) certificate are signed with same Certificate Authority or configured with same self-signed certificate. • In case of a load balancer/firewall, make sure that proper cloud

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
TCT Account Status	ssl not trusted certificate exception	Transparent cloud tiering failed to connect to the cloud provider because it could not locate a trusted server certificate.	Check the trace messages and error logs for further details.
	ssl invalid algorithm exception	Transparent cloud tiering failed to connect to the cloud provider because of invalid or inappropriate SSL algorithm parameters.	Check the trace messages and error logs for further details.
	ssl invalid padding exception	Transparent cloud tiering failed to connect to the cloud provider because of invalid SSL padding.	Check the trace messages and error logs for further details.
	ssl unrecognized message	Transparent cloud tiering failed to connect cloud provider because of an unrecognized SSL message.	Check the trace messages and error logs for further details.
	bad request	Transparent cloud tiering failed to connect to the cloud provider because of a request error.	Check the trace messages and error logs for further details.
	precondition failed	Transparent cloud tiering failed to connect to the cloud provider because of a precondition failed error.	Check the trace messages and error logs for further details.
	default exception	The cloud provider account is not accessible due to an unknown error.	Check the trace messages and error logs for further details.
	container create failed	The cloud provider container creation failed. The cloud provider account might not be authorized to create the container.	Check the trace messages and error logs for further details. Also, check the account create related issues in *Transparent cloud tiering issues* in the *IBM Spectrum Scale: Problem Determination Guide*.
	time skew	The time observed on the Transparent cloud tiering service node is not in sync with the time on the target cloud provider.	Change the Transparent cloud tiering service node time stamp to be in sync with the NTP server and rerun the operation.
	server error	Transparent cloud tiering failed to connect to the cloud provider because of a cloud provider server error (HTTP 503) or the container size reached max storage limit.	Check the trace messages and error logs for further details.
	internal dir not found	Transparent cloud tiering failed because one of its internal directory is not found.	Check the trace messages and error logs for further details.
	db corrupted	The database of Transparent cloud tiering service is corrupted.	Check the trace messages and error logs for further details. Use the **mmcloudgateway files rebuildDB** command to repair it.
TCT File system Status	Not configured	Transparent cloud tiering installed, but the file system is not configured or it was deleted.	Run the **mmcloudgateway filesystem create** command to configure the file system.
	Configured	The Transparent cloud tiering is configured with a file system.	

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
TCT Server Status	Stopped	Transparent cloud tiering service is stopped by CLI command or stopped itself due to some error.	Run the **mmcloudgateway service start** command to start the cloud gateway service.
	Suspended	The cloud service was suspended manually.	Run the **mmcloudgateway service start** command to resume the cloud gateway service.
	Started	The cloud gateway service is running.	
	Not configured	Transparent cloud tiering was either not configured or its services were never started.	Set up the Transparent cloud tiering and start the service.
Security	rkm down	The remote key manager configured for Transparent cloud tiering is not accessible.	Check the trace messages and error logs for further details.
	lkm down	The local key manager who is configured for Transparent cloud tiering is either not found or corrupted.	Check the trace messages and error logs for further details.

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
Cloud Storage Access Point	ONLINE	Cloud storage access point that is configured with Transparent cloud tiering service is active.	The state when the cloud storage access point is reachable, and credentials are valid.
	OFFLINE	The cloud storage access point URL is unreachable due to either it is down or network issues.	`unreachable` The cloud provider end point URL is not reachable.
	OFFLINE	The reason could be because of `HTTP 404 Not Found` error.	`invalid_csp_endpoint_url` The specified end point URL is invalid.
	OFFLINE	The cloud storage access point URL is malformed.	`malformed_url` The URL is malformed.
	OFFLINE	The response from cloud storage access point is invalid.	`no_route_to_csp` There is no route to the CSP, which indicates a possible a firewall issue.
	OFFLINE	The connection is refused remotely by the cloud storage access point.	`connect_exception` There is a connection exception when you connect.
	OFFLINE	Timeout occurs on a socket while it connects to the cloud storage access point URL.	`SOCKET_TIMEOUT` The socket times out.
	OFFLINE	Transparent cloud tiering refuses to connect to the cloud storage access point.	`invalid_cloud_config` There is an invalid Cloud configuration, check the cloud.
	OFFLINE	The Transparent cloud tiering service fails to connect cloud storage access point because the authentication fails.	`credentials_invalid` The credentials that are provided during **`addCloud`** are no longer valid (including if the password is expired).
	OFFLINE	The network of Transparent cloud tiering node is down.	`mcstore_node_network_down` The network is down on the `mcstore` node.
	OFFLINE	The cloud storage access point status fails due to an unknown SSL handshake error.	`ssl_handshake_exception` There is an SSL handshake exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of an untrusted server certificate chain.	`ssl_handshake_cert_exception` There is an SSL handshake certificate exception.

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
Cloud Storage Access Point	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because the remote host closed connection during handshake.	`ssl_handshake_sock_closed_exception` There is an SSL handshake socket closed exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of a bad certificate.	`ssl_handshake_bad_cert_exception` There is an SSL handshake bad certificate exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because it is unable to find a valid certification path.	`ssl_handshake_invalid_path_cert _exception` There is an SSL handshake invalid path certificate exception.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because it could not negotiate the desired level of security.	`ssl_handshake_failure_exception` There is an SSL handshake failure exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of an unknown certificate.	`ssl_handshake_unknown_cert _exception` There is an SSL handshake unknown certificate exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of a bad SSL key or misconfiguration.	`ssl_key_exception` There is an SSL key exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because its identity is not verified.	`ssl_peer_unverified_exception` There is an SSL peer unverified exception.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because of an SSL protocol operation error.	`ssl_protocol_exception` There is an SSL protocol exception.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because of an SSL subsystem error.	`ssl_exception` There is an SSL exception.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because there is no available certificate.	`ssl_no_cert_exception` There is an SSL not trusted certificate exception.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because the server certificate is not trusted.	`ssl_not_trusted_cert_exception` There is an SSL not trusted certificate exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of invalid or inappropriate SSL algorithm parameters.	`ssl_invalid_algo_exception` There is an SSL invalid algorithm exception.

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
Cloud Storage Access Point	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of invalid SSL padding.	`ssl_invalid_padding_exception` There is an SSL invalid padding exception.
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because of an unrecognized SSL message.	`ssl_unrecognized_msg` There is an SSL unrecognized message exception.
	OFFLINE	Transparent cloud tiering fails to the connect cloud storage access point because of a request error.	`bad_request` This is a bad request.
	OFFLINE	Transparent cloud tiering fails to connect to the cloud storage access point because of a precondition failed error.	`precondition_failed` The pre-condition failed.
	OFFLINE	The cloud storage access point account is not accessible due to unknown error.	`default` Some unchecked exception is caught.
	OFFLINE	The cloud provider container creation fails.	`CONTAINER_CREATE_FAILED` The container creation failed.
	OFFLINE	The cloud provider container creation fails because it exists.	`CONTAINER_ALREADY_EXISTS` The container creation failed.
	OFFLINE	The cloud provider container creation fails because it exceeds the maximum limit.	`BUCKET_LIMIT_EXCEEDED` The container creation failed.
	OFFLINE	The cloud provider container does not exist.	CONTAINER_DOES_NOT_EXIST cloud operations failed
	OFFLINE	The time observed on the Transparent cloud tiering service node is not in sync with the time observed on target cloud storage access point.	`TIME_SKEW`
	OFFLINE	Transparent cloud tiering fails to connect cloud storage access point because the cloud storage access point server error or container size reaches the maximum storage limit.	`SERVER_ERROR`
	OFFLINE	Transparent cloud tiering fails because one of its internal directories is not found.	The internal directory is not found: `internal_dir_notfound`

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
Cloud Storage Access Point	OFFLINE	Transparent cloud tiering failed because the resource address file is not found.	The resource address file is not found: `RESOURCE_ADDRESSES_FILE_NOT_FOUND`
	OFFLINE	The database of Transparent cloud tiering service is corrupted	The Transparent cloud tiering service is corrupted: `db_corrupted`
	OFFLINE	The remote key manager who is configured for Transparent cloud tiering is not accessible.	The SKLM server is not accessible. This is valid only when ISKLM is configured
	OFFLINE	The local key manager who is configured for Transparent cloud tiering is either not found or corrupted.	The local `.jks` file is either not found or corrupted. This is valid only when local key manager is configured.
	OFFLINE	The reason could be because of an `HTTP 403 Forbidden` error.	This action is forbidden.
	OFFLINE	Access denied due to authorization error.	Access is denied.
	OFFLINE	The filesystem of Transparent cloud tiering service is corrupted.	The filesystem is corrupted.
	OFFLINE	The directory of Transparent cloud tiering service is corrupted.	There is a directory error.
	OFFLINE	The key manager configured for Transparent cloud tiering is either not found or corrupted.	There is a key manager error.
	OFFLINE	Transparent cloud tiering failed because its container pair root directory not found.	The container pair root directory is not found.
		Transparent cloud tiering service is has too many retries internally.	There are too many retries.
Cloud Service	ENABLED	The cloud service is enabled for cloud operations.	The Cloud service is enabled,
	DISABLED	The cloud service is disabled.	The Cloud service is disabled.
TCT Service	STOPPED	The cloud gateway service is down and could not be started	The server is stopped abruptly, for example, JVM crash.
	SUSPENDED	The cloud gateway service is suspended manually.	The TCT service is stopped intentionally.
	STARTED	The cloud gateway service is up and running.	The TCT service is started up and running.
		The cloud gateway check returns an unknown result.	The TCT service status is unknown value.
		Attempt to restart the cloud gateway process.	The TCT service is restarted.
	NOT_CONFIGURED	The Transparent cloud tiering was either not configured or its service was never started.	The TCT service is recently installed.

Table 84. Cloud services status description (continued)

Entity	Status	Description	Comments
Container pair status	ONLINE	The container pair set is configured with an active Transparent cloud tiering service.	The state when the container pair set is reachable.
	OFFLINE	Transparent cloud tiering failed to connect to the container pair set because of a directory error.	Connection failure due to: • `directory_error` • TCT directory error
		Transparent cloud tiering failed to connect to the container pair set because the container does not exist.	`container_does_not_exist` The cloud container cannot be found but the connection to the cloud is active.
		Transparent cloud tiering failed to connect to the cloud provider because of precondition failed error.	Check trace messages and error logs for further details.
		Transparent cloud tiering failed to connect because resource address file is not found.	Check trace messages and error logs for further details.
		Transparent cloud tiering failed to connect because one of its internal directory is not found.	Check trace messages and error logs for further details.
		Transparent cloud tiering failed because one of its internal directory is not found.	Check trace messages and error logs for further details.
		Transparent cloud tiering failed to connect cloud storage access point because of cloud storage access point service unavailability error.	Check trace messages and error logs for further details.
		Transparent cloud tiering failed to connect cloud storage access point because of a bad certificate.	Check trace messages and error logs for further details.

Cloud services audit events

Every operation that is performed using Cloud services is audited and recorded to meet the regulatory requirements.

The audit details are saved in this file: `/var/MCStore/ras/audit/audit_events.json`. You can parse this JSON file by using some tools and use it for troubleshooting purposes. For the complete list of events, see Table 85. This is an example of the audit entry that is added to the JSON file after you successfully execute the **`mmcloudgateway account create`** command.

```
{"typeURI":"http://schemas.dmtf.org/cloud/audit/1.0/event","eventType":"activity","id":
"b4e9a5a9-0bf7-45ee-9e93-b6f825781328","eventTime":"2017-08-21T18:46:10.439 UTC","action":
"create/create_cloudaccount","outcome":"success","initiator":{"id":"b22ec254-d645-43c4-a402-3e15757d8463",
"typeURI":"data/security/account/admin","name":"root","host":{"address":"192.0.2.0"}},"target":
{"id":"58347894-6a10-4218-a66d-357e4a3f4aaf","typeURI":"service/storage/object/account","name":
"tct.cloudstorageaccesspoint"},"observer":{"id":"target"},"attachments":[{"content":"account-name=
swift-account, cloud-type=openstack-swift, username=admin, tenant=admin, src-keystore-path=null,
src-alias-name=null, src-keystore-type=null","name":"swift-account","contentType":"text"}]}
```

Table 85. Audit events

S.No	Events
1	Add a cloud account – Success

Table 85. Audit events (continued)

S.No	Events
2	Add a cloud account – Failure
3	Create a cloud storage access point – Success
4	Create a cloud storage access point – Failure
5	Create a cloud service – Success
6	Create a cloud service – Failure
7	Create a container pair – Success
8	Create a container pair – Failure
9	Create a key manager – Success
10	Create a key manager – Failure
11	Update a cloud account – Success
12	Update a cloud account - Failure
13	Update cloud storage access point – Success
14	Update cloud storage access point – Failure
15	Update cloud service – Success
16	Update cloud service – Failure
17	Update container pair – Success
18	Update container pair – Failure
19	Update key manager – Success
20	Update key manager – Failure
21	Delete a cloud account – Success
22	Delete a cloud account - Failure
23	Delete cloud storage access point – Success
24	Delete cloud storage access point – Failure
25	Delete cloud service – Success
26	Delete cloud service – Failure
27	Delete container pair – Success
28	Delete container pair – Failure
29	Rotate key manager – Success
30	Rotate key manager – Failure
31	Clean up orphan objects from the orphan table
32	Cloud destroy – Success
33	Cloud destroy – Failure
34	Export files – Success
35	Export files – Failure
36	Import files – Success
37	Import files -Failure
38	Migrate files – Success
39	Migrate files – Failure
40	Recall files – Success
41	Recall files – Failure

Table 85. Audit events (continued)

S.No	Events
42	Remove cloud objects – Success
43	Remove cloud objects – Failure
44	Reconcile files – Success
45	Reconcile files – Failure
46	Rebuild DB – Success
47	Rebuild DB - Failure
48	Restore files – Success
49	Restore files – Failure
50	Run policy (lwe destroy) – Success
51	Run policy (lwe destroy) – Failure
52	Config set – Success
53	Config set - Failure

Messages

This topic contains explanations for GPFS error messages.

Messages for IBM Spectrum Scale RAID in the ranges 6027-1850 – 6027-1899 and 6027-3000 – 6027-3099 are documented in *IBM Spectrum Scale RAID: Administration*.

Message severity tags

GPFS has adopted a message severity tagging convention. This convention applies to some newer messages and to some messages that are being updated and adapted to be more usable by scripts or semi-automated management programs.

A severity tag is a one-character alphabetic code (**A** through **Z**), optionally followed by a colon (**:**) and a number, and surrounded by an opening and closing bracket (**[ ]**). For example:

[E] or **[E:*nnn*]**

If more than one substring within a message matches this pattern (for example, **[A]** or [**A:***nnn*]), the severity tag is the first such matching string.

When the severity tag includes a numeric code (*nnn*), this is an error code associated with the message. If this were the only problem encountered by the command, the command return code would be *nnn*.

If a message does not have a severity tag, the message does not conform to this specification. You can determine the message severity by examining the text or any supplemental information provided in the message catalog, or by contacting the IBM Support Center.

Each message severity tag has an assigned priority that can be used to filter the messages that are sent to the error log on Linux. Filtering is controlled with the **mmchconfig** attribute **systemLogLevel**. The default for **systemLogLevel** is **error**, which means GPFS will send all error **[E]**, critical **[X]**, and alert **[A]** messages to the error log. The values allowed for **systemLogLevel** are: **alert**, **critical**, **error**, **warning**, **notice**, **configuration**, **informational**, **detail**, or **debug**. Additionally, the value **none** can be specified so no messages are sent to the error log.

Alert **[A]** messages have the highest priority, and debug **[B]** messages have the lowest priority. If the **systemLogLevel** default of **error** is changed, only messages with the specified severity and all those with a higher priority are sent to the error log. The following table lists the message severity tags in order of priority:

Table 86. Message severity tags ordered by priority

Severity tag	Type of message (systemLogLevel attribute)	Meaning
A	**alert**	Indicates a problem where action must be taken immediately. Notify the appropriate person to correct the problem.
X	**critical**	Indicates a critical condition that should be corrected immediately. The system discovered an internal inconsistency of some kind. Command execution might be halted or the system might attempt to continue despite the inconsistency. Report these errors to the IBM Support Center.
E	**error**	Indicates an error condition. Command execution might or might not continue, but this error was likely caused by a persistent condition and will remain until corrected by some other program or administrative action. For example, a command operating on a single file or other GPFS object might terminate upon encountering any condition of severity **E**. As another example, a command operating on a list of files, finding that one of the files has permission bits set that disallow the operation, might continue to operate on all other files within the specified list of files.
W	**warning**	Indicates a problem, but command execution continues. The problem can be a transient inconsistency. It can be that the command has skipped some operations on some objects, or is reporting an irregularity that could be of interest. For example, if a multipass command operating on many files discovers during its second pass that a file that was present during the first pass is no longer present, the file might have been removed by another command or program.
N	**notice**	Indicates a normal but significant condition. These events are unusual but not error conditions, and might be summarized in an email to developers or administrators for spotting potential problems. No immediate action is required.
C	**configuration**	Indicates a configuration change; such as, creating a file system or removing a node from the cluster.
I	**informational**	Indicates normal operation. This message by itself indicates that nothing is wrong; no action is required.
D	**detail**	Indicates verbose operational messages; no is action required.
B	**debug**	Indicates debug-level messages that are useful to application developers for debugging purposes. This information is not useful during operations.

6027-000 **Attention: A disk being removed reduces the number of failure groups to** *nFailureGroups***, which is below the number required for replication:** *nReplicas***.**

Explanation: Replication cannot protect data against disk failures when there are insufficient failure groups.

User response: Add more disks in new failure groups to the file system or accept the risk of data loss.

6027-300 [N] mmfsd ready

Explanation: The **mmfsd** server is up and running.

User response: None. Informational message only.

6027-301 **File** *fileName* **could not be run with err** *errno***.**

Explanation: The named shell script could not be executed. This message is followed by the error string that is returned by the exec.

User response: Check file existence and access permissions.

6027-302 [E] Could not execute *script*

Explanation: The **verifyGpfsReady=yes** configuration attribute is set, but the **/var/mmfs/etc/gpfsready** script could not be executed.

User response: Make sure **/var/mmfs/etc/gpfsready** exists and is executable, or disable the **verifyGpfsReady** option via **mmchconfig verifyGpfsReady=no**.

6027-303 [N] *script* **killed by signal** *signal*

Explanation: The **verifyGpfsReady=yes** configuration attribute is set and **/var/mmfs/etc/gpfsready** script did not complete successfully.

User response: Make sure **/var/mmfs/etc/gpfsready** completes and returns a zero exit status, or disable the **verifyGpfsReady** option via **mmchconfig verifyGpfsReady=no**.

6027-304 [W] *script* **ended abnormally**

Explanation: The **verifyGpfsReady=yes** configuration attribute is set and **/var/mmfs/etc/gpfsready** script did not complete successfully.

User response: Make sure **/var/mmfs/etc/gpfsready** completes and returns a zero exit status, or disable the **verifyGpfsReady** option via **mmchconfig verifyGpfsReady=no**.

6027-305 [N] *script* **failed with exit code** *code*

Explanation: The **verifyGpfsReady=yes** configuration attribute is set and **/var/mmfs/etc/gpfsready** script did not complete successfully

User response: Make sure **/var/mmfs/etc/gpfsready** completes and returns a zero exit status, or disable the **verifyGpfsReady** option via **mmchconfig verifyGpfsReady=no**.

6027-306 [E] Could not initialize inter-node communication

Explanation: The GPFS daemon was unable to initialize the communications required to proceed.

User response: User action depends on the return code shown in the accompanying message (**/usr/include/errno.h**). The communications failure that caused the failure must be corrected. One possibility is an **rc** value of 67, indicating that the required port is unavailable. This may mean that a previous version of the **mmfs** daemon is still running. Killing that daemon may resolve the problem.

6027-307 [E] All tries for command thread allocation failed for `msgCommand` minor *commandMinorNumber*

Explanation: The GPFS daemon exhausted all tries and was unable to allocate a thread to process an incoming command message. This might impact the cluster.

User response: Evaluate thread usage using **`mmfsadm`** dump threads. Consider tuning the **`workerThreads`** parameter using the **`mmchconfig`** and then retry the command.

6027-310 [I] *command* **initializing. {Version** *versionName***: Built** *date time***}**

Explanation: The **mmfsd** server has started execution.

User response: None. Informational message only.

6027-311 [N] *programName* **is shutting down.**

Explanation: The stated program is about to terminate.

User response: None. Informational message only.

6027-312 [E] Unknown trace class '*traceClass***'.**

Explanation: The trace class is not recognized.

User response: Specify a valid trace class.

6027-313 [X] Cannot open configuration file *fileName***.**

Explanation: The configuration file could not be opened.

User response: The configuration file is **/var/mmfs/gen/mmfs.cfg**. Verify that this file and **/var/mmfs/gen/mmsdrfs** exist in your system.

6027-314 [E] *command* **requires** *SuperuserName* **authority to execute.**

Explanation: The mmfsd server was started by a user without superuser authority.

User response: Log on as a superuser and reissue the command.

6027-315 [E] Bad config file entry in *fileName***, line** *number***.**

Explanation: The configuration file has an incorrect entry.

User response: Fix the syntax error in the configuration file. Verify that you are not using a configuration file that was created on a release of GPFS subsequent to the one that you are currently running.

6027-316 [E] Unknown config parameter "*parameter*" in *fileName*, line *number*.

Explanation: There is an unknown parameter in the configuration file.

User response: Fix the syntax error in the configuration file. Verify that you are not using a configuration file that was created on a release of GPFS subsequent to the one you are currently running.

6027-317 [A] Old server with PID *pid* still running.

Explanation: An old copy of **mmfsd** is still running.

User response: This message would occur only if the user bypasses the SRC. The normal message in this case would be an SRC message stating that multiple instances are not allowed. If it occurs, stop the previous instance and use the SRC commands to restart the daemon.

6027-318 [E] Watchdog: Some process appears stuck; stopped the daemon process.

Explanation: A high priority process got into a loop.

User response: Stop the old instance of the **mmfs** server, then restart it.

6027-319 Could not create shared segment

Explanation: The shared segment could not be created.

User response: This is an error from the AIX operating system. Check the accompanying error indications from AIX.

6027-320 Could not map shared segment

Explanation: The shared segment could not be attached.

User response: This is an error from the AIX operating system. Check the accompanying error indications from AIX.

6027-321 Shared segment mapped at wrong address (is *value*, should be *value*).

Explanation: The shared segment did not get mapped to the expected address.

User response: Contact the IBM Support Center.

6027-322 Could not map shared segment in kernel extension

Explanation: The shared segment could not be mapped in the kernel.

User response: If an **EINVAL** error message is displayed, the kernel extension could not use the shared segment because it did not have the correct GPFS version number. Unload the kernel extension and restart the GPFS daemon.

6027-323 [A] Error unmapping shared segment.

Explanation: The shared segment could not be detached.

User response: Check reason given by error message.

6027-324 Could not create message queue for main process

Explanation: The message queue for the main process could not be created. This is probably an operating system error.

User response: Contact the IBM Support Center.

6027-328 [W] Value '*value*' for '*parameter*' is out of range in *fileName*. Valid values are *value* through *value*. *value* used.

Explanation: An error was found in the **/var/mmfs/gen/mmfs.cfg** file.

User response: Check the **/var/mmfs/gen/mmfs.cfg** file.

6027-329 Cannot pin the main shared segment: *name*

Explanation: Trying to pin the shared segment during initialization.

User response: Check the **mmfs.cfg** file. The **pagepool** size may be too large. It cannot be more than 80% of real memory. If a previous **mmfsd** crashed, check for processes that begin with the name **mmfs** that may be holding on to an old pinned shared segment. Issue **mmchconfig** command to change the **pagepool** size.

6027-334 [E] Error initializing internal communications.

Explanation: The mailbox system used by the daemon for communication with the kernel cannot be initialized.

User response: Increase the size of available memory using the **mmchconfig** command.

6027-335 [E] Configuration error: check *fileName*.

Explanation: A configuration error is found.

User response: Check the **mmfs.cfg** file and other error messages.

6027-336 [E] **Value '*value*' for configuration parameter '*parameter*' is not valid. Check *fileName*.**

Explanation: A configuration error was found.

User response: Check the **mmfs.cfg** file.

6027-337 [N] **Waiting for resources to be reclaimed before exiting.**

Explanation: The **mmfsd** daemon is attempting to terminate, but cannot because data structures in the daemon shared segment may still be referenced by kernel code. This message may be accompanied by other messages that show which disks still have I/O in progress.

User response: None. Informational message only.

6027-338 [N] **Waiting for *number* user(s) of shared segment to release it.**

Explanation: The **mmfsd** daemon is attempting to terminate, but cannot because some process is holding the shared segment while in a system call. The message will repeat every 30 seconds until the count drops to zero.

User response: Find the process that is not responding, and find a way to get it out of its system call.

6027-339 [E] **Nonnumeric trace value '*value*' after class '*class*'.**

Explanation: The specified trace value is not recognized.

User response: Specify a valid trace integer value.

6027-340 **Child process *file* failed to start due to error *rc*: *errStr*.**

Explanation: A failure occurred when GPFS attempted to start a program.

User response: If the program was a user exit script, verify the script file exists and has appropriate permissions assigned. If the program was not a user exit script, then this is an internal GPFS error or the GPFS installation was altered.

6027-341 [D] **Node *nodeName* is incompatible because its maximum compatible version (*number*) is less than the version of this node (*number*). [*value*/*value*]**

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other daemon is not compatible. Its maximum compatible version is less than the version of the daemon running on this node. The numbers in square brackets are for use by the IBM Support Center.

User response: Verify your GPFS daemon version.

6027-342 [E] **Node *nodeName* is incompatible because its minimum compatible version is greater than the version of this node (*number*). [*value*/*value*]**

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other daemon is not compatible. Its minimum compatible version is greater than the version of the daemon running on this node. The numbers in square brackets are for use by the IBM Support Center.

User response: Verify your GPFS daemon version.

6027-343 [E] **Node *nodeName* is incompatible because its version (*number*) is less than the minimum compatible version of this node (*number*). [*value*/*value*]**

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other daemon is not compatible. Its version is less than the minimum compatible version of the daemon running on this node. The numbers in square brackets are for use by the IBM Support Center.

User response: Verify your GPFS daemon version.

6027-344 [E] **Node *nodeName* is incompatible because its version is greater than the maximum compatible version of this node (*number*). [*value*/*value*]**

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other daemon is not compatible. Its version is greater than the maximum compatible version of the daemon running on this node. The numbers in square brackets are for use by the IBM Support Center.

User response: Verify your GPFS daemon version.

6027-345 **Network error on *ipAddress*, check connectivity.**

Explanation: A TCP error has caused GPFS to exit due to a bad return code from an error. Exiting allows recovery to proceed on another node and resources are not tied up on this node.

User response: Follow network problem determination procedures.

6027-346 [E] **Incompatible daemon version. My version = *number*, repl.my_version = *number***

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other GPFS daemon is not the same version and it sent

a reply indicating its version number is incompatible.

User response: Verify your GPFS daemon version.

6027-347 [E] Remote host *ipAddress* refused connection because IP address *ipAddress* was not in the node list file

Explanation: The GPFS daemon tried to make a connection with another GPFS daemon. However, the other GPFS daemon sent a reply indicating it did not recognize the IP address of the connector.

User response: Add the IP address of the local host to the node list file on the remote host.

6027-348 [E] Bad "subnets" configuration: invalid subnet "*ipAddress*".

Explanation: A subnet specified by the **subnets** configuration parameter could not be parsed.

User response: Run the **mmlsconfig** command and check the value of the **subnets** parameter. Each subnet must be specified as a dotted-decimal IP address. Run the **mmchconfig subnets** command to correct the value.

6027-349 [E] Bad "subnets" configuration: invalid cluster name pattern "*clusterNamePattern*".

Explanation: A cluster name pattern specified by the **subnets** configuration parameter could not be parsed.

User response: Run the **mmlsconfig** command and check the value of the **subnets** parameter. The optional cluster name pattern following subnet address must be a shell-style pattern allowing '*', '/' and '[...]' as wild cards. Run the **mmchconfig subnets** command to correct the value.

6027-350 [E] Bad "subnets" configuration: primary IP address *ipAddress* is on a private subnet. Use a public IP address instead.

Explanation: GPFS is configured to allow multiple IP addresses per node (**subnets** configuration parameter), but the primary IP address of the node (the one specified when the cluster was created or when the node was added to the cluster) was found to be on a private subnet. If multiple IP addresses are used, the primary address must be a public IP address.

User response: Remove the node from the cluster; then add it back using a public IP address.

6027-358 Communication with mmspsecserver through socket *name* failed, err *value*: *errorString*, msgType *messageType*.

Explanation: Communication failed between **spsecClient** (the daemon) and **spsecServer**.

User response: Verify both the communication socket and the **mmspsecserver** process.

6027-359 The mmspsecserver process is shutting down. Reason: *explanation*.

Explanation: The **mmspsecserver** process received a signal from the **mmfsd** daemon or encountered an error on execution.

User response: Verify the reason for shutdown.

6027-360 Disk *name* must be removed from the /etc/filesystems stanza before it can be deleted.

Explanation: A disk being deleted is found listed in the **disks=** list for a file system.

User response: Remove the disk from list.

6027-361 [E] Local access to *disk* failed with EIO, switching to access the disk remotely.

Explanation: Local access to the disk failed. To avoid unmounting of the file system, the disk will now be accessed remotely.

User response: Wait until work continuing on the local node completes. Then determine why local access to the disk failed, correct the problem and restart the daemon. This will cause GPFS to begin accessing the disk locally again.

6027-362 Attention: No disks were deleted, but some data was migrated. The file system may no longer be properly balanced.

Explanation: The **mmdeldisk** command did not complete migrating data off the disks being deleted. The disks were restored to normal **ready**, status, but the migration has left the file system unbalanced. This may be caused by having too many disks unavailable or insufficient space to migrate all of the data to other disks.

User response: Check disk availability and space requirements. Determine the reason that caused the command to end before successfully completing the migration and disk deletion. Reissue the **mmdeldisk** command.

6027-363 I/O error writing disk descriptor for disk *name*.

Explanation: An I/O error occurred when the **mmadddisk** command was writing a disk descriptor on a disk. This could have been caused by either a configuration error or an error in the path to the disk.

User response: Determine the reason the disk is inaccessible for writing and reissue the **mmadddisk** command.

6027-364 Error processing disks.

Explanation: An error occurred when the **mmadddisk** command was reading disks in the file system.

User response: Determine the reason why the disks are inaccessible for reading, then reissue the **mmadddisk** command.

6027-365 [I] Rediscovered local access to *disk*.

Explanation: Rediscovered local access to disk, which failed earlier with **EIO**. For good performance, the disk will now be accessed locally.

User response: Wait until work continuing on the local node completes. This will cause GPFS to begin accessing the disk locally again.

6027-369 I/O error writing file system descriptor for disk *name*.

Explanation: **mmadddisk** detected an I/O error while writing a file system descriptor on a disk.

User response: Determine the reason the disk is inaccessible for writing and reissue the **mmadddisk** command.

6027-370 mmdeldisk completed.

Explanation: The **mmdeldisk** command has completed.

User response: None. Informational message only.

6027-371 Cannot delete all disks in the file system

Explanation: An attempt was made to delete all the disks in a file system.

User response: Either reduce the number of disks to be deleted or use the **mmdelfs** command to delete the file system.

6027-372 Replacement disk must be in the same failure group as the disk being replaced.

Explanation: An improper failure group was specified for **mmrpldisk**.

User response: Specify a failure group in the disk descriptor for the replacement disk that is the same as the failure group of the disk being replaced.

6027-373 Disk *diskName* is being replaced, so status of disk *diskName* must be replacement.

Explanation: The **mmrpldisk** command failed when retrying a replace operation because the new disk does not have the correct status.

User response: Issue the **mmlsdisk** command to display disk status. Then either issue the **mmchdisk** command to change the status of the disk to **replacement** or specify a new disk that has a status of **replacement**.

6027-374 Disk *name* may not be replaced.

Explanation: A disk being replaced with **mmrpldisk** does not have a status of **ready** or **suspended**.

User response: Use the **mmlsdisk** command to display disk status. Issue the **mmchdisk** command to change the status of the disk to be replaced to either **ready** or **suspended**.

6027-375 Disk name *diskName* already in file system.

Explanation: The replacement disk name specified in the **mmrpldisk** command already exists in the file system.

User response: Specify a different disk as the replacement disk.

6027-376 Previous replace command must be completed before starting a new one.

Explanation: The **mmrpldisk** command failed because the status of other disks shows that a replace command did not complete.

User response: Issue the **mmlsdisk** command to display disk status. Retry the failed **mmrpldisk** command or issue the **mmchdisk** command to change the status of the disks that have a status of **replacing** or **replacement**.

6027-377 Cannot replace a disk that is in use.

Explanation: Attempting to replace a disk in place, but the disk specified in the **mmrpldisk** command is still available for use.

User response: Use the **mmchdisk** command to stop GPFS's use of the disk.

6027-378 [I] I/O still in progress near sector *number* on disk *diskName*.

Explanation: The **mmfsd** daemon is attempting to terminate, but cannot because data structures in the daemon shared segment may still be referenced by kernel code. In particular, the daemon has started an I/O that has not yet completed. It is unsafe for the daemon to terminate until the I/O completes, because of asynchronous activity in the device driver that will access data structures belonging to the daemon.

User response: Either wait for the I/O operation to time out, or issue a device-dependent command to terminate the I/O.

6027-379 **Could not invalidate disk(s).**

Explanation: Trying to delete a disk and it could not be written to in order to invalidate its contents.

User response: No action needed if removing that disk permanently. However, if the disk is ever to be used again, the **-v** flag must be specified with a value of **no** when using either the **mmcrfs** or **mmadddisk** command.

6027-380 **Disk name missing from disk descriptor list entry** *name***.**

Explanation: When parsing disk lists, no disks were named.

User response: Check the argument list of the command.

6027-382 **Value** *value* **for the 'sector size' option for disk** *disk* **is not a multiple of** *value***.**

Explanation: When parsing disk lists, the sector size given is not a multiple of the default sector size.

User response: Specify a correct sector size.

6027-383 **Disk name** *name* **appears more than once.**

Explanation: When parsing disk lists, a duplicate name is found.

User response: Remove the duplicate name.

6027-384 **Disk name** *name* **already in file system.**

Explanation: When parsing disk lists, a disk name already exists in the file system.

User response: Rename or remove the duplicate disk.

6027-385 **Value** *value* **for the 'sector size' option for disk** *name* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: When parsing disk lists, the sector size given is not valid.

User response: Specify a correct sector size.

6027-386 **Value** *value* **for the 'sector size' option for disk** *name* **is invalid.**

Explanation: When parsing disk lists, the sector size given is not valid.

User response: Specify a correct sector size.

6027-387 **Value** *value* **for the 'failure group' option for disk** *name* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: When parsing disk lists, the failure group given is not valid.

User response: Specify a correct failure group.

6027-388 **Value** *value* **for the 'failure group' option for disk** *name* **is invalid.**

Explanation: When parsing disk lists, the failure group given is not valid.

User response: Specify a correct failure group.

6027-389 **Value** *value* **for the 'has metadata' option for disk** *name* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: When parsing disk lists, the 'has metadata' value given is not valid.

User response: Specify a correct 'has metadata' value.

6027-390 **Value** *value* **for the 'has metadata' option for disk** *name* **is invalid.**

Explanation: When parsing disk lists, the 'has metadata' value given is not valid.

User response: Specify a correct 'has metadata' value.

6027-391 **Value** *value* **for the 'has data' option for disk** *name* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: When parsing disk lists, the 'has data' value given is not valid.

User response: Specify a correct 'has data' value.

6027-392 **Value** *value* **for the 'has data' option for disk** *name* **is invalid.**

Explanation: When parsing disk lists, the 'has data' value given is not valid.

User response: Specify a correct 'has data' value.

6027-393 **Either the 'has data' option or the 'has metadata' option must be '1' for disk** *diskName***.**

Explanation: When parsing disk lists the 'has data' or 'has metadata' value given is not valid.

User response: Specify a correct 'has data' or 'has metadata' value.

6027-394 Too many disks specified for file system. Maximum = *number***.**

Explanation: Too many disk names were passed in the disk descriptor list.

User response: Check the disk descriptor list or the file containing the list.

6027-399 Not enough items in disk descriptor list entry, need *fields***.**

Explanation: When parsing a disk descriptor, not enough fields were specified for one disk.

User response: Correct the disk descriptor to use the correct disk descriptor syntax.

6027-416 Incompatible file system descriptor version or not formatted.

Explanation: Possible reasons for the error are:

1. A file system descriptor version that is not valid was encountered.
2. No file system descriptor can be found.
3. Disks are not correctly defined on all active nodes.
4. Disks, logical volumes, network shared disks, or virtual shared disks were incorrectly re-configured after creating a file system.

User response: Verify:

1. The disks are correctly defined on all nodes.
2. The paths to the disks are correctly defined and operational.

6027-417 Bad file system descriptor.

Explanation: A file system descriptor that is not valid was encountered.

User response: Verify:

1. The disks are correctly defined on all nodes.
2. The paths to the disks are correctly defined and operational.

6027-418 Inconsistent file system quorum. readQuorum=*value* **writeQuorum=***value* **quorumSize=***value***.**

Explanation: A file system descriptor that is not valid was encountered.

User response: Start any disks that have been stopped by the **mmchdisk** command or by hardware failures. If the problem persists, run offline **mmfsck**.

6027-419 Failed to read a file system descriptor.

Explanation: Not enough valid replicas of the file system descriptor could be read from the file system.

User response: Start any disks that have been stopped by the **mmchdisk** command or by hardware failures. Verify that paths to all disks are correctly defined and operational.

6027-420 Inode size must be greater than zero.

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-421 Inode size must be a multiple of logical sector size.

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-422 Inode size must be at least as large as the logical sector size.

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-423 Minimum fragment size must be a multiple of logical sector size.

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-424 Minimum fragment size must be greater than zero.

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-425 File system block size of *blockSize* **is larger than maxblocksize parameter.**

Explanation: An attempt is being made to mount a file system whose block size is larger than the **maxblocksize** parameter as set by **mmchconfig**.

User response: Use the **mmchconfig maxblocksize**=*xxx* command to increase the maximum allowable block size.

6027-426 **Warning: mount detected unavailable disks. Use mmlsdisk** *fileSystem* **to see details.**

Explanation: The **mount** command detected that some disks needed for the file system are unavailable.

User response: Without file system replication enabled, the mount will fail. If it has replication, the mount may succeed depending on which disks are unavailable. Use **mmlsdisk** to see details of the disk status.

6027-427 **Indirect block size must be at least as large as the minimum fragment size.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-428 **Indirect block size must be a multiple of the minimum fragment size.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-429 **Indirect block size must be less than full data block size.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-430 **Default metadata replicas must be less than or equal to default maximum number of metadata replicas.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-431 **Default data replicas must be less than or equal to default maximum number of data replicas.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-432 **Default maximum metadata replicas must be less than or equal to** *value***.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-433 **Default maximum data replicas must be less than or equal to** *value***.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-434 **Indirect blocks must be at least as big as inodes.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-435 [N] **The file system descriptor quorum has been overridden.**

Explanation: The **mmfsctl exclude** command was previously issued to override the file system descriptor quorum after a disaster.

User response: None. Informational message only.

6027-438 **Duplicate disk name** *name***.**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-439 **Disk** *name* **sector size** *value* **does not match sector size** *value* **of other disk(s).**

Explanation: An internal consistency check has found a problem with file system parameters.

User response: Record the above information. Contact the IBM Support Center.

6027-441 **Unable to open disk '***name***' on node** *nodeName***.**

Explanation: A disk name that is not valid was specified in a GPFS disk command.

User response: Correct the parameters of the executing GPFS disk command.

6027-445 **Value for option '-m' cannot exceed the number of metadata failure groups.**

Explanation: The current number of replicas of metadata cannot be larger than the number of failure groups that are enabled to hold metadata.

User response: Use a smaller value for **-m** on the **mmchfs** command, or increase the number of failure groups by adding disks to the file system.

6027-446 **Value for option '-r' cannot exceed the number of data failure groups.**

Explanation: The current number of replicas of data cannot be larger than the number of failure groups that are enabled to hold data.

User response: Use a smaller value for **-r** on the **mmchfs** command, or increase the number of failure groups by adding disks to the file system.

6027-451 **No disks= list found in mount options.**

Explanation: No 'disks=' clause found in the mount options list when opening a file system.

User response: Check the operating system's file system database and local **mmsdrfs** file for this file system.

6027-452 **No disks found in disks= list.**

Explanation: No disks listed when opening a file system.

User response: Check the operating system's file system database and local **mmsdrfs** file for this file system.

6027-453 **No disk name found in a clause of the list.**

Explanation: No disk name found in a clause of the **disks=** list.

User response: Check the operating system's file system database and local **mmsdrfs** file for this file system.

6027-461 **Unable to find** *name* **device.**

Explanation: Self explanatory.

User response: There must be a **/dev/sg***name* special device defined. Check the error code. This could indicate a configuration error in the specification of disks, logical volumes, network shared disks, or virtual shared disks.

6027-462 *name* **must be a char or block special device.**

Explanation: Opening a file system.

User response: There must be a **/dev/sg***name* special device defined. This could indicate a configuration error in the specification of disks, logical volumes, network shared disks, or virtual shared disks.

6027-463 **SubblocksPerFullBlock was not 32.**

Explanation: The value of the SubblocksPerFullBlock variable was not 32. This situation should never exist, and indicates an internal error.

User response: Record the above information and contact the IBM Support Center.

6027-465 **The average file size must be at least as large as the minimum fragment size.**

Explanation: When parsing the command line of **tscrfs**, it was discovered that the average file size is smaller than the minimum fragment size.

User response: Correct the indicated command parameters.

6027-468 **Disk** *name* **listed in** *fileName* **or local mmsdrfs file, not found in device** *name***. Run: mmcommon recoverfs** *name***.**

Explanation: Tried to access a file system but the disks listed in the operating system's file system database or the local **mmsdrfs** file for the device do not exist in the file system.

User response: Check the configuration and availability of disks. Run the **mmcommon recoverfs** *device* command. If this does not resolve the problem, configuration data in the SDR may be incorrect. If no user modifications have been made to the SDR, contact the IBM Support Center. If user modifications have been made, correct these modifications.

6027-469 **File system** *name* **does not match** *descriptor***.**

Explanation: The file system name found in the descriptor on disk does not match the corresponding device name in **/etc/filesystems**.

User response: Check the operating system's file system database.

6027-470 **Disk** *name* **may still belong to file system** *filesystem***. Created on** *IPandTime***.**

Explanation: The disk being added by the **mmcrfs**, **mmadddisk**, or **mmrpldisk** command appears to still belong to some file system.

User response: Verify that the disks you are adding do not belong to an active file system, and use the **-v no** option to bypass this check. Use this option only if you are sure that no other file system has this disk configured because you may cause data corruption in both file systems if this is not the case.

6027-471 **Disk *diskName*: Incompatible file system descriptor version or not formatted.**

Explanation: Possible reasons for the error are:

1. A file system descriptor version that is not valid was encountered.
2. No file system descriptor can be found.
3. Disks are not correctly defined on all active nodes.
4. Disks, logical volumes, network shared disks, or virtual shared disks were incorrectly reconfigured after creating a file system.

User response: Verify:

1. The disks are correctly defined on all nodes.
2. The paths to the disks are correctly defined and operative.

6027-472 [E] **File system format version *versionString* is not supported.**

Explanation: The current file system format version is not supported.

User response: Verify:

1. The disks are correctly defined on all nodes.
2. The paths to the disks are correctly defined and operative.

6027-473 [X] **File System *fileSystem* unmounted by the system with return code *value* reason code *value***

Explanation: Console log entry caused by a forced unmount due to disk or communication failure.

User response: Correct the underlying problem and remount the file system.

6027-474 [X] **Recovery Log I/O failed, unmounting file system *fileSystem***

Explanation: I/O to the recovery log failed.

User response: Check the paths to all disks making up the file system. Run the **mmlsdisk** command to determine if GPFS has declared any disks unavailable. Repair any paths to disks that have failed. Remount the file system.

6027-475 **The option '--inode-limit' is not enabled. Use option '-V' to enable most recent features.**

Explanation: **mmchfs --inode-limit** is not enabled under the current file system format version.

User response: Run **mmchfs -V**, this will change the file system format to the latest format supported.

6027-476 **Restricted mount using only available file system descriptor.**

Explanation: Fewer than the necessary number of file system descriptors were successfully read. Using the best available descriptor to allow the restricted mount to continue.

User response: Informational message only.

6027-477 **The option -z is not enabled. Use the -V option to enable most recent features.**

Explanation: The file system format version does not support the **-z** option on the **mmchfs** command.

User response: Change the file system format version by issuing **mmchfs -V**.

6027-478 **The option -z could not be changed. *fileSystem* is still in use.**

Explanation: The file system is still mounted or another GPFS administration command (**mm...**) is running against the file system.

User response: Unmount the file system if it is mounted, and wait for any command that is running to complete before reissuing the **mmchfs -z** command.

6027-479 [N] **Mount of *fsName* was blocked by *fileName***

Explanation: The internal or external mount of the file system was blocked by the existence of the specified file.

User response: If the file system needs to be mounted, remove the specified file.

6027-480 **Cannot enable DMAPI in a file system with existing snapshots.**

Explanation: The user is not allowed to enable DMAPI for a file system with existing snapshots.

User response: Delete all existing snapshots in the file system and repeat the **mmchfs** command.

6027-481 [E] Remount failed for mountid *id*: *errnoDescription*

Explanation: **mmfsd** restarted and tried to remount any file systems that the VFS layer thinks are still mounted.

User response: Check the errors displayed and the **errno** description.

6027-482 [E] Remount failed for device *name*: *errnoDescription*

Explanation: **mmfsd** restarted and tried to remount any file systems that the VFS layer thinks are still mounted.

User response: Check the errors displayed and the **errno** description.

6027-483 [N] Remounted *name*

Explanation: **mmfsd** restarted and remounted the specified file system because it was in the kernel's list of previously mounted file systems.

User response: Informational message only.

6027-484 Remount failed for *device* **after daemon restart.**

Explanation: A remount failed after daemon restart. This ordinarily occurs because one or more disks are unavailable. Other possibilities include loss of connectivity to one or more disks.

User response: Issue the **mmlsdisk** command and check for **down** disks. Issue the **mmchdisk** command to start any **down** disks, then remount the file system. If there is another problem with the disks or the connections to the disks, take necessary corrective actions and remount the file system.

6027-485 Perform mmchdisk for any disk failures and re-mount.

Explanation: Occurs in conjunction with 6027-484.

User response: Follow the *User response* for 6027-484.

6027-486 No local device specified for *fileSystemName* **in** *clusterName*.

Explanation: While attempting to mount a remote file system from another cluster, GPFS was unable to determine the local device name for this file system.

User response: There must be a **/dev/sgname** special device defined. Check the error code. This is probably a configuration error in the specification of a remote file system. Run **mmremotefs show** to check that the remote file system is properly configured.

6027-487 Failed to write the file system descriptor to disk *diskName*.

Explanation: An error occurred when **mmfsctl include** was writing a copy of the file system descriptor to one of the disks specified on the command line. This could have been caused by a failure of the corresponding disk device, or an error in the path to the disk.

User response: Verify that the disks are correctly defined on all nodes. Verify that paths to all disks are correctly defined and operational.

6027-488 Error opening the exclusion disk file *fileName*.

Explanation: Unable to retrieve the list of excluded disks from an internal configuration file.

User response: Ensure that GPFS executable files have been properly installed on all nodes. Perform required configuration steps prior to starting GPFS.

6027-489 Attention: The desired replication factor exceeds the number of available *dataOrMetadata* **failure groups. This is allowed, but the files will not be replicated and will therefore be at risk.**

Explanation: You specified a number of replicas that exceeds the number of failure groups available.

User response: Reissue the command with a smaller replication factor, or increase the number of failure groups.

6027-490 [N] The descriptor replica on disk *diskName* **has been excluded.**

Explanation: The file system descriptor quorum has been overridden and, as a result, the specified disk was excluded from all operations on the file system descriptor quorum.

User response: None. Informational message only.

6027-491 Incompatible file system format. Only file systems formatted with GPFS 3.2 or later can be mounted on this platform.

Explanation: User running GPFS on Microsoft Windows tried to mount a file system that was formatted with a version of GPFS that did not have Windows support.

User response: Create a new file system using current GPFS code.

6027-492 **The file system is already at file system version** *number*

Explanation: The user tried to upgrade the file system format using **mmchfs -V --version=v**, but the specified version is smaller than the current version of the file system.

User response: Specify a different value for the **--version** option.

6027-493 **File system version** *number* **is not supported on** *nodeName* **nodes in the cluster.**

Explanation: The user tried to upgrade the file system format using **mmchfs -V**, but some nodes in the local cluster are still running an older GPFS release that does support the new format version.

User response: Install a newer version of GPFS on those nodes.

6027-494 **File system version** *number* **is not supported on the following** *nodeName* **remote nodes mounting the file system:**

Explanation: The user tried to upgrade the file system format using **mmchfs -V**, but the file system is still mounted on some nodes in remote clusters that do not support the new format version.

User response: Unmount the file system on the nodes that do not support the new format version.

6027-495 **You have requested that the file system be upgraded to version** *number***. This will enable new functionality but will prevent you from using the file system with earlier releases of GPFS. Do you want to continue?**

Explanation: Verification request in response to the **mmchfs -V full** command. This is a request to upgrade the file system and activate functions that are incompatible with a previous release of GPFS.

User response: Enter **yes** if you want the conversion to take place.

6027-496 **You have requested that the file system version for local access be upgraded to version** *number***. This will enable some new functionality but will prevent local nodes from using the file system with earlier releases of GPFS. Remote nodes are not affected by this change. Do you want to continue?**

Explanation: Verification request in response to the **mmchfs -V** command. This is a request to upgrade the file system and activate functions that are incompatible with a previous release of GPFS.

User response: Enter **yes** if you want the conversion to take place.

6027-497 **The file system has already been upgraded to** *number* **using -V full. It is not possible to revert back.**

Explanation: The user tried to upgrade the file system format using **mmchfs -V compat**, but the file system has already been fully upgraded.

User response: Informational message only.

6027-498 **Incompatible file system format. Only file systems formatted with GPFS 3.2.1.5 or later can be mounted on this platform.**

Explanation: A user running GPFS on Microsoft Windows tried to mount a file system that was formatted with a version of GPFS that did not have Windows support.

User response: Create a new file system using current GPFS code.

6027-499 [X] **An unexpected Device Mapper path** *dmDevice* **(***nsdId***) has been detected. The new path does not have a Persistent Reserve set up. File system** *fileSystem* **will be internally unmounted.**

Explanation: A new device mapper path is detected or a previously failed path is activated after the local device discovery has finished. This path lacks a Persistent Reserve, and can not be used. All device paths must be active at mount time.

User response: Check the paths to all disks making up the file system. Repair any paths to disks which have failed. Remount the file system.

6027-500 *name* **loaded and configured.**

Explanation: The kernel extension was loaded and configured.

User response: None. Informational message only.

6027-501 *name***:module** *moduleName* **unloaded.**

Explanation: The kernel extension was unloaded.

User response: None. Informational message only.

6027-502 **Incorrect parameter:** *name***.**

Explanation: **mmfsmnthelp** was called with an incorrect parameter.

User response: Contact the IBM Support Center.

6027-504 **Not enough memory to allocate internal data structure.**

Explanation: Self explanatory.

User response: Increase ulimit or paging space

6027-505 **Internal error, aborting.**

Explanation: Self explanatory.

User response: Contact the IBM Support Center.

6027-506 *program*: *loadFile* **is already loaded at** *address*.

Explanation: The program was already loaded at the address displayed.

User response: None. Informational message only.

6027-507 *program*: *loadFile* **is not loaded.**

Explanation: The program could not be loaded.

User response: None. Informational message only.

6027-510 **Cannot mount** *fileSystem* **on** *mountPoint*: *errorString*

Explanation: There was an error mounting the GPFS file system.

User response: Determine action indicated by the error messages and error log entries. Errors in the disk path often cause this problem.

6027-511 **Cannot unmount** *fileSystem*: *errorDescription*

Explanation: There was an error unmounting the GPFS file system.

User response: Take the action indicated by **errno** description.

6027-512 *name* **not listed in /etc/vfs**

Explanation: Error occurred while installing the GPFS kernel extension, or when trying to mount a file system.

User response: Check for the **mmfs** entry in **/etc/vfs**

6027-514 **Cannot mount** *fileSystem* **on** *mountPoint*: **Already mounted.**

Explanation: An attempt has been made to mount a file system that is already mounted.

User response: None. Informational message only.

6027-515 **Cannot mount** *fileSystem* **on** *mountPoint*

Explanation: There was an error mounting the named GPFS file system. Errors in the disk path usually cause this problem.

User response: Take the action indicated by other error messages and error log entries.

6027-516 **Cannot mount** *fileSystem*

Explanation: There was an error mounting the named GPFS file system. Errors in the disk path usually cause this problem.

User response: Take the action indicated by other error messages and error log entries.

6027-517 **Cannot mount** *fileSystem*: *errorString*

Explanation: There was an error mounting the named GPFS file system. Errors in the disk path usually cause this problem.

User response: Take the action indicated by other error messages and error log entries.

6027-518 **Cannot mount** *fileSystem*: **Already mounted.**

Explanation: An attempt has been made to mount a file system that is already mounted.

User response: None. Informational message only.

6027-519 **Cannot mount** *fileSystem* **on** *mountPoint*: **File system table full.**

Explanation: An attempt has been made to mount a file system when the file system table is full.

User response: None. Informational message only.

6027-520 **Cannot mount** *fileSystem*: **File system table full.**

Explanation: An attempt has been made to mount a file system when the file system table is full.

User response: None. Informational message only.

6027-530 **Mount of** *name* **failed: cannot mount restorable file system for read/write.**

Explanation: A file system marked as **enabled** for restore cannot be mounted **read/write**.

User response: None. Informational message only.

6027-531 **The following disks of *name* will be formatted on node *nodeName*: *list*.**

Explanation: Output showing which disks will be formatted by the **mmcrfs** command.

User response: None. Informational message only.

6027-532 [E] **The quota record *recordNumber* in file *fileName* is not valid.**

Explanation: A quota entry contained a checksum that is not valid.

User response: Remount the file system with quotas disabled. Restore the quota file from back up, and run **mmcheckquota**.

6027-533 [W] **Inode space *inodeSpace* in file system *fileSystem* is approaching the limit for the maximum number of inodes.**

Explanation: The number of files created is approaching the file system limit.

User response: Use the **mmchfileset** command to increase the maximum number of files to avoid reaching the inode limit and possible performance degradation.

6027-534 **Cannot create a snapshot in a DMAPI-enabled file system, rc=*returnCode*.**

Explanation: You cannot create a snapshot in a DMAPI-enabled file system.

User response: Use the **mmchfs** command to disable DMAPI, and reissue the command.

6027-535 **Disks up to size *size* can be added to storage pool *pool*.**

Explanation: Based on the parameters given to **mmcrfs** and the size and number of disks being formatted, GPFS has formatted its allocation maps to allow disks up the given size to be added to this storage pool by the **mmadddisk** command.

User response: None. Informational message only. If the reported maximum disk size is smaller than necessary, delete the file system with **mmdelfs** and rerun **mmcrfs** with either larger disks or a larger value for the **-n** parameter.

6027-536 **Insufficient system memory to run GPFS daemon. Reduce page pool memory size with the mmchconfig command or add additional RAM to system.**

Explanation: Insufficient memory for GPFS internal data structures with current system and GPFS configuration.

User response: Reduce page pool usage with the **mmchconfig** command, or add additional RAM to system.

6027-537 **Disks up to size *size* can be added to this file system.**

Explanation: Based on the parameters given to the **mmcrfs** command and the size and number of disks being formatted, GPFS has formatted its allocation maps to allow disks up the given size to be added to this file system by the **mmadddisk** command.

User response: None, informational message only. If the reported maximum disk size is smaller than necessary, delete the file system with **mmdelfs** and reissue the **mmcrfs** command with larger disks or a larger value for the **-n** parameter.

6027-538 **Error accessing disks.**

Explanation: The **mmcrfs** command encountered an error accessing one or more of the disks.

User response: Verify that the disk descriptors are coded correctly and that all named disks exist and are online.

6027-539 **Unable to clear descriptor areas for *fileSystem*.**

Explanation: The **mmdelfs** command encountered an error while invalidating the file system control structures on one or more disks in the file system being deleted.

User response: If the problem persists, specify the **-p** option on the **mmdelfs** command.

6027-540 **Formatting file system.**

Explanation: The **mmcrfs** command began to write file system data structures onto the new disks.

User response: None. Informational message only.

6027-541 **Error formatting file system.**

Explanation: **mmcrfs** command encountered an error while formatting a new file system. This is often an I/O error.

User response: Check the subsystems in the path to the disk. Follow the instructions from other messages that appear with this one.

6027-542 [N] Fileset in file system *fileSystem*:*filesetName* (id *filesetId*) has been incompletely deleted.

Explanation: A fileset delete operation was interrupted, leaving this fileset in an incomplete state.

User response: Reissue the fileset delete command.

6027-543 Error writing file system descriptor for *fileSystem*.

Explanation: The **mmcrfs** command could not successfully write the file system descriptor in a particular file system. Check the subsystems in the path to the disk. This is often an I/O error.

User response: Check system error log, rerun **mmcrfs**.

6027-544 Could not invalidate *disk* of *fileSystem*.

Explanation: A disk could not be written to invalidate its contents. Check the subsystems in the path to the disk. This is often an I/O error.

User response: Ensure the indicated logical volume is writable.

6027-545 Error processing fileset metadata file.

Explanation: There is no I/O path to critical metadata or metadata has been corrupted.

User response: Verify that the I/O paths to all disks are valid and that all disks are either in the 'recovering' or 'up' availability states. If all disks are available and the problem persists, issue the **mmfsck** command to repair damaged metadata

6027-546 Error processing allocation map for storage pool *poolName*.

Explanation: There is no I/O path to critical metadata, or metadata has been corrupted.

User response: Verify that the I/O paths to all disks are valid, and that all disks are either in the 'recovering' or 'up' availability. Issue the **mmlsdisk** command.

6027-547 Fileset *filesetName* was unlinked.

Explanation: Fileset was already unlinked.

User response: None. Informational message only.

6027-548 Fileset *filesetName* unlinked from *filesetName*.

Explanation: A fileset being deleted contains junctions to other filesets. The cited fileset were unlinked.

User response: None. Informational message only.

6027-549 [E] Failed to open *name*.

Explanation: The **mount** command was unable to access a file system. Check the subsystems in the path to the disk. This is often an I/O error.

User response: Follow the suggested actions for the other messages that occur with this one.

6027-550 [X] Allocation manager for *fileSystem* failed to revoke ownership from node *nodeName*.

Explanation: An irrecoverable error occurred trying to revoke ownership of an allocation region. The allocation manager has panicked the file system to prevent corruption of on-disk data.

User response: Remount the file system.

6027-551 *fileSystem* is still in use.

Explanation: The **mmdelfs** or **mmcrfs** command found that the named file system is still mounted or that another GPFS command is running against the file system.

User response: Unmount the file system if it is mounted, or wait for GPFS commands in progress to terminate before retrying the command.

6027-552 Scan completed successfully.

Explanation: The scan function has completed without error.

User response: None. Informational message only.

6027-553 Scan failed on *number* user or system files.

Explanation: Data may be lost as a result of pointers that are not valid or unavailable disks.

User response: Some files may have to be restored from backup copies. Issue the **mmlsdisk** command to check the availability of all the disks that make up the file system.

6027-554 Scan failed on *number* out of *number* user or system files.

Explanation: Data may be lost as a result of pointers that are not valid or unavailable disks.

User response: Some files may have to be restored from backup copies. Issue the **mmlsdisk** command to check the availability of all the disks that make up the file system.

6027-555 The desired replication factor exceeds the number of available failure groups.

Explanation: You have specified a number of replicas that exceeds the number of failure groups available.

User response: Reissue the command with a smaller replication factor or increase the number of failure groups.

6027-556 Not enough space for the desired number of replicas.

Explanation: In attempting to restore the correct replication, GPFS ran out of space in the file system. The operation can continue but some data is not fully replicated.

User response: Make additional space available and reissue the command.

6027-557 Not enough space or available disks to properly balance the file.

Explanation: In attempting to stripe data within the file system, data was placed on a disk other than the desired one. This is normally not a problem.

User response: Run **mmrestripefs** to rebalance all files.

6027-558 Some data are unavailable.

Explanation: An I/O error has occurred or some disks are in the stopped state.

User response: Check the availability of all disks by issuing the **mmlsdisk** command and check the path to all disks. Reissue the command.

6027-559 Some data could not be read or written.

Explanation: An I/O error has occurred or some disks are in the stopped state.

User response: Check the availability of all disks and the path to all disks, and reissue the command.

6027-560 File system is already suspended.

Explanation: The **tsfsctl** command was asked to suspend a suspended file system.

User response: None. Informational message only.

6027-561 Error migrating log.

Explanation: There are insufficient available disks to continue operation.

User response: Restore the unavailable disks and reissue the command.

6027-562 Error processing inodes.

Explanation: There is no I/O path to critical metadata or metadata has been corrupted.

User response: Verify that the I/O paths to all disks are valid and that all disks are either in the `recovering` or `up` availability. Issue the **mmlsdisk** command.

6027-563 File system is already running.

Explanation: The **tsfsctl** command was asked to resume a file system that is already running.

User response: None. Informational message only.

6027-564 Error processing inode allocation map.

Explanation: There is no I/O path to critical metadata or metadata has been corrupted.

User response: Verify that the I/O paths to all disks are valid and that all disks are either in the `recovering` or `up` availability. Issue the **mmlsdisk** command.

6027-565 Scanning user file metadata ...

Explanation: Progress information.

User response: None. Informational message only.

6027-566 Error processing user file metadata.

Explanation: Error encountered while processing user file metadata.

User response: None. Informational message only.

6027-567 Waiting for pending file system scan to finish ...

Explanation: Progress information.

User response: None. Informational message only.

6027-568 Waiting for *number* pending file system scans to finish ...

Explanation: Progress information.

User response: None. Informational message only.

6027-569 Incompatible parameters. Unable to allocate space for file system metadata. Change one or more of the following as suggested and try again:

Explanation: Incompatible file system parameters were detected.

User response: Refer to the details given and correct the file system parameters.

6027-570 **Incompatible parameters. Unable to create file system. Change one or more of the following as suggested and try again:**

Explanation: Incompatible file system parameters were detected.

User response: Refer to the details given and correct the file system parameters.

6027-571 **Logical sector size** *value* **must be the same as disk sector size.**

Explanation: This message is produced by the **mmcrfs** command if the sector size given by the **-l** option is not the same as the sector size given for disks in the **-d** option.

User response: Correct the options and reissue the command.

6027-572 **Completed creation of file system** *fileSystem***.**

Explanation: The **mmcrfs** command has successfully completed.

User response: None. Informational message only.

6027-573 **All data on the following disks of** *fileSystem* **will be destroyed:**

Explanation: Produced by the **mmdelfs** command to list the disks in the file system that is about to be destroyed. Data stored on the disks will be lost.

User response: None. Informational message only.

6027-574 **Completed deletion of file system** *fileSystem***.**

Explanation: The **mmdelfs** command has successfully completed.

User response: None. Informational message only.

6027-575 **Unable to complete low level format for** *fileSystem***. Failed with error** *errorCode*

Explanation: The **mmcrfs** command was unable to create the low level file structures for the file system.

User response: Check other error messages and the error log. This is usually an error accessing disks.

6027-576 **Storage pools have not been enabled for file system** *fileSystem***.**

Explanation: User invoked a command with a storage pool option (**-p** or **-P**) before storage pools were enabled.

User response: Enable storage pools with the **mmchfs** **-V** command, or correct the command invocation and reissue the command.

6027-577 **Attention:** *number* **user or system files are not properly replicated.**

Explanation: GPFS has detected files that are not replicated correctly due to a previous failure.

User response: Issue the **mmrestripefs** command at the first opportunity.

6027-578 **Attention:** *number* **out of** *number* **user or system files are not properly replicated:**

Explanation: GPFS has detected files that are not replicated correctly

6027-579 **Some unreplicated file system metadata has been lost. File system usable only in restricted mode.**

Explanation: A disk was deleted that contained vital file system metadata that was not replicated.

User response: Mount the file system in restricted mode (**-o rs**) and copy any user data that may be left on the file system. Then delete the file system.

6027-580 **Unable to access vital system metadata. Too many disks are unavailable.**

Explanation: Metadata is unavailable because the disks on which the data reside are stopped, or an attempt was made to delete them.

User response: Either start the stopped disks, try to delete the disks again, or recreate the file system.

6027-581 **Unable to access vital system metadata, file system corrupted.**

Explanation: When trying to access the files system, the metadata was unavailable due to a disk being deleted.

User response: Determine why a disk is unavailable.

6027-582 **Some data has been lost.**

Explanation: An I/O error has occurred or some disks are in the stopped state.

User response: Check the availability of all disks by issuing the **mmlsdisk** command and check the path to all disks. Reissue the command.

6027-584 **Incompatible parameters. Unable to allocate space for root directory. Change one or more of the following as suggested and try again:**

Explanation: Inconsistent parameters have been

passed to the **mmcrfs** command, which would result in the creation of an inconsistent file system. Suggested parameter changes are given.

User response: Reissue the **mmcrfs** command with the suggested parameter changes.

6027-585 **Incompatible parameters. Unable to allocate space for ACL data. Change one or more of the following as suggested and try again:**

Explanation: Inconsistent parameters have been passed to the **mmcrfs** command, which would result in the creation of an inconsistent file system. The parameters entered require more space than is available. Suggested parameter changes are given.

User response: Reissue the **mmcrfs** command with the suggested parameter changes.

6027-586 **Quota server initialization failed.**

Explanation: Quota server initialization has failed. This message may appear as part of the detail data in the quota error log.

User response: Check status and availability of the disks. If quota files have been corrupted, restore them from the last available backup. Finally, reissue the command.

6027-587 **Unable to initialize quota client because there is no quota server. Please check error log on the file system manager node. The mmcheckquota command must be run with the file system unmounted before retrying the command.**

Explanation: `startQuotaClient` failed.

User response: If the quota file could not be read (check error log on file system manager. Issue the **mmlsmgr** command to determine which node is the file system manager), then the **mmcheckquota** command must be run with the file system unmounted.

6027-588 **No more than** *number* **nodes can mount a file system.**

Explanation: The limit of the number of nodes that can mount a file system was exceeded.

User response: Observe the stated limit for how many nodes can mount a file system.

6027-589 **Scanning file system metadata, phase** *number* **...**

Explanation: Progress information.

User response: None. Informational message only.

6027-590 [W] **GPFS is experiencing a shortage of pagepool. This message will not be repeated for at least one hour.**

Explanation: Pool starvation occurs, buffers have to be continually stolen at high aggressiveness levels.

User response: Issue the **mmchconfig** command to increase the size of **pagepool**.

6027-591 **Unable to allocate sufficient inodes for file system metadata. Increase the value for** *option* **and try again.**

Explanation: Too few inodes have been specified on the **-N** option of the **mmcrfs** command.

User response: Increase the size of the **-N** option and reissue the **mmcrfs** command.

6027-592 **Mount of** *fileSystem* **is waiting for the mount disposition to be set by some data management application.**

Explanation: Data management utilizing DMAPI is enabled for the file system, but no data management application has set a disposition for the mount event.

User response: Start the data management application and verify that the application sets the mount disposition.

6027-593 [E] **The root quota entry is not found in its assigned record**

Explanation: On mount, the root entry is not found in the first record of the quota file.

User response: Issue the **mmcheckquota** command to verify that the use of root has not been lost.

6027-594 **Disk** *diskName* **cannot be added to storage pool** *poolName***. Allocation map cannot accommodate disks larger than** *size* **MB.**

Explanation: The specified disk is too large compared to the disks that were initially used to create the storage pool.

User response: Specify a smaller disk or add the disk to a new storage pool.

6027-595 [E] **While creating quota files, file *fileName*, with no valid quota information was found in the root directory. Remove files with reserved quota file names (for example, user.quota) without valid quota information from the root directory by: - mounting the file system without quotas, - removing the files, and - remounting the file system with quotas to recreate new quota files. To use quota file names other than the reserved names, use the mmcheckquota command.**

Explanation: While mounting a file system, the state of the file system descriptor indicates that quota files do not exist. However, files that do not contain quota information but have one of the reserved names: **user.quota**, **group.quota**, or **fileset.quota** exist in the root directory.

User response: To mount the file system so that new quota files will be created, perform these steps:

1. Mount the file system without quotas.
2. Verify that there are no files in the root directory with the reserved names: **user.quota**, **group.quota**, or **fileset.quota**.
3. Remount the file system with quotas. To mount the file system with other files used as quota files, issue the **mmcheckquota** command.

6027-596 [I] **While creating quota files, file *fileName* containing quota information was found in the root directory. This file will be used as *quotaType* quota file.**

Explanation: While mounting a file system, the state of the file system descriptor indicates that quota files do not exist. However, files that have one of the reserved names **user.quota**, **group.quota**, or **fileset.quota** and contain quota information, exist in the root directory. The file with the reserved name will be used as the quota file.

User response: None. Informational message.

6027-597 [E] **The quota command was requested to process quotas for a type (user, group, or fileset), which is not enabled.**

Explanation: A quota command was requested to process quotas for a **user**, **group**, or **fileset** quota type, which is not enabled.

User response: Verify that the **user**, **group**, or **fileset** quota type is enabled and reissue the command.

6027-598 [E] **The supplied file does not contain quota information.**

Explanation: A file supplied as a quota file does not contain quota information.

User response: Change the file so it contains valid quota information and reissue the command.

To mount the file system so that new quota files are created:

1. Mount the file system without quotas.
2. Verify there are no files in the root directory with the reserved **user.quota** or **group.quota** name.
3. Remount the file system with quotas.

6027-599 [E] **File supplied to the command does not exist in the root directory.**

Explanation: The user-supplied name of a new quota file has not been found.

User response: Ensure that a file with the supplied name exists. Then reissue the command.

6027-600 **On node *nodeName* an earlier error may have caused some file system data to be inaccessible at this time. Check error log for additional information. After correcting the problem, the file system can be mounted again to restore normal data access.**

Explanation: An earlier error may have caused some file system data to be inaccessible at this time.

User response: Check the error log for additional information. After correcting the problem, the file system can be mounted again.

6027-601 **Error changing pool size.**

Explanation: The **mmchconfig** command failed to change the pool size to the requested value.

User response: Follow the suggested actions in the other messages that occur with this one.

6027-602 **ERROR: file system not mounted. Mount file system *fileSystem* and retry command.**

Explanation: A GPFS command that requires the file system be mounted was issued.

User response: Mount the file system and reissue the command.

6027-603 **Current pool size:** *value***K = ***value***M, max block size:** *value***K = ***value***M.**

Explanation: Displays the current pool size.

User response: None. Informational message only.

6027-604 [E] **Parameter incompatibility. File system block size is larger than maxblocksize parameter.**

Explanation: An attempt is being made to mount a file system whose block size is larger than the **maxblocksize** parameter as set by **mmchconfig**.

User response: Use the **mmchconfig maxblocksize=xxx** command to increase the maximum allowable block size.

6027-605 [N] **File system has been renamed.**

Explanation: Self-explanatory.

User response: None. Informational message only.

6027-606 [E] **The node number** *nodeNumber* **is not defined in the node list**

Explanation: A node matching *nodeNumber* was not found in the GPFS configuration file.

User response: Perform required configuration steps prior to starting GPFS on the node.

6027-607 **mmcommon getEFOptions** *fileSystem* **failed. Return code** *value***.**

Explanation: The **mmcommon getEFOptions** command failed while looking up the names of the disks in a file system. This error usually occurs during **mount** processing.

User response: Check the preceding messages. A frequent cause for such errors is lack of space in **/var**.

6027-608 [E] **File system manager takeover failed.**

Explanation: An attempt to takeover as file system manager failed. The file system is unmounted to allow another node to try.

User response: Check the return code. This is usually due to network or disk connectivity problems. Issue the **mmlsdisk** command to determine if the paths to the disk are unavailable, and issue the **mmchdisk** if necessary.

6027-609 **File system** *fileSystem* **unmounted because it does not have a manager.**

Explanation: The file system had to be unmounted because a file system manager could not be assigned. An accompanying message tells which node was the last manager.

User response: Examine error log on the last file system manager. Issue the **mmlsdisk** command to determine if a number of disks are down. Examine the other error logs for an indication of network, disk, or virtual shared disk problems. Repair the base problem and issue the **mmchdisk** command if required.

6027-610 **Cannot mount file system** *fileSystem* **because it does not have a manager.**

Explanation: The file system had to be unmounted because a file system manager could not be assigned. An accompanying message tells which node was the last manager.

User response: Examine error log on the last file system manager node. Issue the **mmlsdisk** command to determine if a number of disks are down. Examine the other error logs for an indication of disk or network shared disk problems. Repair the base problem and issue the **mmchdisk** command if required.

6027-611 [I] **Recovery:** *fileSystem***, delay** *number* **sec. for safe recovery.**

Explanation: Informational. When disk leasing is in use, wait for the existing lease to expire before performing log and token manager recovery.

User response: None.

6027-612 **Unable to run** *command* **while the file system is suspended.**

Explanation: A command that can alter data in a file system was issued while the file system was suspended.

User response: Resume the file system and reissue the command.

6027-613 [N] **Expel** *node* **request from** *node***. Expelling:** *node*

Explanation: One node is asking to have another node expelled from the cluster, usually because they have communications problems between them. The cluster manager node will decide which one will be expelled.

User response: Check that the communications paths are available between the two nodes.

6027-614 **Value** *value* **for option** *name* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: The value for an option in the command line arguments is out of range.

User response: Correct the command line and reissue the command.

6027-615 mmcommon getContactNodes *clusterName* failed. Return code *value*.

Explanation: **mmcommon getContactNodes** failed while looking up contact nodes for a remote cluster, usually while attempting to mount a file system from a remote cluster.

User response: Check the preceding messages, and consult the earlier chapters of this document. A frequent cause for such errors is lack of space in **/var**.

6027-616 [X] Duplicate address *ipAddress* in node list

Explanation: The IP address appears more than once in the node list file.

User response: Check the node list shown by the **mmlscluster** command.

6027-617 [I] Recovered *number* nodes for cluster *clusterName*.

Explanation: The asynchronous part (phase 2) of node failure recovery has completed.

User response: None. Informational message only.

6027-618 [X] Local host not found in node list (local ip interfaces: *interfaceList*)

Explanation: The local host specified in the node list file could not be found.

User response: Check the node list shown by the **mmlscluster** command.

6027-619 Negative grace times are not allowed.

Explanation: The **mmedquota** command received a negative value for the **-t** option.

User response: Reissue the **mmedquota** command with a nonnegative value for grace time.

6027-620 Hard quota limit must not be less than soft limit.

Explanation: The hard quota limit must be greater than or equal to the soft quota limit.

User response: Reissue the **mmedquota** command and enter valid values when editing the information.

6027-621 Negative quota limits are not allowed.

Explanation: The quota value must be positive.

User response: Reissue the **mmedquota** command and enter valid values when editing the information.

6027-622 [E] Failed to join remote cluster *clusterName*

Explanation: The node was not able to establish communication with another cluster, usually while attempting to mount a file system from a remote cluster.

User response: Check other console messages for additional information. Verify that contact nodes for the remote cluster are set correctly. Run **mmremotefs show** and **mmremotecluster show** to display information about the remote cluster.

6027-623 All disks up and ready

Explanation: Self-explanatory.

User response: None. Informational message only.

6027-624 No disks

Explanation: Self-explanatory.

User response: None. Informational message only.

6027-625 File system manager takeover already pending.

Explanation: A request to migrate the file system manager failed because a previous migrate request has not yet completed.

User response: None. Informational message only.

6027-626 Migrate to node *nodeName* already pending.

Explanation: A request to migrate the file system manager failed because a previous migrate request has not yet completed.

User response: None. Informational message only.

6027-627 Node *nodeName* is already manager for *fileSystem*.

Explanation: A request has been made to change the file system manager node to the node that is already the manager.

User response: None. Informational message only.

6027-628 Sending migrate request to current manager node *nodeName*.

Explanation: A request has been made to change the file system manager node.

User response: None. Informational message only.

6027-629 [N] Node *nodeName* **resigned as manager for** *fileSystem***.**

Explanation: Progress report produced by the **mmchmgr** command.

User response: None. Informational message only.

6027-630 [N] Node *nodeName* **appointed as manager for** *fileSystem***.**

Explanation: The **mmchmgr** command successfully changed the node designated as the file system manager.

User response: None. Informational message only.

6027-631 Failed to appoint node *nodeName* **as manager for** *fileSystem***.**

Explanation: A request to change the file system manager node has failed.

User response: Accompanying messages will describe the reason for the failure. Also, see the **mmfs.log** file on the target node.

6027-632 Failed to appoint new manager for *fileSystem***.**

Explanation: An attempt to change the file system manager node has failed.

User response: Accompanying messages will describe the reason for the failure. Also, see the **mmfs.log** file on the target node.

6027-633 The best choice node *nodeName* **is already the manager for** *fileSystem***.**

Explanation: Informational message about the progress and outcome of a migrate request.

User response: None. Informational message only.

6027-634 Node name or number *node* **is not valid.**

Explanation: A node number, IP address, or host name that is not valid has been entered in the configuration file or as input for a command.

User response: Validate your configuration information and the condition of your network. This message may result from an inability to translate a node name.

6027-635 [E] The current file system manager failed and no new manager will be appointed.

Explanation: The file system manager node could not be replaced. This is usually caused by other system errors, such as disk or communication errors.

User response: See accompanying messages for the base failure.

6027-636 [E] Disk marked as stopped or offline.

Explanation: A disk continues to be marked **down** due to a previous error and was not opened again.

User response: Check the disk status by issuing the **mmlsdisk** command, then issue the **mmchdisk start** command to restart the disk.

6027-637 [E] RVSD is not active.

Explanation: The RVSD subsystem needs to be activated.

User response: See the appropriate IBM Reliable Scalable Cluster Technology (RSCT) document (www.ibm.com/support/knowledgecenter/SGVKBA/welcome) and search on *diagnosing IBM Virtual Shared Disk problems*.

6027-638 [E] File system *fileSystem* **unmounted by node** *nodeName*

Explanation: Produced in the console log on a forced unmount of the file system caused by disk or communication failures.

User response: Check the error log on the indicated node. Correct the underlying problem and remount the file system.

6027-639 [E] File system cannot be mounted in restricted mode and ro or rw concurrently

Explanation: There has been an attempt to concurrently mount a file system on separate nodes in both a normal mode and in 'restricted' mode.

User response: Decide which mount mode you want to use, and use that mount mode on both nodes.

6027-640 [E] File system is mounted

Explanation: A command has been issued that requires that the file system be unmounted.

User response: Unmount the file system and reissue the command.

6027-641 [E] Unable to access vital system metadata. Too many disks are unavailable or the file system is corrupted.

Explanation: An attempt has been made to access a file system, but the metadata is unavailable. This can be caused by:

1. The disks on which the metadata resides are either stopped or there was an unsuccessful attempt to delete them.

2. The file system is corrupted.

User response: To access the file system:

1. If the disks are the problem either start the stopped disks or try to delete them.
2. If the file system has been corrupted, you will have to recreate it from backup medium.

6027-642 [N] File system has been deleted.

Explanation: Self-explanatory.

User response: None. Informational message only.

6027-643 [I] Node *nodeName* completed take over for *fileSystem*.

Explanation: The **mmchmgr** command completed successfully.

User response: None. Informational message only.

6027-644 The previous error was detected on node *nodeName*.

Explanation: An unacceptable error was detected. This usually occurs when attempting to retrieve file system information from the operating system's file system database or the cached GPFS system control data. The message identifies the node where the error was encountered.

User response: See accompanying messages for the base failure. A common cause for such errors is lack of space in **/var**.

6027-645 Attention: mmcommon getEFOptions *fileSystem* failed. Checking *fileName*.

Explanation: The names of the disks in a file system were not found in the cached GPFS system data, therefore an attempt will be made to get the information from the operating system's file system database.

User response: If the command fails, see "File system fails to mount" on page 317. A common cause for such errors is lack of space in **/var**.

6027-646 [E] File system unmounted due to loss of cluster membership.

Explanation: Quorum was lost, causing file systems to be unmounted.

User response: Get enough nodes running the GPFS daemon to form a quorum.

6027-647 [E] File *fileName* could not be run with err *errno*.

Explanation: The specified shell script could not be run. This message is followed by the error string that is returned by the exec.

User response: Check file existence and access permissions.

6027-648 EDITOR environment variable must be full pathname.

Explanation: The value of the EDITOR environment variable is not an absolute path name.

User response: Change the value of the EDITOR environment variable to an absolute path name.

6027-649 Error reading the mmpmon command file.

Explanation: An error occurred when reading the **mmpmon** command file.

User response: Check file existence and access permissions.

6027-650 [X] The mmfs daemon is shutting down abnormally.

Explanation: The GPFS daemon is shutting down as a result of an irrecoverable condition, typically a resource shortage.

User response: Review error log entries, correct a resource shortage condition, and restart the GPFS daemon.

6027-660 Error displaying message from mmfsd.

Explanation: GPFS could not properly display an output string sent from the **mmfsd** daemon due to some error. A description of the error follows.

User response: Check that GPFS is properly installed.

6027-661 mmfsd waiting for primary node *nodeName*.

Explanation: The **mmfsd** server has to wait during start up because **mmfsd** on the primary node is not yet ready.

User response: None. Informational message only.

6027-662 mmfsd timed out waiting for primary node *nodeName*.

Explanation: The **mmfsd** server is about to terminate.

User response: Ensure that the **mmfs.cfg** configuration file contains the correct host name or IP

address of the primary node. Check **mmfsd** on the primary node.

6027-663 Lost connection to file system daemon.

Explanation: The connection between a GPFS command and the **mmfsd** daemon has broken. The daemon has probably crashed.

User response: Ensure that the **mmfsd** daemon is running. Check the error log.

6027-664 Unexpected message from file system daemon.

Explanation: The version of the **mmfsd** daemon does not match the version of the GPFS command.

User response: Ensure that all GPFS software components are at the same version.

6027-665 Failed to connect to file system daemon: *errorString*

Explanation: An error occurred while trying to create a session with **mmfsd**.

User response: Ensure that the **mmfsd** daemon is running. Also, only root can run most GPFS commands. The mode bits of the commands must be **set-user-id** to **root**.

6027-666 Failed to determine file system manager.

Explanation: While running a GPFS command in a multiple node configuration, the local file system daemon is unable to determine which node is managing the file system affected by the command.

User response: Check internode communication configuration and ensure that enough GPFS nodes are up to form a quorum.

6027-667 Could not set up socket

Explanation: One of the calls to create or bind the socket used for sending parameters and messages between the command and the daemon failed.

User response: Check additional error messages.

6027-668 Could not send message to file system daemon

Explanation: Attempt to send a message to the file system failed.

User response: Check if the file system daemon is up and running.

6027-669 Could not connect to file system daemon.

Explanation: The TCP connection between the command and the daemon could not be established.

User response: Check additional error messages.

6027-670 Value for '*option***' is not valid. Valid values are** *list***.**

Explanation: The specified value for the given command option was not valid. The remainder of the line will list the valid keywords.

User response: Correct the command line.

6027-671 Keyword missing or incorrect.

Explanation: A missing or incorrect keyword was encountered while parsing command line arguments

User response: Correct the command line.

6027-672 Too few arguments specified.

Explanation: Too few arguments were specified on the command line.

User response: Correct the command line.

6027-673 Too many arguments specified.

Explanation: Too many arguments were specified on the command line.

User response: Correct the command line.

6027-674 Too many values specified for option *name***.**

Explanation: Too many values were specified for the given option on the command line.

User response: Correct the command line.

6027-675 Required value for *option* **is missing.**

Explanation: A required value was not specified for the given option on the command line.

User response: Correct the command line.

6027-676 Option *option* **specified more than once.**

Explanation: The named option was specified more than once on the command line.

User response: Correct the command line.

6027-677 **Option** *option* **is incorrect.**

Explanation: An incorrect option was specified on the command line.

User response: Correct the command line.

6027-678 **Misplaced or incorrect parameter** *name***.**

Explanation: A misplaced or incorrect parameter was specified on the command line.

User response: Correct the command line.

6027-679 **Device** *name* **is not valid.**

Explanation: An incorrect device name was specified on the command line.

User response: Correct the command line.

6027-680 [E] **Disk failure. Volume** *name***. rc =** *value***. Physical volume** *name***.**

Explanation: An I/O request to a disk or a request to fence a disk has failed in such a manner that GPFS can no longer use the disk.

User response: Check the disk hardware and the software subsystems in the path to the disk.

6027-681 **Required option** *name* **was not specified.**

Explanation: A required option was not specified on the command line.

User response: Correct the command line.

6027-682 **Device argument is missing.**

Explanation: The device argument was not specified on the command line.

User response: Correct the command line.

6027-683 **Disk** *name* **is invalid.**

Explanation: An incorrect disk name was specified on the command line.

User response: Correct the command line.

6027-684 **Value** *value* **for** *option* **is incorrect.**

Explanation: An incorrect value was specified for the named option.

User response: Correct the command line.

6027-685 **Value** *value* **for option** *option* **is out of range. Valid values are** *number* **through** *number***.**

Explanation: An out of range value was specified for the named option.

User response: Correct the command line.

6027-686 *option* **(***value***) exceeds option (***value***).**

Explanation: The value of the first option exceeds the value of the second option. This is not permitted.

User response: Correct the command line.

6027-687 **Disk** *name* **is specified more than once.**

Explanation: The named disk was specified more than once on the command line.

User response: Correct the command line.

6027-688 **Failed to read file system descriptor.**

Explanation: The disk block containing critical information about the file system could not be read from disk.

User response: This is usually an error in the path to the disks. If there are associated messages indicating an I/O error such as **ENODEV** or **EIO**, correct that error and retry the operation. If there are no associated I/O errors, then run the **mmfsck** command with the file system unmounted.

6027-689 **Failed to update file system descriptor.**

Explanation: The disk block containing critical information about the file system could not be written to disk.

User response: This is a serious error, which may leave the file system in an unusable state. Correct any I/O errors, then run the **mmfsck** command with the file system unmounted to make repairs.

6027-690 **Failed to allocate I/O buffer.**

Explanation: Could not obtain enough memory (RAM) to perform an operation.

User response: Either retry the operation when the **mmfsd** daemon is less heavily loaded, or increase the size of one or more of the memory pool parameters by issuing the **mmchconfig** command.

6027-691 **Failed to send message to node** *nodeName***.**

Explanation: A message to another file system node could not be sent.

User response: Check additional error message and

the internode communication configuration.

6027-692 **Value for *option* is not valid. Valid values are yes, no.**

Explanation: An option that is required to be **yes** or **no** is neither.

User response: Correct the command line.

6027-693 **Cannot open disk *name*.**

Explanation: Could not access the given disk.

User response: Check the disk hardware and the path to the disk.

6027-694 **Disk not started; disk *name* has a bad volume label.**

Explanation: The volume label on the disk does not match that expected by GPFS.

User response: Check the disk hardware. For hot-pluggable drives, ensure that the proper drive has been plugged in.

6027-695 [E] File system is read-only.

Explanation: An operation was attempted that would require modifying the contents of a file system, but the file system is read-only.

User response: Make the file system R/W before retrying the operation.

6027-696 [E] Too many disks are unavailable.

Explanation: A file system operation failed because all replicas of a data or metadata block are currently unavailable.

User response: Issue the **mmlsdisk** command to check the availability of the disks in the file system; correct disk hardware problems, and then issue the **mmchdisk** command with the **start** option to inform the file system that the disk or disks are available again.

6027-697 [E] No log available.

Explanation: A file system operation failed because no space for logging metadata changes could be found.

User response: Check additional error message. A likely reason for this error is that all disks with available log space are currently unavailable.

6027-698 [E] Not enough memory to allocate internal data structure.

Explanation: A file system operation failed because no memory is available for allocating internal data structures.

User response: Stop other processes that may have main memory pinned for their use.

6027-699 [E] Inconsistency in file system metadata.

Explanation: File system metadata on disk has been corrupted.

User response: This is an extremely serious error that may cause loss of data. Issue the **mmfsck** command with the file system unmounted to make repairs. There will be a **POSSIBLE FILE CORRUPTION** entry in the system error log that should be forwarded to the IBM Support Center.

6027-700 [E] Log recovery failed.

Explanation: An error was encountered while restoring file system metadata from the log.

User response: Check additional error message. A likely reason for this error is that none of the replicas of the log could be accessed because too many disks are currently unavailable. If the problem persists, issue the **mmfsck** command with the file system unmounted.

6027-701 [X] Some file system data are inaccessible at this time.

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an occurred that left the file system in an unusable state.

User response: Possible reasons include too many unavailable disks or insufficient memory for file system control structures. Check other error messages as well as the error log for additional information. Unmount the file system and correct any I/O errors. Then remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-702 [X] Some file system data are inaccessible at this time. Check error log for additional information. After correcting the problem, the file system must be unmounted and then mounted to restore normal data access.

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state.

User response: Possible reasons include too many unavailable disks or insufficient memory for file system control structures. Check other error messages as well as the error log for additional information. Unmount the file system and correct any I/O errors. Then remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-703 [X] **Some file system data are inaccessible at this time. Check error log for additional information.**

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state.

User response: Possible reasons include too many unavailable disks or insufficient memory for file system control structures. Check other error messages as well as the error log for additional information. Unmount the file system and correct any I/O errors. Then remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-704 **Attention: Due to an earlier error normal access to this file system has been disabled. Check error log for additional information. After correcting the problem, the file system must be unmounted and then mounted again to restore normal data access.**

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state.

User response: Possible reasons include too many unavailable disks or insufficient memory for file system control structures. Check other error messages as well as the error log for additional information. Unmount the file system and correct any I/O errors. Then remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-705 **Error code** *value***.**

Explanation: Provides additional information about an error.

User response: See accompanying error messages.

6027-706 **The device** *name* **has no corresponding entry in** *fileName* **or has an incomplete entry.**

Explanation: The command requires a device that has a file system associated with it.

User response: Check the operating system's file system database (the given file) for a valid device entry.

6027-707 **Unable to open file** *fileName***.**

Explanation: The named file cannot be opened.

User response: Check that the file exists and has the correct permissions.

6027-708 **Keyword** *name* **is incorrect. Valid values are** *list***.**

Explanation: An incorrect keyword was encountered.

User response: Correct the command line.

6027-709 **Incorrect response. Valid responses are "yes", "no", or "noall"**

Explanation: A question was asked that requires a **yes** or **no** answer. The answer entered was neither **yes**, **no**, nor **noall**.

User response: Enter a valid response.

6027-710 **Attention:**

Explanation: Precedes an attention messages.

User response: None. Informational message only.

6027-711 [E] **Specified entity, such as a disk or file system, does not exist.**

Explanation: A file system operation failed because the specified entity, such as a disk or file system, could not be found.

User response: Specify existing disk, file system, etc.

6027-712 [E] **Error in communications between mmfsd daemon and client program.**

Explanation: A message sent between the **mmfsd** daemon and the client program had an incorrect format or content.

User response: Verify that the **mmfsd** daemon is running.

6027-713 **Unable to start because conflicting program** *name* **is running. Waiting until it completes.**

Explanation: A program detected that it cannot start because a conflicting program is running. The program will automatically start once the conflicting program has ended, as long as there are no other conflicting programs running at that time.

User response: None. Informational message only.

6027-714 **Terminating because conflicting program** *name* **is running.**

Explanation: A program detected that it must terminate because a conflicting program is running.

User response: Reissue the command once the conflicting program has ended.

6027-715 *command* **is finished waiting. Starting execution now.**

Explanation: A program detected that it can now begin running because a conflicting program has ended.

User response: None. Information message only.

6027-716 [E] Some file system data or metadata has been lost.

Explanation: Unable to access some piece of file system data that has been lost due to the deletion of disks beyond the replication factor.

User response: If the function did not complete, try to mount the file system in **restricted** mode.

6027-717 [E] Must execute mmfsck before mount.

Explanation: An attempt has been made to mount a file system on which an incomplete **mmfsck** command was run.

User response: Reissue the **mmfsck** command to the repair file system, then reissue the **mount** command.

6027-718 The mmfsd daemon is not ready to handle commands yet.

Explanation: The **mmfsd** daemon is not accepting messages because it is restarting or stopping.

User response: None. Informational message only.

6027-719 [E] Device type not supported.

Explanation: A disk being added to a file system with the **mmadddisk** or **mmcrfs** command is not a character mode special file, or has characteristics not recognized by GPFS.

User response: Check the characteristics of the disk being added to the file system.

6027-720 [E] Actual sector size does not match given sector size.

Explanation: A disk being added to a file system with the **mmadddisk** or **mmcrfs** command has a physical sector size that differs from that given in the disk description list.

User response: Check the physical sector size of the disk being added to the file system.

6027-721 [E] Host '*name***' in** *fileName* **is not valid.**

Explanation: A host name or IP address that is not valid was found in a configuration file.

User response: Check the configuration file specified in the error message.

6027-722 Attention: Due to an earlier error normal access to this file system has been disabled. Check error log for additional information. The file system must be mounted again to restore normal data access.

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state. Possible reasons include too many unavailable disks or insufficient memory for file system control structures.

User response: Check other error messages as well as the error log for additional information. Correct any I/O errors. Then, remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-723 Attention: Due to an earlier error normal access to this file system has been disabled. Check error log for additional information. After correcting the problem, the file system must be mounted again to restore normal data access.

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state. Possible reasons include too many unavailable disks or insufficient memory for file system control structures.

User response: Check other error messages as well as the error log for additional information. Correct any I/O errors. Then, remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-724 [E] Incompatible file system format.

Explanation: An attempt was made to access a file system that was formatted with an older version of the product that is no longer compatible with the version currently running.

User response: To change the file system format version to the current version, issue the **-V** option on the **mmchfs** command.

6027-725 The mmfsd daemon is not ready to handle commands yet. Waiting for quorum.

Explanation: The GPFS **mmfsd** daemon is not accepting messages because it is waiting for quorum.

User response: Determine why insufficient nodes have

joined the group to achieve quorum and rectify the problem.

6027-726 [E] Quota initialization/start-up failed.

Explanation: Quota manager initialization was unsuccessful. The file system manager finished without quotas. Subsequent client mount requests will fail.

User response: Check the error log and correct I/O errors. It may be necessary to issue the **mmcheckquota** command with the file system unmounted.

6027-727 Specified driver type *type* **does not match disk** *name* **driver type** *type***.**

Explanation: The driver type specified on the **mmchdisk** command does not match the current driver type of the disk.

User response: Verify the driver type and reissue the command.

6027-728 Specified sector size *value* **does not match disk** *name* **sector size** *value***.**

Explanation: The sector size specified on the **mmchdisk** command does not match the current sector size of the disk.

User response: Verify the sector size and reissue the command.

6027-729 Attention: No changes for disk *name* **were specified.**

Explanation: The disk descriptor in the **mmchdisk** command does not specify that any changes are to be made to the disk.

User response: Check the disk descriptor to determine if changes are needed.

6027-730 *command* **on** *fileSystem***.**

Explanation: Quota was activated or deactivated as stated as a result of the **mmquotaon**, **mmquotaoff**, **mmdefquotaon**, or **mmdefquotaoff** commands.

User response: None, informational only. This message is enabled with the **-v** option on the **mmquotaon**, **mmquotaoff**, **mmdefquotaon**, or **mmdefquotaoff** commands.

6027-731 Error *number* **while performing** *command* **for** *name* **quota on** *fileSystem*

Explanation: An error occurred when switching quotas of a certain type on or off. If errors were returned for multiple file systems, only the error code is shown.

User response: Check the error code shown by the message to determine the reason.

6027-732 Error while performing *command* **on** *fileSystem***.**

Explanation: An error occurred while performing the stated command when listing or reporting quotas.

User response: None. Informational message only.

6027-733 Edit quota: Incorrect format!

Explanation: The format of one or more edited quota limit entries was not correct.

User response: Reissue the **mmedquota** command. Change only the values for the limits and follow the instructions given.

6027-734 [W] Quota check for '*fileSystem***' ended prematurely.**

Explanation: The user interrupted and terminated the command.

User response: If ending the command was not intended, reissue the **mmcheckquota** command.

6027-735 Error editing string from mmfsd.

Explanation: An internal error occurred in the **mmfsd** when editing a string.

User response: None. Informational message only.

6027-736 Attention: Due to an earlier error normal access to this file system has been disabled. Check error log for additional information. The file system must be unmounted and then mounted again to restore normal data access.

Explanation: The file system has encountered an error that is serious enough to make some or all data inaccessible. This message indicates that an error occurred that left the file system in an unusable state. Possible reasons include too many unavailable disks or insufficient memory for file system control structures.

User response: Check other error messages as well as the error log for additional information. Unmount the file system and correct any I/O errors. Then, remount the file system and try the operation again. If the problem persists, issue the **mmfsck** command with the file system unmounted to make repairs.

6027-737 Attention: No metadata disks remain.

Explanation: The **mmchdisk** command has been issued, but no metadata disks remain.

User response: None. Informational message only.

6027-738 **Attention: No data disks remain.**

Explanation: The **mmchdisk** command has been issued, but no data disks remain.

User response: None. Informational message only.

6027-739 **Attention: Due to an earlier configuration change the file system is no longer properly balanced.**

Explanation: The **mmlsdisk** command found that the file system is not properly balanced.

User response: Issue the **mmrestripefs -b** command at your convenience.

6027-740 **Attention: Due to an earlier configuration change the file system is no longer properly replicated.**

Explanation: The **mmlsdisk** command found that the file system is not properly replicated.

User response: Issue the **mmrestripefs -r** command at your convenience

6027-741 **Attention: Due to an earlier configuration change the file system may contain data that is at risk of being lost.**

Explanation: The **mmlsdisk** command found that critical data resides on disks that are suspended or being deleted.

User response: Issue the **mmrestripefs -m** command as soon as possible.

6027-742 **Error occurred while executing a command for** *fileSystem***.**

Explanation: A quota command encountered a problem on a file system. Processing continues with the next file system.

User response: None. Informational message only.

6027-743 **Initial disk state was updated successfully, but another error may have changed the state again.**

Explanation: The **mmchdisk** command encountered an error after the disk status or availability change was already recorded in the file system configuration. The most likely reason for this problem is that too many disks have become unavailable or are still unavailable after the disk state change.

User response: Issue an **mmchdisk start** command when more disks are available.

6027-744 **Unable to run** *command* **while the file system is mounted in restricted mode.**

Explanation: A command that can alter the data in a file system was issued while the file system was mounted in restricted mode.

User response: Mount the file system in read-only or read-write mode or unmount the file system and then reissue the command.

6027-745 *fileSystem***: no** *quotaType* **quota management enabled.**

Explanation: A quota command of the cited type was issued for the cited file system when no quota management was enabled.

User response: Enable quota management and reissue the command.

6027-746 **Editing quota limits for this user or group not permitted.**

Explanation: The **root** user or **system** group was specified for quota limit editing in the **mmedquota** command.

User response: Specify a valid user or group in the **mmedquota** command. Editing quota limits for the **root** user or **system** group is prohibited.

6027-747 [E] **Too many nodes in cluster (max** *number***) or file system (max** *number***).**

Explanation: The operation cannot succeed because too many nodes are involved.

User response: Reduce the number of nodes to the applicable stated limit.

6027-748 *fileSystem***: no quota management enabled**

Explanation: A quota command was issued for the cited file system when no quota management was enabled.

User response: Enable quota management and reissue the command.

6027-749 **Pool size changed to** *number* **K =** *number* **M.**

Explanation: Pool size successfully changed.

User response: None. Informational message only.

6027-750 [E] The node address *ipAddress* is not defined in the node list

Explanation: An address does not exist in the GPFS configuration file.

User response: Perform required configuration steps prior to starting GPFS on the node.

6027-751 [E] Error code *value*

Explanation: Provides additional information about an error.

User response: See accompanying error messages.

6027-752 [E] Lost membership in cluster *clusterName*. Unmounting file systems.

Explanation: This node has lost membership in the cluster. Either GPFS is no longer available on enough nodes to maintain quorum, or this node could not communicate with other members of the quorum. This could be caused by a communications failure between nodes, or multiple GPFS failures.

User response: See associated error logs on the failed nodes for additional problem determination information.

6027-753 [E] Could not run command *command*

Explanation: The GPFS daemon failed to run the specified command.

User response: Verify correct installation.

6027-754 Error reading string for mmfsd.

Explanation: GPFS could not properly read an input string.

User response: Check that GPFS is properly installed.

6027-755 [I] Waiting for challenge *challengeValue* (node *nodeNumber*, sequence *sequenceNumber*) to be responded during disk election

Explanation: The node has challenged another node, which won the previous election and is waiting for the challenger to respond.

User response: None. Informational message only.

6027-756 [E] Configuration invalid or inconsistent between different nodes.

Explanation: Self-explanatory.

User response: Check cluster and file system configuration.

6027-757 *name* is not an excluded disk.

Explanation: Some of the disks passed to the **mmfsctl include** command are not marked as **excluded** in the **mmsdrfs** file.

User response: Verify the list of disks supplied to this command.

6027-758 Disk(s) not started; disk *name* has a bad volume label.

Explanation: The volume label on the disk does not match that expected by GPFS.

User response: Check the disk hardware. For hot-pluggable drives, make sure the proper drive has been plugged in.

6027-759 *fileSystem* is still in use.

Explanation: The **mmfsctl include** command found that the named file system is still mounted, or another GPFS command is running against the file system.

User response: Unmount the file system if it is mounted, or wait for GPFS commands in progress to terminate before retrying the command.

6027-760 [E] Unable to perform i/o to the disk. This node is either fenced from accessing the disk or this node's disk lease has expired.

Explanation: A read or write to the disk failed due to either being fenced from the disk or no longer having a disk lease.

User response: Verify disk hardware fencing setup is correct if being used. Ensure network connectivity between this node and other nodes is operational.

6027-761 [W] Attention: excessive timer drift between *node* and *node* (*number* over *number* sec).

Explanation: GPFS has detected an unusually large difference in the rate of clock ticks (as returned by the **times()** system call) between two nodes. Another node's TOD clock and tick rate changed dramatically relative to this node's TOD clock and tick rate.

User response: Check error log for hardware or device driver problems that might cause timer interrupts to be lost or a recent large adjustment made to the TOD clock.

6027-762 No quota enabled file system found.

Explanation: There is no quota-enabled file system in this cluster.

User response: None. Informational message only.

6027-763 **uidInvalidate: Incorrect option** *option***.**

Explanation: An incorrect option passed to the **uidinvalidate** command.

User response: Correct the command invocation.

6027-764 **Error invalidating UID remapping cache for** *domain***.**

Explanation: An incorrect domain name passed to the **uidinvalidate** command.

User response: Correct the command invocation.

6027-765 [W] **Tick value hasn't changed for nearly** *number* **seconds**

Explanation: Clock ticks incremented by AIX have not been incremented.

User response: Check the error log for hardware or device driver problems that might cause timer interrupts to be lost.

6027-766 [N] **This node will be expelled from cluster** *cluster* **due to expel msg from** *node*

Explanation: This node is being expelled from the cluster.

User response: Check the network connection between this node and the node specified above.

6027-767 [N] **Request sent to** *node* **to expel** *node* **from cluster** *cluster*

Explanation: This node sent an expel request to the cluster manager node to expel another node.

User response: Check network connection between this node and the node specified above.

6027-768 **Wrong number of operands for mmpmon command '***command***'.**

Explanation: The command read from the input file has the wrong number of operands.

User response: Correct the command invocation and reissue the command.

6027-769 **Malformed mmpmon command '***command***'.**

Explanation: The command read from the input file is malformed, perhaps with an unknown keyword.

User response: Correct the command invocation and reissue the command.

6027-770 **Error writing user.quota file.**

Explanation: An error occurred while writing the cited quota file.

User response: Check the status and availability of the disks and reissue the command.

6027-771 **Error writing group.quota file.**

Explanation: An error occurred while writing the cited quota file.

User response: Check the status and availability of the disks and reissue the command.

6027-772 **Error writing fileset.quota file.**

Explanation: An error occurred while writing the cited quota file.

User response: Check the status and availability of the disks and reissue the command.

6027-773 *fileSystem***: quota check may be incomplete because of SANergy activity on** *number* **files.**

Explanation: The online quota check may be incomplete due to active SANergy activities on the file system.

User response: Reissue the quota check when there is no SANergy activity.

6027-774 *fileSystem***: quota management is not enabled, or one or more quota clients are not available.**

Explanation: An attempt was made to perform quotas commands without quota management enabled, or one or more quota clients failed during quota check.

User response: Correct the cause of the problem, and then reissue the quota command.

6027-775 **During mmcheckquota processing,** *number* **node(s) failed. It is recommended that mmcheckquota be repeated.**

Explanation: Nodes failed while an online quota check was running.

User response: Reissue the quota check command.

6027-776 *fileSystem***: There was not enough space for the report. Please repeat quota check!**

Explanation: The **vflag** is set in the **tscheckquota** command, but either no space or not enough space could be allocated for the differences to be printed.

User response: Correct the space problem and reissue the quota check.

6027-777 [I] Recovering nodes: *nodeList*

Explanation: Recovery for one or more nodes has begun.

User response: No response is needed if this message is followed by 'recovered nodes' entries specifying the nodes. If this message is not followed by such a message, determine why recovery did not complete.

6027-778 [I] Recovering nodes in cluster *cluster***:** *nodeList*

Explanation: Recovery for one or more nodes in the cited cluster has begun.

User response: No response is needed if this message is followed by 'recovered nodes' entries on the cited cluster specifying the nodes. If this message is not followed by such a message, determine why recovery did not complete.

6027-779 Incorrect fileset name *filesetName***.**

Explanation: The fileset name provided on the command line is incorrect.

User response: Correct the fileset name and reissue the command.

6027-780 Incorrect path to fileset junction *junctionName***.**

Explanation: The path to the fileset junction is incorrect.

User response: Correct the junction path and reissue the command.

6027-781 Storage pools have not been enabled for file system *fileSystem***.**

Explanation: The user invoked a command with a storage pool option (**-p** or **-P**) before storage pools were enabled.

User response: Enable storage pools with the **mmchfs -V** command, or correct the command invocation and reissue the command.

6027-784 [E] Device not ready.

Explanation: A device is not ready for operation.

User response: Check previous messages for further information.

6027-785 [E] Cannot establish connection.

Explanation: This node cannot establish a connection to another node.

User response: Check previous messages for further information.

6027-786 [E] Message failed because the destination node refused the connection.

Explanation: This node sent a message to a node that refuses to establish a connection.

User response: Check previous messages for further information.

6027-787 [E] Security configuration data is inconsistent or unavailable.

Explanation: There was an error configuring security on this node.

User response: Check previous messages for further information.

6027-788 [E] Failed to load or initialize security library.

Explanation: There was an error loading or initializing the security library on this node.

User response: Check previous messages for further information.

6027-789 Unable to read offsets *offset* **to** *offset* **for inode** *inode* **snap** *snap***, from disk** *diskName***, sector** *sector***.**

Explanation: The **mmdeldisk -c** command found that the cited addresses on the cited disk represent data that is no longer readable.

User response: Save this output for later use in cleaning up failing disks.

6027-790 Specified storage pool *poolName* **does not match disk** *diskName* **storage pool** *poolName***. Use mmdeldisk and mmadddisk to change a disk's storage pool.**

Explanation: An attempt was made to change a disk's storage pool assignment using the **mmchdisk** command. This can only be done by deleting the disk from its current storage pool and then adding it to the new pool.

User response: Delete the disk from its current storage pool and then add it to the new pool.

6027-792 **Policies have not been enabled for file system** *fileSystem***.**

Explanation: The cited file system must be upgraded to use policies.

User response: Upgrade the file system via the **mmchfs -V** command.

6027-793 **No policy file was installed for file system** *fileSystem***.**

Explanation: No policy file was installed for this file system.

User response: Install a policy file.

6027-794 **Failed to read policy file for file system** *fileSystem***.**

Explanation: Failed to read the policy file for the requested file system.

User response: Reinstall the policy file.

6027-795 **Failed to open** *fileName***:** *errorCode***.**

Explanation: An incorrect file name was specified to **tschpolicy**.

User response: Correct the command invocation and reissue the command.

6027-796 **Failed to read** *fileName***:** *errorCode***.**

Explanation: An incorrect file name was specified to **tschpolicy**.

User response: Correct the command invocation and reissue the command.

6027-797 **Failed to stat** *fileName***:** *errorCode***.**

Explanation: An incorrect file name was specified to **tschpolicy**.

User response: Correct the command invocation and reissue the command.

6027-798 **Policy files are limited to** *number* **bytes.**

Explanation: A user-specified policy file exceeded the maximum-allowed length.

User response: Install a smaller policy file.

6027-799 **Policy `***policyName***' installed and broadcast to all nodes.**

Explanation: Self-explanatory.

User response: None. Informational message only.

6027-850 **Unable to issue this command from a non-root user.**

Explanation: **tsiostat** requires root privileges to run.

User response: Get the system administrator to change the executable to set the UID to 0.

6027-851 **Unable to process interrupt received.**

Explanation: An interrupt occurred that **tsiostat** cannot process.

User response: Contact the IBM Support Center.

6027-852 **interval and count must be positive integers.**

Explanation: Incorrect values were supplied for **tsiostat** parameters.

User response: Correct the command invocation and reissue the command.

6027-853 **interval must be less than 1024.**

Explanation: An incorrect value was supplied for the interval parameter.

User response: Correct the command invocation and reissue the command.

6027-854 **count must be less than 1024.**

Explanation: An incorrect value was supplied for the count parameter.

User response: Correct the command invocation and reissue the command.

6027-855 **Unable to connect to server, mmfsd is not started.**

Explanation: The **tsiostat** command was issued but the file system is not started.

User response: Contact your system administrator.

6027-856 **No information to report.**

Explanation: The **tsiostat** command was issued but no file systems are mounted.

User response: Contact your system administrator.

6027-857 **Error retrieving values.**

Explanation: The **tsiostat** command was issued and an internal error occurred.

User response: Contact the IBM Support Center.

6027-858 File system not mounted.

Explanation: The requested file system is not mounted.

User response: Mount the file system and reattempt the failing operation.

6027-859 Set DIRECTIO failed

Explanation: The **tsfattr** call failed.

User response: Check for additional error messages. Resolve the problems before reattempting the failing operation.

6027-860 -d is not appropriate for an NFSv4 ACL

Explanation: Produced by the **mmgetacl** or **mmputacl** commands when the **-d** option was specified, but the object has an NFS Version 4 ACL (does not have a default).

User response: None. Informational message only.

6027-861 Set afm ctl failed

Explanation: The **tsfattr** call failed.

User response: Check for additional error messages. Resolve the problems before reattempting the failing operation.

6027-862 Incorrect storage pool name *poolName***.**

Explanation: An incorrect storage pool name was provided.

User response: Determine the correct storage pool name and reissue the command.

6027-863 File cannot be assigned to storage pool '*poolName***'.**

Explanation: The file cannot be assigned to the specified pool.

User response: Determine the correct storage pool name and reissue the command.

6027-864 Set storage pool failed.

Explanation: An incorrect storage pool name was provided.

User response: Determine the correct storage pool name and reissue the command.

6027-865 Restripe file data failed.

Explanation: An error occurred while restriping the file data.

User response: Check the error code and reissue the command.

6027-866 [E] Storage pools have not been enabled for this file system.

Explanation: The user invoked a command with a storage pool option (**-p** or **-P**) before storage pools were enabled.

User response: Enable storage pools via **mmchfs -V**, or correct the command invocation and reissue the command.

6027-867 Change storage pool is not permitted.

Explanation: The user tried to change a file's assigned storage pool but was not root or superuser.

User response: Reissue the command as root or superuser.

6027-868 mmchattr failed.

Explanation: An error occurred while changing a file's attributes.

User response: Check the error code and reissue the command.

6027-869 File replication exceeds number of failure groups in destination storage pool.

Explanation: The **tschattr** command received incorrect command line arguments.

User response: Correct the command invocation and reissue the command.

6027-870 [E] Error on getcwd(): *errorString***. Try an absolute path instead of just** *pathName*

Explanation: The **getcwd** system call failed.

User response: Specify an absolute path starting with '/' on the command invocation, so that the command will not need to invoke **getcwd**.

6027-871 [E] Error on gpfs_get_pathname_from_ fssnaphandle(*pathName***):** *errorString***.**

Explanation: An error occurred during a **gpfs_get_pathname_from_fssnaphandle** operation.

User response: Verify the invocation parameters and make sure the command is running under a user ID with sufficient authority (**root** or administrator privileges). Specify a GPFS file system device name or a GPFS directory path name as the first argument. Correct the command invocation and reissue the command.

6027-872 [E] *pathName* **is not within a mounted GPFS file system.**

Explanation: An error occurred while attempting to access the named GPFS file system or path.

User response: Verify the invocation parameters and make sure the command is running under a user ID with sufficient authority (**root** or administrator privileges). Mount the GPFS file system. Correct the command invocation and reissue the command.

6027-873 [W] **Error on gpfs_stat_inode([***pathName***/***fileName***],***inodeNumber***.***genNumber***):** *errorString*

Explanation: An error occurred during a **gpfs_stat_inode** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-874 [E] **Error: incorrect Date@Time (YYYY-MM-DD@HH:MM:SS) specification:** *specification*

Explanation: The *Date@Time* command invocation argument could not be parsed.

User response: Correct the command invocation and try again. The syntax should look similar to: **2005-12-25@07:30:00**.

6027-875 [E] **Error on gpfs_stat(***pathName***):** *errorString*

Explanation: An error occurred while attempting to **stat()** the cited path name.

User response: Determine whether the cited path name exists and is accessible. Correct the command arguments as necessary and reissue the command.

6027-876 [E] **Error starting directory scan(***pathName***):** *errorString*

Explanation: The specified path name is not a directory.

User response: Determine whether the specified path name exists and is an accessible directory. Correct the command arguments as necessary and reissue the command.

6027-877 [E] **Error opening** *pathName***:** *errorString*

Explanation: An error occurred while attempting to open the named file. Its pool and replication attributes remain unchanged.

User response: Investigate the file and possibly reissue the command. The file may have been removed or locked by another application.

6027-878 [E] **Error on gpfs_fcntl(***pathName***):** *errorString* **(offset=***offset***)**

Explanation: An error occurred while attempting **fcntl** on the named file. Its pool or replication attributes may not have been adjusted.

User response: Investigate the file and possibly reissue the command. Use the **mmlsattr** and **mmchattr** commands to examine and change the pool and replication attributes of the named file.

6027-879 [E] **Error deleting** *pathName***:** *errorString*

Explanation: An error occurred while attempting to delete the named file.

User response: Investigate the file and possibly reissue the command. The file may have been removed or locked by another application.

6027-880 **Error on gpfs_seek_inode(***inodeNumber***):** *errorString*

Explanation: An error occurred during a **gpfs_seek_inode** operation.

User response: Reissue the command. If the problem persists, contact the contact the IBM Support Center

6027-881 [E] **Error on gpfs_iopen([***rootPath***/***pathName***],***inodeNumber***):** *errorString*

Explanation: An error occurred during a **gpfs_iopen** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-882 [E] **Error on gpfs_ireaddir(***rootPath***/***pathName***):** *errorString*

Explanation: An error occurred during a **gpfs_ireaddir()** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-883 **Error on gpfs_next_inode(***maxInodeNumber***):** *errorString*

Explanation: An error occurred during a **gpfs_next_inode** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-884 [E:nnn] Error during directory scan

Explanation: A terminal error occurred during the directory scan phase of the command.

User response: Verify the command arguments. Reissue the command. If the problem persists, contact the IBM Support Center.

6027-885 [E:nnn] Error during inode scan: *errorString*

Explanation: A terminal error occurred during the inode scan phase of the command.

User response: Verify the command arguments. Reissue the command. If the problem persists, contact the IBM Support Center.

6027-886 [E:nnn] Error during policy decisions scan

Explanation: A terminal error occurred during the policy decisions phase of the command.

User response: Verify the command arguments. Reissue the command. If the problem persists, contact the IBM Support Center.

6027-887 [W] Error on gpfs_igetstoragepool(*dataPoolId***):** *errorString*

Explanation: An error occurred during a **gpfs_igetstoragepool** operation. Possible inode corruption.

User response: Use **mmfsck** command. If the problem persists, contact the IBM Support Center.

6027-888 [W] Error on gpfs_igetfilesetname(*filesetId***):** *errorString*

Explanation: An error occurred during a **gpfs_igetfilesetname** operation. Possible inode corruption.

User response: Use **mmfsck** command. If the problem persists, contact the IBM Support Center.

6027-889 [E] Error on gpfs_get_fssnaphandle(*rootPath***):** *errorString***.**

Explanation: An error occurred during a **gpfs_get_fssnaphandle** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-890 [E] Error on gpfs_open_inodescan(*rootPath***):** *errorString*

Explanation: An error occurred during a **gpfs_open_inodescan()** operation.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-891 [X] WEIGHT(*thresholdValue***) UNKNOWN** *pathName*

Explanation: The named file was assigned the indicated weight, but the rule type is **UNKNOWN**.

User response: Contact the IBM Support Center.

6027-892 [E] Error on pthread_create: *where* **#***threadNumber_or_portNumber_or_socketNumber***:** *errorString*

Explanation: An error occurred while creating the thread during a **pthread_create** operation.

User response: Consider some of the command parameters that might affect memory usage. For further assistance, contact the IBM Support Center.

6027-893 [X] Error on pthread_mutex_init: *errorString*

Explanation: An error occurred during a **pthread_mutex_init** operation.

User response: Contact the IBM Support Center.

6027-894 [X] Error on pthread_mutex_lock: *errorString*

Explanation: An error occurred during a **pthread_mutex_lock** operation.

User response: Contact the IBM Support Center.

6027-895 [X] Error on pthread_mutex_unlock: *errorString*

Explanation: An error occurred during a **pthread_mutex_unlock** operation.

User response: Contact the IBM Support Center.

6027-896 [X] Error on pthread_cond_init: *errorString*

Explanation: An error occurred during a **pthread_cond_init** operation.

User response: Contact the IBM Support Center.

6027-897 [X] Error on pthread_cond_signal: *errorString*

Explanation: An error occurred during a **pthread_cond_signal** operation.

User response: Contact the IBM Support Center.

6027-898 [X] Error on pthread_cond_broadcast: *errorString*

Explanation: An error occurred during a **pthread_cond_broadcast** operation.

User response: Contact the IBM Support Center.

6027-899 [X] Error on pthread_cond_wait: *errorString*

Explanation: An error occurred during a **pthread_cond_wait** operation.

User response: Contact the IBM Support Center.

6027-900 [E] Error opening work file *fileName***:** *errorString*

Explanation: An error occurred while attempting to open the named work file.

User response: Investigate the file and possibly reissue the command. Check that the path name is defined and accessible.

6027-901 [E] Error writing to work file *fileName***:** *errorString*

Explanation: An error occurred while attempting to write to the named work file.

User response: Investigate the file and possibly reissue the command. Check that there is sufficient free space in the file system.

6027-902 [E] Error parsing work file *fileName***. Service index:** *number*

Explanation: An error occurred while attempting to read the specified work file.

User response: Investigate the file and possibly reissue the command. Make sure that there is enough free space in the file system. If the error persists, contact the IBM Support Center.

6027-903 [E:nnn] Error while loading policy rules.

Explanation: An error occurred while attempting to read or parse the policy file, which may contain syntax errors. Subsequent messages include more information about the error.

User response: Read all of the related error messages and try to correct the problem.

6027-904 [E] Error *returnCode* **from PD writer for inode=***inodeNumber* **pathname=***pathName*

Explanation: An error occurred while writing the policy decision for the candidate file with the indicated inode number and path name to a work file. There probably will be related error messages.

User response: Read all the related error messages. Attempt to correct the problems.

6027-905 [E] Error: Out of memory. Service index: *number*

Explanation: The command has exhausted virtual memory.

User response: Consider some of the command parameters that might affect memory usage. For further assistance, contact the IBM Support Center.

6027-906 [E:nnn] Error on system(*command***)**

Explanation: An error occurred during the system call with the specified argument string.

User response: Read and investigate related error messages.

6027-907 [E:nnn] Error from sort_file(*inodeListname***,** *sortCommand***,***sortInodeOptions***,***tempDir***)**

Explanation: An error occurred while sorting the named work file using the named **sort** command with the given options and working directory.

User response: Check these:

- The **sort** command is installed on your system.
- The **sort** command supports the given options.
- The working directory is accessible.
- The file system has sufficient free space.

6027-908 [W] Attention: In RULE '*ruleName***' (***ruleNumber***), the pool named by "***poolName* '*poolType*"' is not defined in the file system.**

Explanation: The cited pool is not defined in the file system.

User response: Correct the rule and reissue the command.

This is not an irrecoverable error; the command will continue to run. Of course it will not find any files in an incorrect **FROM POOL** and it will not be able to migrate any files to an incorrect **TO POOL**.

6027-909 [E] Error on pthread_join: *where* **#***threadNumber***:** *errorString*

Explanation: An error occurred while reaping the thread during a **pthread_join** operation.

User response: Contact the IBM Support Center.

6027-910 [E:nnn] Error during policy execution

Explanation: A terminating error occurred during the policy execution phase of the command.

User response: Verify the command arguments and reissue the command. If the problem persists, contact the IBM Support Center.

6027-911 [E] Error on *changeSpecification* **change for** *pathName***.** *errorString*

Explanation: This message provides more details about a **gpfs_fcntl()** error.

User response: Use the **mmlsattr** and **mmchattr** commands to examine the file, and then reissue the change command.

6027-912 [E] Error on restriping of *pathName***.** *errorString*

Explanation: This provides more details on a **gpfs_fcntl()** error.

User response: Use the **mmlsattr** and **mmchattr** commands to examine the file and then reissue the restriping command.

6027-913 Desired replication exceeds number of failure groups.

Explanation: While restriping a file, the **tschattr** or **tsrestripefile** command found that the desired replication exceeded the number of failure groups.

User response: Reissue the command after adding or restarting file system disks.

6027-914 Insufficient space in one of the replica failure groups.

Explanation: While restriping a file, the **tschattr** or **tsrestripefile** command found there was insufficient space in one of the replica failure groups.

User response: Reissue the command after adding or restarting file system disks.

6027-915 Insufficient space to properly balance file.

Explanation: While restriping a file, the **tschattr** or **tsrestripefile** command found that there was insufficient space to properly balance the file.

User response: Reissue the command after adding or restarting file system disks.

6027-916 Too many disks unavailable to properly balance file.

Explanation: While restriping a file, the **tschattr** or **tsrestripefile** command found that there were too many disks unavailable to properly balance the file.

User response: Reissue the command after adding or restarting file system disks.

6027-917 All replicas of a data block were previously deleted.

Explanation: While restriping a file, the **tschattr** or **tsrestripefile** command found that all replicas of a data block were previously deleted.

User response: Reissue the command after adding or restarting file system disks.

6027-918 Cannot make this change to a nonzero length file.

Explanation: GPFS does not support the requested change to the replication attributes.

User response: You may want to create a new file with the desired attributes and then copy your data to that file and rename it appropriately. Be sure that there are sufficient disks assigned to the pool with different failure groups to support the desired replication attributes.

6027-919 Replication parameter range error (*value***,** *value***).**

Explanation: Similar to message **6027-918**. The (a,b) numbers are the allowable range of the replication attributes.

User response: You may want to create a new file with the desired attributes and then copy your data to that file and rename it appropriately. Be sure that there are sufficient disks assigned to the pool with different failure groups to support the desired replication attributes.

6027-920 [E] Error on pthread_detach(self): *where***:** *errorString*

Explanation: An error occurred during a **pthread_detach** operation.

User response: Contact the IBM Support Center.

6027-921 [E] Error on socket *socketName***(***hostName***):** *errorString*

Explanation: An error occurred during a socket operation.

User response: Verify any command arguments related to interprocessor communication and then reissue the command. If the problem persists, contact the IBM Support Center.

6027-922 [X] Error in Mtconx - p_accepts should not be empty

Explanation: The program discovered an inconsistency or logic error within itself.

User response: Contact the IBM Support Center.

6027-923 [W] Error - command client is an incompatible version: *hostName protocolVersion*

Explanation: While operating in master/client mode, the command discovered that the client is running an incompatible version.

User response: Ensure the same version of the command software is installed on all nodes in the clusters and then reissue the command.

6027-924 [X] Error - unrecognized client response from *hostName*: *clientResponse*

Explanation: Similar to message **6027-923**, except this may be an internal logic error.

User response: Ensure the latest, same version software is installed on all nodes in the clusters and then reissue the command. If the problem persists, contact the IBM Support Center.

6027-925 Directory cannot be assigned to storage pool '*poolName***'.**

Explanation: The file cannot be assigned to the specified pool.

User response: Determine the correct storage pool name and reissue the command.

6027-926 Symbolic link cannot be assigned to storage pool '*poolName***'.**

Explanation: The file cannot be assigned to the specified pool.

User response: Determine the correct storage pool name and reissue the command.

6027-927 System file cannot be assigned to storage pool '*poolName***'.**

Explanation: The file cannot be assigned to the specified pool.

User response: Determine the correct storage pool name and reissue the command.

6027-928 [E] Error: File system or device *fileSystem* **has no global snapshot with name** *snapshotName*.

Explanation: The specified file system does not have a global snapshot with the specified snapshot name.

User response: Use the **mmlssnapshot** command to list the snapshot names for the file system. Alternatively, specify the full pathname of the desired snapshot directory instead of using the **-S** option.

6027-929 [W] Attention: In RULE '*ruleName***' (***ruleNumber***), both pools '***poolName***' and '***poolName***' are EXTERNAL. This is not a supported migration.**

Explanation: The command does not support migration between two **EXTERNAL** pools.

User response: Correct the rule and reissue the command.

Note: This is not an unrecoverable error. The command will continue to run.

6027-930 [W] Attention: In RULE '*ruleName***' LIST name '***listName***' appears, but there is no corresponding EXTERNAL LIST '***listName***' EXEC ... OPTS ... rule to specify a program to process the matching files.**

Explanation: There should be an **EXTERNAL LIST** rule for every list named by your **LIST** rules.

User response: Add an "**EXTERNAL LIST** *listName* **EXEC** *scriptName* **OPTS** *opts*" rule.

Note: This is not an unrecoverable error. For execution with **-I defer**, file lists are generated and saved, so **EXTERNAL LIST** rules are not strictly necessary for correct execution.

6027-931 [E] Error - The policy evaluation phase did not complete.

Explanation: One or more errors prevented the policy evaluation phase from examining all of the files.

User response: Consider other messages emitted by the command. Take appropriate action and then reissue the command.

6027-932 [E] Error - The policy execution phase did not complete.

Explanation: One or more errors prevented the policy execution phase from operating on each chosen file.

User response: Consider other messages emitted by the command. Take appropriate action and then reissue the command.

6027-933 [W] EXEC '*wouldbeScriptPathname***' of EXTERNAL POOL or LIST '***PoolOrListName***' fails TEST with code** *scriptReturnCode* **on this node.**

Explanation: Each EXEC defined in an **EXTERNAL POOL** or **LIST** rule is run in **TEST** mode on each node. Each invocation that fails with a nonzero return code is reported. Command execution is terminated on any node that fails any of these tests.

User response: Correct the **EXTERNAL POOL** or **LIST** rule, the EXEC script, or do nothing because this is not necessarily an error. The administrator may suppress execution of the **mmapplypolicy** command on some nodes by deliberately having one or more EXECs return nonzero codes.

6027-934 [W] Attention: Specified snapshot: '*SnapshotName***' will be ignored because the path specified: '***PathName***' is not within that snapshot.**

Explanation: The command line specified both a path name to be scanned and a snapshot name, but the snapshot name was not consistent with the path name.

User response: If you wanted the entire snapshot, just specify the GPFS file system name or device name. If you wanted a directory within a snapshot, specify a path name within that snapshot (for example, **/gpfs/FileSystemName/.snapshots/SnapShotName/Directory**).

6027-935 [W] Attention: In RULE '*ruleName***' (***ruleNumber***) LIMIT or REPLICATE clauses are ignored; not supported for migration to EXTERNAL pool '***storagePoolName***'.**

Explanation: GPFS does not support the **LIMIT** or **REPLICATE** clauses during migration to external pools.

User response: Correct the policy rule to avoid this warning message.

6027-936 [W] Error - command master is an incompatible version.

Explanation: While operating in master/client mode, the command discovered that the master is running an incompatible version.

User response: Upgrade the command software on all nodes and reissue the command.

6027-937 [E] Error creating shared temporary sub-directory *subDirName***:** *subDirPath*

Explanation: The **mkdir** command failed on the named subdirectory path.

User response: Specify an existing writable shared directory as the shared temporary directory argument to the policy command. The policy command will create a subdirectory within that.

6027-938 [E] Error closing work file *fileName***:** *errorString*

Explanation: An error occurred while attempting to close the named work file or socket.

User response: Record the above information. Contact the IBM Support Center.

6027-939 [E] Error on gpfs_quotactl(*pathName***,***commandCode***,** *resourceId***):** *errorString*

Explanation: An error occurred while attempting **gpfs_quotactl()**.

User response: Correct the policy rules and/or enable GPFS quota tracking. If problem persists contact the IBM Support Center.

6027-940 Open failed.

Explanation: The **open()** system call was not successful.

User response: Check additional error messages.

6027-941 Set replication failed.

Explanation: The **open()** system call was not successful.

User response: Check additional error messages.

6027-943 -M and -R are only valid for zero length files.

Explanation: The **mmchattr** command received command line arguments that were not valid.

User response: Correct command line and reissue the command.

6027-944 -m *value* **exceeds number of failure groups for metadata.**

Explanation: The **mmchattr** command received command line arguments that were not valid.

User response: Correct command line and reissue the command.

6027-945 -r *value* **exceeds number of failure groups for data.**

Explanation: The **mmchattr** command received command line arguments that were not valid.

User response: Correct command line and reissue the command.

6027-946 Not a regular file or directory.

Explanation: An **mmlsattr** or **mmchattr** command error occurred.

User response: Correct the problem and reissue the command.

6027-947 **Stat failed: A file or directory in the path name does not exist.**

Explanation: A file or directory in the path name does not exist.

User response: Correct the problem and reissue the command.

6027-948 [E:nnn] *fileName*: **get clone attributes failed:** *errorString*

Explanation: The **tsfattr** call failed.

User response: Check for additional error messages. Resolve the problems before reattempting the failing operation.

6027-949 [E] *fileName*: **invalid clone attributes.**

Explanation: Self explanatory.

User response: Check for additional error messages. Resolve the problems before reattempting the failing operation.

6027-950 [E:nnn] **File cloning requires the 'fastea' feature to be enabled.**

Explanation: The file system **fastea** feature is not enabled.

User response: Enable the **fastea** feature by issuing the **mmchfs -V** and **mmmigratefs --fastea** commands.

6027-951 [E] **Error on** *operationName* **to work file** *fileName*: *errorString*

Explanation: An error occurred while attempting to do a (write-like) operation on the named work file.

User response: Investigate the file and possibly reissue the command. Check that there is sufficient free space in the file system.

6027-953 **Failed to get a handle for fileset** *filesetName*, **snapshot** *snapshotName* **in file system** *fileSystem*. *errorMessage*.

Explanation: Failed to get a handle for a specific fileset snapshot in the file system.

User response: Correct the command line and reissue the command. If the problem persists, contact the IBM Support Center.

6027-954 **Failed to get the maximum inode number in the active file system.** *errorMessage*.

Explanation: Failed to get the maximum inode number in the current active file system.

User response: Correct the command line and reissue the command. If the problem persists, contact the IBM Support Center.

6027-955 **Failed to set the maximum allowed memory for the specified** *fileSystem* **command.**

Explanation: Failed to set the maximum allowed memory for the specified command.

User response: Correct the command line and reissue the command. If the problem persists, contact the IBM Support Center.

6027-956 **Cannot allocate enough buffer to record different items.**

Explanation: Cannot allocate enough buffer to record different items which are used in the next phase.

User response: Correct the command line and reissue the command. If the problem persists, contact the system administrator.

6027-957 **Failed to get the root directory inode of fileset** *filesetName*

Explanation: Failed to get the root directory inode of a fileset.

User response: Correct the command line and reissue the command. If the problem persists, contact the IBM Support Center.

6027-959 '*fileName*' **is not a regular file.**

Explanation: Only regular files are allowed to be clone parents.

User response: This file is not a valid target for **mmclone** operations.

6027-960 **cannot access '***fileName***':** *errorString*.

Explanation: This message provides more details about a **stat()** error.

User response: Correct the problem and reissue the command.

6027-961 **Cannot execute** *command*.

Explanation: The **mmeditacl** command cannot invoke the **mmgetacl** or **mmputacl** command.

User response: Contact your system administrator.

6027-962 **Failed to list fileset** *filesetName*.

Explanation: Failed to list specific fileset.

User response: None.

6027-963 EDITOR environment variable not set

Explanation: Self-explanatory.

User response: Set the EDITOR environment variable and reissue the command.

6027-964 EDITOR environment variable must be an absolute path name

Explanation: Self-explanatory.

User response: Set the EDITOR environment variable correctly and reissue the command.

6027-965 Cannot create temporary file

Explanation: Self-explanatory.

User response: Contact your system administrator.

6027-966 Cannot access *fileName*

Explanation: Self-explanatory.

User response: Verify file permissions.

6027-967 Should the modified ACL be applied? (yes) or (no)

Explanation: Self-explanatory.

User response: Respond **yes** if you want to commit the changes, **no** otherwise.

6027-971 Cannot find *fileName*

Explanation: Self-explanatory.

User response: Verify the file name and permissions.

6027-972 *name* **is not a directory (-d not valid).**

Explanation: Self-explanatory.

User response: None, only directories are allowed to have default ACLs.

6027-973 Cannot allocate *number* **byte buffer for ACL.**

Explanation: There was not enough available memory to process the request.

User response: Contact your system administrator.

6027-974 Failure reading ACL (rc=*number***).**

Explanation: An unexpected error was encountered by **mmgetacl** or **mmeditacl**.

User response: Examine the return code, contact the IBM Support Center if necessary.

6027-976 Failure writing ACL (rc=*number***).**

Explanation: An unexpected error encountered by **mmputacl** or **mmeditacl**.

User response: Examine the return code, Contact the IBM Support Center if necessary.

6027-977 Authorization failure

Explanation: An attempt was made to create or modify the ACL for a file that you do not own.

User response: Only the owner of a file or the root user can create or change the access control list for a file.

6027-978 Incorrect, duplicate, or missing access control entry detected.

Explanation: An access control entry in the ACL that was created had incorrect syntax, one of the required access control entries is missing, or the ACL contains duplicate access control entries.

User response: Correct the problem and reissue the command.

6027-979 Incorrect ACL entry: *entry***.**

Explanation: Self-explanatory.

User response: Correct the problem and reissue the command.

6027-980 *name* **is not a valid user name.**

Explanation: Self-explanatory.

User response: Specify a valid user name and reissue the command.

6027-981 *name* **is not a valid group name.**

Explanation: Self-explanatory.

User response: Specify a valid group name and reissue the command.

6027-982 *name* **is not a valid ACL entry type.**

Explanation: Specify a valid ACL entry type and reissue the command.

User response: Correct the problem and reissue the command.

6027-983 *name* **is not a valid permission set.**

Explanation: Specify a valid permission set and reissue the command.

User response: Correct the problem and reissue the command.

6027-985 **An error was encountered while deleting the ACL (rc=*value*).**

Explanation: An unexpected error was encountered by **tsdelacl**.

User response: Examine the return code and contact the IBM Support Center, if necessary.

6027-986 **Cannot open *fileName*.**

Explanation: Self-explanatory.

User response: Verify the file name and permissions.

6027-987 ***name* is not a valid special name.**

Explanation: Produced by the **mmputacl** command when the NFS V4 'special' identifier is followed by an unknown special id string. *name* is one of the following: 'owner@', 'group@', 'everyone@'.

User response: Specify a valid NFS V4 special name and reissue the command.

6027-988 ***type* is not a valid NFS V4 type.**

Explanation: Produced by the **mmputacl** command when the type field in an ACL entry is not one of the supported NFS Version 4 type values. *type* is one of the following: 'allow' or 'deny'.

User response: Specify a valid NFS V4 type and reissue the command.

6027-989 ***name* is not a valid NFS V4 flag.**

Explanation: A flag specified in an ACL entry is not one of the supported values, or is not valid for the type of object (inherit flags are valid for directories only). Valid values are **FileInherit**, **DirInherit**, and **InheritOnly**.

User response: Specify a valid NFS V4 option and reissue the command.

6027-990 **Missing permissions (*value* found, *value* are required).**

Explanation: The permissions listed are less than the number required.

User response: Add the missing permissions and reissue the command.

6027-991 **Combining FileInherit and DirInherit makes the mask ambiguous.**

Explanation: Produced by the **mmputacl** command when WRITE/CREATE is specified without **MKDIR** (or the other way around), and both the **FILE_INHERIT** and **DIR_INHERIT** flags are specified.

User response: Make separate **FileInherit** and **DirInherit** entries and reissue the command.

6027-992 **Subdirectory *name* already exists. Unable to create snapshot.**

Explanation: **tsbackup** was unable to create a snapshot because the snapshot subdirectory already exists. This condition sometimes is caused by issuing a IBM Spectrum Protect restore operation without specifying a different subdirectory as the target of the restore.

User response: Remove or rename the existing subdirectory and then retry the command.

6027-993 **Keyword *aclType* is incorrect. Valid values are: 'posix', 'nfs4', 'native'.**

Explanation: One of the **mm*acl** commands specified an incorrect value with the **-k** option.

User response: Correct the *aclType* value and reissue the command.

6027-994 **ACL permissions cannot be denied to the file owner.**

Explanation: The **mmputacl** command found that the **READ_ACL**, **WRITE_ACL**, **READ_ATTR**, or **WRITE_ATTR** permissions are explicitly being denied to the file owner. This is not permitted, in order to prevent the file being left with an ACL that cannot be modified.

User response: Do not select the **READ_ACL**, **WRITE_ACL**, **READ_ATTR**, or **WRITE_ATTR** permissions on **deny** ACL entries for the **OWNER**.

6027-995 **This command will run on a remote node, *nodeName*.**

Explanation: The **mmputacl** command was invoked for a file that resides on a file system in a remote cluster, and UID remapping is enabled. To parse the user and group names from the ACL file correctly, the command will be run transparently on a node in the remote cluster.

User response: None. Informational message only.

6027-996 [E:nnn] Error reading policy text from: *fileName*

Explanation: An error occurred while attempting to open or read the specified policy file. The policy file may be missing or inaccessible.

User response: Read all of the related error messages and try to correct the problem.

6027-997 [W] **Attention: RULE '***ruleName***' attempts to redefine EXTERNAL** *POOLorLISTliteral* **'***poolName***', ignored.**

Explanation: Execution continues as if the specified rule was not present.

User response: Correct or remove the policy rule.

6027-998 [E] **Error in FLR/PDR serving for client** *clientHostNameAndPortNumber***: FLRs=***numOfFileListRecords* **PDRs=***numOfPolicyDecisionResponses* **pdrs=***numOfPolicyDecisionResponseRecords*

Explanation: A protocol error has been detected among cooperating **mmapplypolicy** processes.

User response: Reissue the command. If the problem persists, contact the IBM Support Center.

6027-999 [E] **Authentication failed:** *myNumericNetworkAddress* **with** *partnersNumericNetworkAddress* **(code=***codeIndicatingProtocolStepSequence* **rc=***errnoStyleErrorCode***)**

Explanation: Two processes at the specified network addresses failed to authenticate. The cooperating processes should be on the same network; they should not be separated by a firewall.

User response: Correct the configuration and try the operation again. If the problem persists, contact the IBM Support Center.

6027-1004 **Incorrect [***nodelist***] format in file:** *nodeListLine*

Explanation: A [*nodelist*] line in the input stream is not a comma-separated list of nodes.

User response: Fix the format of the [*nodelist*] line in the **mmfs.cfg** input file. This is usually the *NodeFile* specified on the **mmchconfig** command.

If no user-specified [*nodelist*] lines are in error, contact the IBM Support Center.

If user-specified [*nodelist*] lines are in error, correct these lines.

6027-1005 **Common is not sole item on [] line** *number***.**

Explanation: A [*nodelist*] line in the input stream contains common plus any other names.

User response: Fix the format of the [*nodelist*] line in the **mmfs.cfg** input file. This is usually the *NodeFile* specified on the **mmchconfig** command.

If no user-specified [*nodelist*] lines are in error, contact the IBM Support Center.

If user-specified [*nodelist*] lines are in error, correct these lines.

6027-1006 **Incorrect custom [ ] line** *number***.**

Explanation: A [*nodelist*] line in the input stream is not of the format: [*nodelist*]. This covers syntax errors not covered by messages 6027-1004 and 6027-1005.

User response: Fix the format of the list of nodes in the **mmfs.cfg** input file. This is usually the *NodeFile* specified on the **mmchconfig** command.

If no user-specified lines are in error, contact the IBM Support Center.

If user-specified lines are in error, correct these lines.

6027-1007 **attribute found in common multiple times:** *attribute***.**

Explanation: The attribute specified on the command line is in the main input stream multiple times. This is occasionally legal, such as with the trace attribute. These attributes, however, are not meant to be repaired by **mmfixcfg**.

User response: Fix the configuration file (**mmfs.cfg** or **mmfscfg1** in the SDR). All attributes modified by GPFS configuration commands may appear only once in common sections of the configuration file.

6027-1008 **Attribute found in custom multiple times:** *attribute***.**

Explanation: The attribute specified on the command line is in a custom section multiple times. This is occasionally legal. These attributes are not meant to be repaired by **mmfixcfg**.

User response: Fix the configuration file (**mmfs.cfg** or **mmfscfg1** in the SDR). All attributes modified by GPFS configuration commands may appear only once in custom sections of the configuration file.

6027-1022 **Missing mandatory arguments on command line.**

Explanation: Some, but not enough, arguments were specified to the **mmcrfsc** command.

User response: Specify all arguments as per the usage statement that follows.

6027-1023 **invalid maxBlockSize parameter:** *value*

Explanation: The first argument to the **mmcrfsc** command is maximum block size and should be greater than 0.

User response: The maximum block size should be greater than 0. The **mmcrfs** command should never call the **mmcrfsc** command without a valid maximum block size argument. Contact the IBM Support Center.

6027-1028 Incorrect value for *-name* flag.

Explanation: An incorrect argument was specified with an option that requires one of a limited number of allowable options (for example, **-s** or any of the **yes | no** options).

User response: Use one of the valid values for the specified option.

6027-1029 Incorrect characters in integer field for *-name* option.

Explanation: An incorrect character was specified with the indicated option.

User response: Use a valid integer for the indicated option.

6027-1030 Value below minimum for *-optionLetter* option. Valid range is from *value* to *value*

Explanation: The value specified with an option was below the minimum.

User response: Use an integer in the valid range for the indicated option.

6027-1031 Value above maximum for option *-optionLetter*. Valid range is from *value* to *value*.

Explanation: The value specified with an option was above the maximum.

User response: Use an integer in the valid range for the indicated option.

6027-1032 Incorrect option *optionName*.

Explanation: An unknown option was specified.

User response: Use only the options shown in the syntax.

6027-1033 Option *optionName* specified twice.

Explanation: An option was specified more than once on the command line.

User response: Use options only once.

6027-1034 Missing argument after *optionName* option.

Explanation: An option was not followed by an argument.

User response: All options need an argument. Specify one.

6027-1035 Option *-optionName* is mandatory.

Explanation: A mandatory input option was not specified.

User response: Specify all mandatory options.

6027-1036 Option expected at *string*.

Explanation: Something other than an expected option was encountered on the latter portion of the command line.

User response: Follow the syntax shown. Options may not have multiple values. Extra arguments are not allowed.

6027-1038 IndirectSize must be <= BlockSize and must be a multiple of LogicalSectorSize (512).

Explanation: The IndirectSize specified was not a multiple of 512 or the IndirectSize specified was larger than BlockSize.

User response: Use valid values for IndirectSize and BlockSize.

6027-1039 InodeSize must be a multiple of LocalSectorSize (512).

Explanation: The specified InodeSize was not a multiple of 512.

User response: Use a valid value for InodeSize.

6027-1040 InodeSize must be less than or equal to Blocksize.

Explanation: The specified InodeSize was not less than or equal to Blocksize.

User response: Use a valid value for InodeSize.

6027-1042 DefaultMetadataReplicas must be less than or equal to MaxMetadataReplicas.

Explanation: The specified DefaultMetadataReplicas was greater than MaxMetadataReplicas.

User response: Specify a valid value for DefaultMetadataReplicas.

6027-1043 DefaultDataReplicas must be less than or equal MaxDataReplicas.

Explanation: The specified DefaultDataReplicas was greater than MaxDataReplicas.

User response: Specify a valid value for DefaultDataReplicas.

6027-1055 LogicalSectorSize must be a multiple of 512

Explanation: The specified *LogicalSectorSize* was not a multiple of 512.

User response: Specify a valid *LogicalSectorSize*.

6027-1056 *Blocksize* must be a multiple of *LogicalSectorSize* × 32

Explanation: The specified *Blocksize* was not a multiple of *LogicalSectorSize* × 32.

User response: Specify a valid value for *Blocksize*.

6027-1057 InodeSize must be less than or equal to Blocksize.

Explanation: The specified *InodeSize* was not less than or equal to *Blocksize*.

User response: Specify a valid value for *InodeSize*.

6027-1059 Mode must be M or S: *mode*

Explanation: The first argument provided in the **mmcrfsc** command was not M or S.

User response: The **mmcrfsc** command should not be called by a user. If any other command produces this error, contact the IBM Support Center.

6027-1084 The specified block size (*value*K) exceeds the maximum allowed block size currently in effect (*value*K). Either specify a smaller value for the -B parameter, or increase the maximum block size by issuing: mmchconfig maxblocksize=*value*K and restart the GPFS daemon.

Explanation: The specified value for block size was greater than the value of the **maxblocksize** configuration parameter.

User response: Specify a valid value or increase the value of the allowed block size by specifying a larger value on the **maxblocksize** parameter of the **mmchconfig** command.

6027-1113 Incorrect option: *option*.

Explanation: The specified command option is not valid.

User response: Specify a valid option and reissue the command.

6027-1119 Obsolete option: *option*.

Explanation: A command received an option that is not valid any more.

User response: Correct the command line and reissue the command.

6027-1120 Interrupt received: No changes made.

Explanation: A GPFS administration command (**mm...**) received an interrupt before committing any changes.

User response: None. Informational message only.

6027-1123 Disk name must be specified in disk descriptor.

Explanation: The disk name positional parameter (the first field) in a disk descriptor was empty. The bad disk descriptor is displayed following this message.

User response: Correct the input and rerun the command.

6027-1124 Disk usage must be dataOnly, metadataOnly, descOnly, or dataAndMetadata.

Explanation: The disk usage parameter has a value that is not valid.

User response: Correct the input and reissue the command.

6027-1132 Interrupt received: changes not propagated.

Explanation: An interrupt was received after changes were committed but before the changes could be propagated to all the nodes.

User response: All changes will eventually propagate as nodes recycle or other GPFS administration commands are issued. Changes can be activated now by manually restarting the GPFS daemons.

6027-1133 Interrupt received. Only a subset of the parameters were changed.

Explanation: An interrupt was received in **mmchfs** before all of the requested changes could be completed.

User response: Use **mmlsfs** to see what the currently active settings are. Reissue the command if you want to change additional parameters.

6027-1135 Restriping may not have finished.

Explanation: An interrupt occurred during restriping.

User response: Restart the restripe. Verify that the file system was not damaged by running the **mmfsck** command.

6027-1136 *option* **option specified twice.**

Explanation: An option was specified multiple times on a command line.

User response: Correct the error on the command line and reissue the command.

6027-1137 *option* **value must be yes or no.**

Explanation: A yes or no option was used with something other than **yes** or **no**.

User response: Correct the error on the command line and reissue the command.

6027-1138 **Incorrect extra argument:** *argument*

Explanation: Non-option arguments followed the mandatory arguments.

User response: Unlike most POSIX commands, the main arguments come first, followed by the optional arguments. Correct the error and reissue the command.

6027-1140 **Incorrect integer for** *option***:** *number***.**

Explanation: An option requiring an integer argument was followed by something that cannot be parsed as an integer.

User response: Specify an integer with the indicated option.

6027-1141 **No disk descriptor file specified.**

Explanation: An **-F** flag was not followed by the path name of a disk descriptor file.

User response: Specify a valid disk descriptor file.

6027-1142 **File** *fileName* **already exists.**

Explanation: The specified file already exists.

User response: Rename the file or specify a different file name and reissue the command.

6027-1143 **Cannot open** *fileName***.**

Explanation: A file could not be opened.

User response: Verify that the specified file exists and that you have the proper authorizations.

6027-1144 **Incompatible cluster types. You cannot move file systems that were created by GPFS cluster type** *sourceCluster* **into GPFS cluster type** *targetCluster***.**

Explanation: The source and target cluster types are incompatible.

User response: Contact the IBM Support Center for assistance.

6027-1145 *parameter* **must be greater than 0:** *value*

Explanation: A negative value had been specified for the named parameter, which requires a positive value.

User response: Correct the input and reissue the command.

6027-1147 **Error converting** *diskName* **into an NSD.**

Explanation: Error encountered while converting a disk into an NSD.

User response: Check the preceding messages for more information.

6027-1148 **File system** *fileSystem* **already exists in the cluster. Use mmchfs -W to assign a new device name for the existing file system.**

Explanation: You are trying to import a file system into the cluster but there is already a file system with the same name in the cluster.

User response: Remove or rename the file system with the conflicting name.

6027-1149 *fileSystem* **is defined to have mount point** *mountpoint***. There is already such a mount point in the cluster. Use mmchfs -T to assign a new mount point to the existing file system.**

Explanation: The cluster into which the file system is being imported already contains a file system with the same mount point as the mount point of the file system being imported.

User response: Use the **-T** option of the **mmchfs** command to change the mount point of the file system that is already in the cluster and then rerun the **mmimportfs** command.

6027-1150 **Error encountered while importing disk** *diskName***.**

Explanation: The **mmimportfs** command encountered problems while processing the disk.

User response: Check the preceding messages for more information.

6027-1151 **Disk** *diskName* **already exists in the cluster.**

Explanation: You are trying to import a file system that has a disk with the same name as some disk from a file system that is already in the cluster.

User response: Remove or replace the disk with the conflicting name.

6027-1152 Block size must be 64K, 128K, 256K, 512K, 1M, 2M, 4M, 8M or 16M.

Explanation: The specified block size value is not valid.

User response: Specify a valid block size value.

6027-1153 At least one node in the cluster must be defined as a quorum node.

Explanation: All nodes were explicitly designated or allowed to default to be nonquorum.

User response: Specify which of the nodes should be considered quorum nodes and reissue the command.

6027-1154 Incorrect node *node* specified for command.

Explanation: The user specified a node that is not valid.

User response: Specify a valid node.

6027-1155 The NSD servers for the following disks from file system *fileSystem* were reset or not defined: *diskList*

Explanation: Either the **mmimportfs** command encountered disks with no NSD servers, or was forced to reset the NSD server information for one or more disks.

User response: After the **mmimportfs** command finishes, use the **mmchnsd** command to assign NSD server nodes to the disks as needed.

6027-1156 The NSD servers for the following free disks were reset or not defined: *diskList*

Explanation: Either the **mmimportfs** command encountered disks with no NSD servers, or was forced to reset the NSD server information for one or more disks.

User response: After the **mmimportfs** command finishes, use the **mmchnsd** command to assign NSD server nodes to the disks as needed.

6027-1157 Use the mmchnsd command to assign NSD servers as needed.

Explanation: Either the **mmimportfs** command encountered disks with no NSD servers, or was forced to reset the NSD server information for one or more disks. Check the preceding messages for detailed information.

User response: After the **mmimportfs** command finishes, use the **mmchnsd** command to assign NSD server nodes to the disks as needed.

6027-1159 The following file systems were not imported: *fileSystemList*

Explanation: The **mmimportfs** command was not able to import the specified file systems. Check the preceding messages for error information.

User response: Correct the problems and reissue the **mmimportfs** command.

6027-1160 The drive letters for the following file systems have been reset: *fileSystemList*.

Explanation: The drive letters associated with the specified file systems are already in use by existing file systems and have been reset.

User response: After the **mmimportfs** command finishes, use the **-t** option of the **mmchfs** command to assign new drive letters as needed.

6027-1161 Use the dash character (-) to separate multiple node designations.

Explanation: A command detected an incorrect character used as a separator in a list of node designations.

User response: Correct the command line and reissue the command.

6027-1162 Use the semicolon character (;) to separate the disk names.

Explanation: A command detected an incorrect character used as a separator in a list of disk names.

User response: Correct the command line and reissue the command.

6027-1163 GPFS is still active on *nodeName*.

Explanation: The GPFS daemon was discovered to be active on the specified node during an operation that requires the daemon to be stopped.

User response: Stop the daemon on the specified node and rerun the command.

6027-1164 Use mmchfs -t to assign drive letters as needed.

Explanation: The **mmimportfs** command was forced to reset the drive letters associated with one or more file systems. Check the preceding messages for detailed information.

User response: After the **mmimportfs** command finishes, use the **-t** option of the **mmchfs** command to assign new drive letters as needed.

6027-1165 **The PR attributes for the following disks from file system** *fileSystem* **were reset or not yet established:** *diskList*

Explanation: The **mmimportfs** command disabled the Persistent Reserve attribute for one or more disks.

User response: After the **mmimportfs** command finishes, use the **mmchconfig** command to enable Persistent Reserve in the cluster as needed.

6027-1166 **The PR attributes for the following free disks were reset or not yet established:** *diskList*

Explanation: The **mmimportfs** command disabled the Persistent Reserve attribute for one or more disks.

User response: After the **mmimportfs** command finishes, use the **mmchconfig** command to enable Persistent Reserve in the cluster as needed.

6027-1167 **Use mmchconfig to enable Persistent Reserve in the cluster as needed.**

Explanation: The **mmimportfs** command disabled the Persistent Reserve attribute for one or more disks.

User response: After the **mmimportfs** command finishes, use the **mmchconfig** command to enable Persistent Reserve in the cluster as needed.

6027-1168 **Inode size must be 512, 1K or 4K.**

Explanation: The specified inode size is not valid.

User response: Specify a valid inode size.

6027-1169 *attribute* **must be** *value***.**

Explanation: The specified value of the given attribute is not valid.

User response: Specify a valid value.

6027-1178 *parameter* **must be from** *value* **to** *value***:** *valueSpecified*

Explanation: A parameter value specified was out of range.

User response: Keep the specified value within the range shown.

6027-1188 **Duplicate disk specified:** *disk*

Explanation: A disk was specified more than once on the command line.

User response: Specify each disk only once.

6027-1189 **You cannot delete all the disks.**

Explanation: The number of disks to delete is greater than or equal to the number of disks in the file system.

User response: Delete only some of the disks. If you want to delete them all, use the **mmdelfs** command.

6027-1197 *parameter* **must be greater than** *value***:** *value***.**

Explanation: An incorrect value was specified for the named parameter.

User response: Correct the input and reissue the command.

6027-1200 **tscrfs failed. Cannot create** *device*

Explanation: The internal **tscrfs** command failed.

User response: Check the error message from the command that failed.

6027-1201 **Disk** *diskName* **does not belong to file system** *fileSystem***.**

Explanation: The specified disk was not found to be part of the cited file system.

User response: If the disk and file system were specified as part of a GPFS command, reissue the command with a disk that belongs to the specified file system.

6027-1202 **Active disks are missing from the GPFS configuration data.**

Explanation: A GPFS disk command found that one or more active disks known to the GPFS daemon are not recorded in the GPFS configuration data. A list of the missing disks follows.

User response: Contact the IBM Support Center.

6027-1203 **Attention: File system** *fileSystem* **may have some disks that are in a non-ready state. Issue the command: mmcommon recoverfs** *fileSystem*

Explanation: The specified file system may have some disks that are in a non-ready state.

User response: Run **mmcommon recoverfs** *fileSystem* to ensure that the GPFS configuration data for the file system is current, and then display the states of the disks in the file system using the **mmlsdisk** command.

If any disks are in a non-ready state, steps should be taken to bring these disks into the ready state, or to remove them from the file system. This can be done by mounting the file system, or by using the **mmchdisk** command for a mounted or unmounted file system. When maintenance is complete or the failure has been

repaired, use the **mmchdisk** command with the **start** option. If the failure cannot be repaired without loss of data, you can use the **mmdeldisk** command to delete the disks.

6027-1204 *command* **failed.**

Explanation: An internal command failed. This is usually a call to the GPFS daemon.

User response: Check the error message from the command that failed.

6027-1205 **Failed to connect to remote cluster** *clusterName***.**

Explanation: Attempt to establish a connection to the specified cluster was not successful. This can be caused by a number of reasons: GPFS is down on all of the contact nodes, the contact node list is obsolete, the owner of the remote cluster revoked authorization, and so forth.

User response: If the error persists, contact the administrator of the remote cluster and verify that the contact node information is current and that the authorization key files are current as well.

6027-1206 **File system** *fileSystem* **belongs to cluster** *clusterName***. Command is not allowed for remote file systems.**

Explanation: The specified file system is not local to the cluster, but belongs to the cited remote cluster.

User response: Choose a local file system, or issue the command on a node in the remote cluster.

6027-1207 **There is already an existing file system using** *value***.**

Explanation: The mount point or device name specified matches that of an existing file system. The device name and mount point must be unique within a GPFS cluster.

User response: Choose an unused name or path.

6027-1208 **File system** *fileSystem* **not found in cluster** *clusterName***.**

Explanation: The specified file system does not belong to the cited remote cluster. The local information about the file system is not current. The file system may have been deleted, renamed, or moved to a different cluster.

User response: Contact the administrator of the remote cluster that owns the file system and verify the accuracy of the local information. Use the **mmremotefs show** command to display the local information about the file system. Use the **mmremotefs update** command to make the necessary changes.

6027-1209 **GPFS is down on this node.**

Explanation: GPFS is not running on this node.

User response: Ensure that GPFS is running and reissue the command.

6027-1210 **GPFS is not ready to handle commands yet.**

Explanation: GPFS is in the process of initializing or waiting for quorum to be reached.

User response: Reissue the command.

6027-1211 *fileSystem* **refers to file system** *fileSystem* **in cluster** *clusterName***.**

Explanation: Informational message.

User response: None.

6027-1212 **File system** *fileSystem* **does not belong to cluster** *clusterName***.**

Explanation: The specified file system refers to a file system that is remote to the cited cluster. Indirect remote file system access is not allowed.

User response: Contact the administrator of the remote cluster that owns the file system and verify the accuracy of the local information. Use the **mmremotefs show** command to display the local information about the file system. Use the **mmremotefs update** command to make the necessary changes.

6027-1213 *command* **failed. Error code** *errorCode***.**

Explanation: An internal command failed. This is usually a call to the GPFS daemon.

User response: Examine the error code and other messages to determine the reason for the failure. Correct the problem and reissue the command.

6027-1214 **Unable to enable Persistent Reserve on the following disks:** *diskList*

Explanation: The command was unable to set up all of the disks to use Persistent Reserve.

User response: Examine the disks and the additional error information to determine if the disks should have supported Persistent Reserve. Correct the problem and reissue the command.

6027-1215 **Unable to reset the Persistent Reserve attributes on one or more disks on the following nodes:** *nodeList*

Explanation: The command could not reset Persistent Reserve on at least one disk on the specified nodes.

User response: Examine the additional error

information to determine whether nodes were down or if there was a disk error. Correct the problems and reissue the command.

6027-1216 **File** *fileName* **contains additional error information.**

Explanation: The command generated a file containing additional error information.

User response: Examine the additional error information.

6027-1217 **A disk descriptor contains an incorrect separator character.**

Explanation: A command detected an incorrect character used as a separator in a disk descriptor.

User response: Correct the disk descriptor and reissue the command.

6027-1218 **Node** *nodeName* **does not have a GPFS server license designation.**

Explanation: The function that you are assigning to the node requires the node to have a GPFS server license.

User response: Use the **mmchlicense** command to assign a valid GPFS license to the node or specify a different node.

6027-1219 **NSD discovery on node** *nodeName* **failed with return code** *value***.**

Explanation: The NSD discovery process on the specified node failed with the specified return code.

User response: Determine why the node cannot access the specified NSDs. Correct the problem and reissue the command.

6027-1220 **Node** *nodeName* **cannot be used as an NSD server for Persistent Reserve disk** *diskName* **because it is not an AIX node.**

Explanation: The node shown was specified as an NSD server for *diskName*, but the node does not support Persistent Reserve.

User response: Specify a node that supports Persistent Reserve as an NSD server.

6027-1221 **The number of NSD servers exceeds the maximum (***value***) allowed.**

Explanation: The number of NSD servers in the disk descriptor exceeds the maximum allowed.

User response: Change the disk descriptor to specify no more NSD servers than the maximum allowed.

6027-1222 **Cannot assign a minor number for file system** *fileSystem* **(major number** *deviceMajorNumber***).**

Explanation: The command was not able to allocate a minor number for the new file system.

User response: Delete unneeded **/dev** entries for the specified major number and reissue the command.

6027-1223 *ipAddress* **cannot be used for NFS serving; it is used by the GPFS daemon.**

Explanation: The IP address shown has been specified for use by the GPFS daemon. The same IP address cannot be used for NFS serving because it cannot be failed over.

User response: Specify a different IP address for NFS use and reissue the command.

6027-1224 **There is no file system with drive letter** *driveLetter***.**

Explanation: No file system in the GPFS cluster has the specified drive letter.

User response: Reissue the command with a valid file system.

6027-1225 **Explicit drive letters are supported only in a Windows environment. Specify a mount point or allow the default settings to take effect.**

Explanation: An explicit drive letter was specified on the **mmmount** command but the target node does not run the Windows operating system.

User response: Specify a mount point or allow the default settings for the file system to take effect.

6027-1226 **Explicit mount points are not supported in a Windows environment. Specify a drive letter or allow the default settings to take effect.**

Explanation: An explicit mount point was specified on the **mmmount** command but the target node runs the Windows operating system.

User response: Specify a drive letter or allow the default settings for the file system to take effect.

6027-1227 **The main GPFS cluster configuration file is locked. Retrying ...**

Explanation: Another GPFS administration command has locked the cluster configuration file. The current process will try to obtain the lock a few times before giving up.

User response: None. Informational message only.

6027-1228 Lock creation successful.

Explanation: The holder of the lock has released it and the current process was able to obtain it.

User response: None. Informational message only. The command will now continue.

6027-1229 Timed out waiting for lock. Try again later.

Explanation: Another GPFS administration command kept the main GPFS cluster configuration file locked for over a minute.

User response: Try again later. If no other GPFS administration command is presently running, see "GPFS cluster configuration data file issues" on page 298.

6027-1230 *diskName* **is a tiebreaker disk and cannot be deleted.**

Explanation: A request was made to GPFS to delete a node quorum tiebreaker disk.

User response: Specify a different disk for deletion.

6027-1231 GPFS detected more than eight quorum nodes while node quorum with tiebreaker disks is in use.

Explanation: A GPFS command detected more than eight quorum nodes, but this is not allowed while node quorum with tiebreaker disks is in use.

User response: Reduce the number of quorum nodes to a maximum of eight, or use the normal node quorum algorithm.

6027-1232 GPFS failed to initialize the tiebreaker disks.

Explanation: A GPFS command unsuccessfully attempted to initialize the node quorum tiebreaker disks.

User response: Examine prior messages to determine why GPFS was unable to initialize the tiebreaker disks and correct the problem. After that, reissue the command.

6027-1233 Incorrect keyword: *value*.

Explanation: A command received a keyword that is not valid.

User response: Correct the command line and reissue the command.

6027-1234 Adding node *node* **to the cluster will exceed the quorum node limit.**

Explanation: An attempt to add the cited node to the cluster resulted in the quorum node limit being exceeded.

User response: Change the command invocation to not exceed the node quorum limit, and reissue the command.

6027-1235 The *fileName* **kernel extension does not exist. Use the `mmbuildgpl` command to create the needed kernel extension for your kernel or copy the binaries from another node with the identical environment.**

Explanation: The cited kernel extension does not exist.

User response: Create the needed kernel extension by compiling a custom **mmfslinux** module for your kernel (see steps in **/usr/lpp/mmfs/src/README**), or copy the binaries from another node with the identical environment.

6027-1236 Unable to verify kernel/module configuration.

Explanation: The **mmfslinux** kernel extension does not exist.

User response: Create the needed kernel extension by compiling a custom **mmfslinux** module for your kernel (see steps in **/usr/lpp/mmfs/src/README**), or copy the binaries from another node with the identical environment.

6027-1237 The GPFS daemon is still running; use the mmshutdown command.

Explanation: An attempt was made to unload the GPFS kernel extensions while the GPFS daemon was still running.

User response: Use the **mmshutdown** command to shut down the daemon.

6027-1238 Module *fileName* **is still in use. Unmount all GPFS file systems and issue the command: mmfsadm cleanup**

Explanation: An attempt was made to unload the cited module while it was still in use.

User response: Unmount all GPFS file systems and issue the command **mmfsadm cleanup**. If this does not solve the problem, reboot the machine.

6027-1239 **Error unloading module** *moduleName***.**

Explanation: GPFS was unable to unload the cited module.

User response: Unmount all GPFS file systems and issue the command **mmfsadm cleanup**. If this does not solve the problem, reboot the machine.

6027-1240 **Module** *fileName* **is already loaded.**

Explanation: An attempt was made to load the cited module, but it was already loaded.

User response: None. Informational message only.

6027-1241 *diskName* **was not found in /proc/partitions.**

Explanation: The cited disk was not found in **/proc/partitions**.

User response: Take steps to cause the disk to appear in **/proc/partitions**, and then reissue the command.

6027-1242 **GPFS is waiting for** *requiredCondition*

Explanation: GPFS is unable to come up immediately due to the stated required condition not being satisfied yet.

User response: This is an informational message. As long as the required condition is not satisfied, this message will repeat every five minutes. You may want to stop the GPFS daemon after a while, if it will be a long time before the required condition will be met.

6027-1243 *command***: Processing user configuration file** *fileName*

Explanation: Progress information for the **mmcrcluster** command.

User response: None. Informational message only.

6027-1244 *configParameter* **is set by the mmcrcluster processing. Line in error:** *configLine***. The line will be ignored; processing continues.**

Explanation: The specified parameter is set by the **mmcrcluster** command and cannot be overridden by the user.

User response: None. Informational message only.

6027-1245 *configParameter* **must be set with the** *command* **command. Line in error:** *configLine***. The line is ignored; processing continues.**

Explanation: The specified parameter has additional dependencies and cannot be specified prior to the completion of the **mmcrcluster** command.

User response: After the cluster is created, use the specified command to establish the desired configuration parameter.

6027-1246 *configParameter* **is an obsolete parameter. Line in error:** *configLine***. The line is ignored; processing continues.**

Explanation: The specified parameter is not used by GPFS anymore.

User response: None. Informational message only.

6027-1247 *configParameter* **cannot appear in a node-override section. Line in error:** *configLine***. The line is ignored; processing continues.**

Explanation: The specified parameter must have the same value across all nodes in the cluster.

User response: None. Informational message only.

6027-1248 **Mount point can not be a relative path name:** *path*

Explanation: The mount point does not begin with /.

User response: Specify the absolute path name for the mount point.

6027-1249 *operand* **can not be a relative path name:** *path***.**

Explanation: The specified path name does not begin with '/'.

User response: Specify the absolute path name.

6027-1250 **Key file is not valid.**

Explanation: While attempting to establish a connection to another node, GPFS detected that the format of the public key file is not valid.

User response: Use the **mmremotecluster** command to specify the correct public key.

6027-1251 **Key file mismatch.**

Explanation: While attempting to establish a connection to another node, GPFS detected that the public key file does not match the public key file of the cluster to which the file system belongs.

User response: Use the **mmremotecluster** command to specify the correct public key.

6027-1252 **Node *nodeName* already belongs to the GPFS cluster.**

Explanation: A GPFS command found that a node to be added to a GPFS cluster already belongs to the cluster.

User response: Specify a node that does not already belong to the GPFS cluster.

6027-1253 **Incorrect value for *option* option.**

Explanation: The provided value for the specified option is not valid.

User response: Correct the error and reissue the command.

6027-1254 **Warning: Not all nodes have proper GPFS license designations. Use the mmchlicense command to designate licenses as needed.**

Explanation: Not all nodes in the cluster have valid license designations.

User response: Use **mmlslicense** to see the current license designations. Use **mmchlicense** to assign valid GPFS licenses to all nodes as needed.

6027-1255 **There is nothing to commit. You must first run: *command*.**

Explanation: You are attempting to commit an SSL private key but such a key has not been generated yet.

User response: Run the specified command to generate the public/private key pair.

6027-1256 **The current authentication files are already committed.**

Explanation: You are attempting to commit public/private key files that were previously generated with the **mmauth** command. The files have already been committed.

User response: None. Informational message.

6027-1257 **There are uncommitted authentication files. You must first run: *command*.**

Explanation: You are attempting to generate new public/private key files but previously generated files have not been committed yet.

User response: Run the specified command to commit the current public/private key pair.

6027-1258 **You must establish a cipher list first. Run: *command*.**

Explanation: You are attempting to commit an SSL private key but a cipher list has not been established yet.

User response: Run the specified command to specify a cipher list.

6027-1259 ***command* not found. Ensure the OpenSSL code is properly installed.**

Explanation: The specified command was not found.

User response: Ensure the OpenSSL code is properly installed and reissue the command.

6027-1260 **File *fileName* does not contain any *typeOfStanza* stanzas.**

Explanation: The input file should contain at least one specified stanza.

User response: Correct the input file and reissue the command.

6027-1261 ***descriptorField* must be specified in *descriptorType* descriptor.**

Explanation: A required field of the descriptor was empty. The incorrect descriptor is displayed following this message.

User response: Correct the input and reissue the command.

6027-1262 **Unable to obtain the GPFS configuration file lock. Retrying ...**

Explanation: A command requires the lock for the GPFS system data but was not able to obtain it.

User response: None. Informational message only.

6027-1263 **Unable to obtain the GPFS configuration file lock.**

Explanation: A command requires the lock for the GPFS system data but was not able to obtain it.

User response: Check the preceding messages, if any. Follow the procedure in "GPFS cluster configuration data file issues" on page 298, and then reissue the command.

6027-1264 **GPFS is unresponsive on this node.**

Explanation: GPFS is up but not responding to the GPFS commands.

User response: Wait for GPFS to be active again. If the problem persists, perform the problem determination procedures and contact the IBM Support Center.

6027-1268 Missing arguments.

Explanation: A GPFS administration command received an insufficient number of arguments.

User response: Correct the command line and reissue the command.

6027-1269 The device name *device* starts with a slash, but not /dev/.

Explanation: The device name does not start with **/dev/**.

User response: Correct the device name.

6027-1270 The device name *device* contains a slash, but not as its first character.

Explanation: The specified device name contains a slash, but the first character is not a slash.

User response: The device name must be an unqualified device name or an absolute device path name, for example: **fs0** or **/dev/fs0**.

6027-1271 Unexpected error from *command*. Return code: *value*

Explanation: A GPFS administration command (**mm...**) received an unexpected error code from an internally called command.

User response: Perform problem determination. See "GPFS commands are unsuccessful" on page 306.

6027-1272 Unknown user name *userName*.

Explanation: The specified value cannot be resolved to a valid user ID (UID).

User response: Reissue the command with a valid user name.

6027-1273 Unknown group name *groupName*.

Explanation: The specified value cannot be resolved to a valid group ID (GID).

User response: Reissue the command with a valid group name.

6027-1274 Unexpected error obtaining the *lockName* lock.

Explanation: GPFS cannot obtain the specified lock.

User response: Examine any previous error messages. Correct any problems and reissue the command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-1275 Daemon node adapter *Node* was not found on admin node *Node*.

Explanation: An input node descriptor was found to be incorrect. The node adapter specified for GPFS daemon communications was not found to exist on the cited GPFS administrative node.

User response: Correct the input node descriptor and reissue the command.

6027-1276 Command failed for disks: *diskList*.

Explanation: A GPFS command was unable to complete successfully on the listed disks.

User response: Correct the problems and reissue the command.

6027-1277 No contact nodes were provided for cluster *clusterName*.

Explanation: A GPFS command found that no contact nodes have been specified for the cited cluster.

User response: Use the **mmremotecluster** command to specify some contact nodes for the cited cluster.

6027-1278 None of the contact nodes in cluster *clusterName* can be reached.

Explanation: A GPFS command was unable to reach any of the contact nodes for the cited cluster.

User response: Determine why the contact nodes for the cited cluster cannot be reached and correct the problem, or use the **mmremotecluster** command to specify some additional contact nodes that can be reached.

6027-1287 Node *nodeName* returned ENODEV for disk *diskName*.

Explanation: The specified node returned **ENODEV** for the specified disk.

User response: Determine the cause of the **ENODEV** error for the specified disk and rectify it. The **ENODEV** may be due to disk fencing or the removal of a device that previously was present.

6027-1288 Remote cluster *clusterName* was not found.

Explanation: A GPFS command found that the cited cluster has not yet been identified to GPFS as a remote cluster.

User response: Specify a remote cluster known to GPFS, or use the **mmremotecluster** command to make the cited cluster known to GPFS.

6027-1289 **Name** *name* **is not allowed. It contains the following invalid special character:** *char*

Explanation: The cited name is not allowed because it contains the cited invalid special character.

User response: Specify a name that does not contain an invalid special character, and reissue the command.

6027-1290 **GPFS configuration data for file system** *fileSystem* **may not be in agreement with the on-disk data for the file system. Issue the command: mmcommon recoverfs** *fileSystem*

Explanation: GPFS detected that the GPFS configuration database data for the specified file system may not be in agreement with the on-disk data for the file system. This may be caused by a GPFS disk command that did not complete normally.

User response: Issue the specified command to bring the GPFS configuration database into agreement with the on-disk data.

6027-1291 **Options** *name* **and** *name* **cannot be specified at the same time.**

Explanation: Incompatible options were specified on the command line.

User response: Select one of the options and reissue the command.

6027-1292 **The -N option cannot be used with attribute** *name***.**

Explanation: The specified configuration attribute cannot be changed on only a subset of nodes. This attribute must be the same on all nodes in the cluster.

User response: Certain attributes, such as **autoload**, may not be customized from node to node. Change the attribute for the entire cluster.

6027-1293 **There are no remote file systems.**

Explanation: A value of **all** was specified for the remote file system operand of a GPFS command, but no remote file systems are defined.

User response: None. There are no remote file systems on which to operate.

6027-1294 **Remote file system** *fileSystem* **is not defined.**

Explanation: The specified file system was used for the remote file system operand of a GPFS command, but the file system is not known to GPFS.

User response: Specify a remote file system known to GPFS.

6027-1295 **The GPFS configuration information is incorrect or not available.**

Explanation: A problem has been encountered while verifying the configuration information and the execution environment.

User response: Check the preceding messages for more information. Correct the problem and restart GPFS.

6027-1296 **Device name cannot be 'all'.**

Explanation: A device name of **all** was specified on a GPFS command.

User response: Reissue the command with a valid device name.

6027-1297 **Each device specifies metadataOnly for disk usage. This file system could not store data.**

Explanation: All disk descriptors specify **metadataOnly** for disk usage.

User response: Change at least one disk descriptor in the file system to indicate the usage of **dataOnly** or **dataAndMetadata**.

6027-1298 **Each device specifies dataOnly for disk usage. This file system could not store metadata.**

Explanation: All disk descriptors specify **dataOnly** for disk usage.

User response: Change at least one disk descriptor in the file system to indicate a usage of **metadataOnly** or **dataAndMetadata**.

6027-1299 **Incorrect value** *value* **specified for failure group.**

Explanation: The specified failure group is not valid.

User response: Correct the problem and reissue the command.

6027-1300 **No file systems were found.**

Explanation: A GPFS command searched for file systems, but none were found.

User response: Create a GPFS file system before reissuing the command.

6027-1301 **The NSD servers specified in the disk descriptor do not match the NSD servers currently in effect.**

Explanation: The set of NSD servers specified in the disk descriptor does not match the set that is currently in effect.

User response: Specify the same set of NSD servers in the disk descriptor as is currently in effect or omit it from the disk descriptor and then reissue the command. Use the **mmchnsd** command to change the NSD servers as needed.

6027-1302 *clusterName* **is the name of the local cluster.**

Explanation: The cited cluster name was specified as the name of a remote cluster, but it is already being used as the name of the local cluster.

User response: Use the **mmchcluster** command to change the name of the local cluster, and then reissue the command that failed.

6027-1303 **This function is not available in the GPFS Express Edition.**

Explanation: The requested function is not part of the GPFS Express Edition.

User response: Install the GPFS Standard Edition on all nodes in the cluster, and then reissue the command.

6027-1304 **Missing argument after** *option* **option.**

Explanation: The specified command option requires a value.

User response: Specify a value and reissue the command.

6027-1305 **Prerequisite libraries not found or correct version not installed. Ensure** *productName* **is properly installed.**

Explanation: The specified software product is missing or is not properly installed.

User response: Verify that the product is installed properly.

6027-1306 **Command** *command* **failed with return code** *value***.**

Explanation: A command was not successfully processed.

User response: Correct the failure specified by the command and reissue the command.

6027-1307 **Disk** *disk* **on node** *nodeName* **already has a volume group** *vgName* **that does not appear to have been created by this program in a prior invocation. Correct the descriptor file or remove the volume group and retry.**

Explanation: The specified disk already belongs to a volume group.

User response: Either remove the volume group or remove the disk descriptor and retry.

6027-1308 *feature* **is not available in the GPFS Express Edition.**

Explanation: The specified function or feature is not part of the GPFS Express Edition.

User response: Install the GPFS Standard Edition on all nodes in the cluster, and then reissue the command.

6027-1309 **Storage pools are not available in the GPFS Express Edition.**

Explanation: Support for multiple storage pools is not part of the GPFS Express Edition.

User response: Install the GPFS Standard Edition on all nodes in the cluster, and then reissue the command.

6027-1332 **Cannot find** *disk* **with** *command***.**

Explanation: The specified disk cannot be found.

User response: Specify a correct disk name.

6027-1333 **The following nodes could not be restored:** *nodeList***. Correct the problems and use the mmsdrrestore command to recover these nodes.**

Explanation: The **mmsdrrestore** command was unable to restore the configuration information for the listed nodes.

User response: Correct the problems and reissue the **mmsdrrestore** command for these nodes.

6027-1334 **Incorrect value for option** *option***. Valid values are:** *validValues***.**

Explanation: An incorrect argument was specified with an option requiring one of a limited number of legal options.

User response: Use one of the legal values for the indicated option.

6027-1335 Command completed: Not all required changes were made.

Explanation: Some, but not all, of the required changes were made.

User response: Examine the preceding messages, correct the problems, and reissue the command.

6027-1338 Command is not allowed for remote file systems.

Explanation: A command for which a remote file system is not allowed was issued against a remote file system.

User response: Choose a local file system, or issue the command on a node in the cluster that owns the file system.

6027-1339 Disk usage *value* is incompatible with storage pool *name*.

Explanation: A disk descriptor specified a disk usage involving metadata and a storage pool other than **system**.

User response: Change the descriptor's disk usage field to **dataOnly**, or do not specify a storage pool name.

6027-1340 File *fileName* not found. Recover the file or run mmauth genkey.

Explanation: The cited file was not found.

User response: Recover the file or run the **mmauth genkey** command to recreate it.

6027-1341 Starting force unmount of GPFS file systems

Explanation: Progress information for the **mmshutdown** command.

User response: None. Informational message only.

6027-1342 Unmount not finished after *value* seconds. Waiting *value* more seconds.

Explanation: Progress information for the **mmshutdown** command.

User response: None. Informational message only.

6027-1343 Unmount not finished after *value* seconds.

Explanation: Progress information for the **mmshutdown** command.

User response: None. Informational message only.

6027-1344 Shutting down GPFS daemons

Explanation: Progress information for the **mmshutdown** command.

User response: None. Informational message only.

6027-1345 Finished

Explanation: Progress information for the **mmshutdown** command.

User response: None. Informational message only.

6027-1347 Disk with NSD volume id *NSD volume id* no longer exists in the GPFS cluster configuration data but the NSD volume id was not erased from the disk. To remove the NSD volume id, issue: mmdelnsd -p *NSD volume id*

Explanation: A GPFS administration command (**mm...**) successfully removed the disk with the specified NSD volume id from the GPFS cluster configuration data but was unable to erase the NSD volume id from the disk.

User response: Issue the specified command to remove the NSD volume id from the disk.

6027-1348 Disk with NSD volume id *NSD volume id* no longer exists in the GPFS cluster configuration data but the NSD volume id was not erased from the disk. To remove the NSD volume id, issue: mmdelnsd -p *NSD volume id* -N *nodeNameList*

Explanation: A GPFS administration command (**mm...**) successfully removed the disk with the specified NSD volume id from the GPFS cluster configuration data but was unable to erase the NSD volume id from the disk.

User response: Issue the specified command to remove the NSD volume id from the disk.

6027-1352 *fileSystem* is not a remote file system known to GPFS.

Explanation: The cited file system is not the name of a remote file system known to GPFS.

User response: Use the **mmremotefs** command to identify the cited file system to GPFS as a remote file system, and then reissue the command that failed.

6027-1357 An internode connection between GPFS nodes was disrupted.

Explanation: An internode connection between GPFS nodes was disrupted, preventing its successful completion.

User response: Reissue the command. If the problem

recurs, determine and resolve the cause of the disruption. If the problem persists, contact the IBM Support Center.

6027-1358 **No clusters are authorized to access this cluster.**

Explanation: Self-explanatory.

User response: This is an informational message.

6027-1359 **Cluster** *clusterName* **is not authorized to access this cluster.**

Explanation: Self-explanatory.

User response: This is an informational message.

6027-1361 **Attention: There are no available valid VFS type values for mmfs in /etc/vfs.**

Explanation: An out of range number was used as the vfs number for GPFS.

User response: The valid range is 8 through 32. Check **/etc/vfs** and remove unneeded entries.

6027-1362 **There are no remote cluster definitions.**

Explanation: A value of **all** was specified for the remote cluster operand of a GPFS command, but no remote clusters are defined.

User response: None. There are no remote clusters on which to operate.

6027-1363 **Remote cluster** *clusterName* **is not defined.**

Explanation: The specified cluster was specified for the remote cluster operand of a GPFS command, but the cluster is not known to GPFS.

User response: Specify a remote cluster known to GPFS.

6027-1364 **No disks specified**

Explanation: There were no disks in the descriptor list or file.

User response: Specify at least one disk.

6027-1365 **Disk** *diskName* **already belongs to file system** *fileSystem***.**

Explanation: The specified disk name is already assigned to a GPFS file system. This may be because the disk was specified more than once as input to the command, or because the disk was assigned to a GPFS file system in the past.

User response: Specify the disk only once as input to the command, or specify a disk that does not belong to a file system.

6027-1366 **File system** *fileSystem* **has some disks that are in a non-ready state.**

Explanation: The specified file system has some disks that are in a non-ready state.

User response: Run **mmcommon recoverfs** *fileSystem* to ensure that the GPFS configuration data for the file system is current. If some disks are still in a non-ready state, display the states of the disks in the file system using the **mmlsdisk** command. Any disks in an undesired non-ready state should be brought into the ready state by using the **mmchdisk** command or by mounting the file system. If these steps do not bring the disks into the ready state, use the **mmdeldisk** command to delete the disks from the file system.

6027-1367 **Attention: Not all disks were marked as available.**

Explanation: The process of marking the disks as available could not be completed.

User response: Before adding these disks to a GPFS file system, you should either reformat them, or use the **-v no** option on the **mmcrfs** or **mmadddisk** command.

6027-1368 **This GPFS cluster contains declarations for remote file systems and clusters. You cannot delete the last node. First use the delete option of the mmremotecluster and mmremotefs commands.**

Explanation: An attempt has been made to delete a GPFS cluster that still has declarations for remote file systems and clusters.

User response: Before deleting the last node of a GPFS cluster, delete all remote cluster and file system information. Use the **delete** option of the **mmremotecluster** and **mmremotefs** commands.

6027-1370 **The following nodes could not be reached:**

Explanation: A GPFS command was unable to communicate with one or more nodes in the cluster. A list of the nodes that could not be reached follows.

User response: Determine why the reported nodes could not be reached and resolve the problem.

6027-1371 **Propagating the cluster configuration data to all affected nodes. This is an asynchronous process.**

Explanation: A process is initiated to distribute the cluster configuration data to other nodes in the cluster.

User response: This is an informational message. The

command does not wait for the distribution to finish.

6027-1373 There is no file system information in input file *fileName*.

Explanation: The cited input file passed to the **mmimportfs** command contains no file system information. No file system can be imported.

User response: Reissue the **mmimportfs** command while specifying a valid input file.

6027-1374 File system *fileSystem* was not found in input file *fileName*.

Explanation: The specified file system was not found in the input file passed to the **mmimportfs** command. The file system cannot be imported.

User response: Reissue the **mmimportfs** command while specifying a file system that exists in the input file.

6027-1375 The following file systems were not imported: *fileSystem*.

Explanation: The **mmimportfs** command was unable to import one or more of the file systems in the input file. A list of the file systems that could not be imported follows.

User response: Examine the preceding messages, rectify the problems that prevented the importation of the file systems, and reissue the **mmimportfs** command.

6027-1377 Attention: Unknown attribute specified: *name*. Press the ENTER key to continue.

Explanation: The **mmchconfig** command received an unknown attribute.

User response: Unless directed otherwise by the IBM Support Center, press any key to bypass this attribute.

6027-1378 Incorrect record found in the mmsdrfs file (code *value*):

Explanation: A line that is not valid was detected in the main GPFS cluster configuration file **/var/mmfs/gen/mmsdrfs**.

User response: The data in the cluster configuration file is incorrect. If no user modifications have been made to this file, contact the IBM Support Center. If user modifications have been made, correct these modifications.

6027-1379 There is no file system with mount point *mountpoint*.

Explanation: No file system in the GPFS cluster has the specified mount point.

User response: Reissue the command with a valid file system.

6027-1380 File system *fileSystem* is already mounted at *mountpoint*.

Explanation: The specified file system is mounted at a mount point different than the one requested on the **mmmount** command.

User response: Unmount the file system and reissue the command.

6027-1381 Mount point cannot be specified when mounting all file systems.

Explanation: A device name of **all** and a mount point were specified on the **mmmount** command.

User response: Reissue the command with a device name for a single file system or do not specify a mount point.

6027-1382 This node does not belong to a GPFS cluster.

Explanation: The specified node does not appear to belong to a GPFS cluster, or the GPFS configuration information on the node has been lost.

User response: Informational message. If you suspect that there is corruption of the GPFS configuration information, recover the data following the procedures outlined in “Recovery from loss of GPFS cluster configuration data file” on page 299.

6027-1383 There is no record for this node in file *fileName*. Either the node is not part of the cluster, the file is for a different cluster, or not all of the node's adapter interfaces have been activated yet.

Explanation: The **mmsdrrestore** command cannot find a record for this node in the specified cluster configuration file. The search of the file is based on the currently active IP addresses of the node as reported by the **ifconfig** command.

User response: Ensure that all adapter interfaces are properly functioning. Ensure that the correct GPFS configuration file is specified on the command line. If the node indeed is not a member of the cluster, use the **mmaddnode** command instead.

6027-1386 Unexpected value for Gpfs object: *value*.

Explanation: A function received a value that is not allowed for the Gpfs object.

User response: Perform problem determination.

6027-1388 File system *fileSystem* is not known to the GPFS cluster.

Explanation: The file system was not found in the GPFS cluster.

User response: If the file system was specified as part of a GPFS command, reissue the command with a valid file system.

6027-1390 Node *node* does not belong to the GPFS cluster, or was specified as input multiple times.

Explanation: Nodes that are not valid were specified.

User response: Verify the list of nodes. All specified nodes must belong to the GPFS cluster, and each node can be specified only once.

6027-1393 Incorrect node designation specified: *type*.

Explanation: A node designation that is not valid was specified. Valid values are **client** or **manager**.

User response: Correct the command line and reissue the command.

6027-1394 Operation not allowed for the local cluster.

Explanation: The requested operation cannot be performed for the local cluster.

User response: Specify the name of a remote cluster.

6027-1450 Could not allocate storage.

Explanation: Sufficient memory cannot be allocated to run the **mmsanrepairfs** command.

User response: Increase the amount of memory available.

6027-1500 [E] Open *devicetype device* failed with error:

Explanation: The "open" of a device failed. Operation of the file system may continue unless this device is needed for operation. If this is a replicated disk device, it will often not be needed. If this is a block or character device for another subsystem (such as /dev/VSD0) then GPFS will discontinue operation.

User response: Problem diagnosis will depend on the subsystem that the device belongs to. For instance device "/dev/VSD0" belongs to the IBM Virtual Shared Disk subsystem and problem determination should follow guidelines in that subsystem's documentation. If this is a normal disk device then take needed repair action on the specified disk.

6027-1501 [X] Volume label of disk *name* is *name*, should be *uid*.

Explanation: The UID in the disk descriptor does not match the expected value from the file system descriptor. This could occur if a disk was overwritten by another application or if the IBM Virtual Shared Disk subsystem incorrectly identified the disk.

User response: Check the disk configuration.

6027-1502 [X] Volume label of disk *diskName* is corrupt.

Explanation: The disk descriptor has a bad magic number, version, or checksum. This could occur if a disk was overwritten by another application or if the IBM Virtual Shared Disk subsystem incorrectly identified the disk.

User response: Check the disk configuration.

6027-1503 Completed adding disks to file system *fileSystem*.

Explanation: The **mmadddisk** command successfully completed.

User response: None. Informational message only.

6027-1504 File *name* could not be run with err *error*.

Explanation: A failure occurred while trying to run an external program.

User response: Make sure the file exists. If it does, check its access permissions.

6027-1505 Could not get minor number for *name*.

Explanation: Could not obtain a minor number for the specified block or character device.

User response: Problem diagnosis will depend on the subsystem that the device belongs to. For example, device **/dev/VSD0** belongs to the IBM Virtual Shared Disk subsystem and problem determination should follow guidelines in that subsystem's documentation.

6027-1507 READ_KEYS ioctl failed with errno=*returnCode*, tried *timesTried* times. Related values are scsi_status=*scsiStatusValue*, sense_key=*senseKeyValue*, scsi_asc=*scsiAscValue*, scsi_ascq=*scsiAscqValue*.

Explanation: A READ_KEYS **ioctl** call failed with the

errno= and related values shown.

User response: Check the reported **errno=** value and try to correct the problem. If the problem persists, contact the IBM Support Center.

6027-1508 **Registration failed with errno=***returnCode*, **tried** *timesTried* **times. Related values are scsi_status=***scsiStatusValue*, **sense_key=***senseKeyValue*, **scsi_asc=***scsiAscValue*, **scsi_ascq=***scsiAscqValue*.

Explanation: A REGISTER **ioctl** call failed with the **errno=** and related values shown.

User response: Check the reported **errno=** value and try to correct the problem. If the problem persists, contact the IBM Support Center.

6027-1509 **READRES ioctl failed with errno=***returnCode*, **tried** *timesTried* **times. Related values are scsi_status=***scsiStatusValue*, **sense_key=***senseKeyValue*, **scsi_asc=***scsiAscValue*, **scsi_ascq=***scsiAscqValue*.

Explanation: A READRES **ioctl** call failed with the **errno=** and related values shown.

User response: Check the reported **errno=** value and try to correct the problem. If the problem persists, contact the IBM Support Center.

6027-1510 [E] Error mounting file system *stripeGroup* **on** *mountPoint*; *errorQualifier* **(***gpfsErrno***)**

Explanation: An error occurred while attempting to mount a GPFS file system on Windows.

User response: Examine the error details, previous errors, and the GPFS message log to identify the cause.

6027-1511 [E] Error unmounting file system *stripeGroup*; *errorQualifier* **(***gpfsErrno***)**

Explanation: An error occurred while attempting to unmount a GPFS file system on Windows.

User response: Examine the error details, previous errors, and the GPFS message log to identify the cause.

6027-1512 [E] WMI query for *queryType* **failed;** *errorQualifier* **(***gpfsErrno***)**

Explanation: An error occurred while running a WMI query on Windows.

User response: Examine the error details, previous errors, and the GPFS message log to identify the cause.

6027-1513 *DiskName* **is not an sg device, or sg driver is older than sg3**

Explanation: The disk is not a SCSI disk, or supports SCSI standard older than SCSI 3.

User response: Correct the command invocation and try again.

6027-1514 **ioctl failed with rc=***returnCode*. **Related values are SCSI status=***scsiStatusValue*, **host_status=***hostStatusValue*, **driver_status=***driverStatsValue*.

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1515 **READ KEY ioctl failed with rc=***returnCode*. **Related values are SCSI status=***scsiStatusValue*, **host_status=***hostStatusValue*, **driver_status=***driverStatsValue*.

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1516 **REGISTER ioctl failed with rc=***returnCode*. **Related values are SCSI status=***scsiStatusValue*, **host_status=***hostStatusValue*, **driver_status=***driverStatsValue*.

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1517 **READ RESERVE ioctl failed with rc=***returnCode*. **Related values are SCSI status=***scsiStatusValue*, **host_status=***hostStatusValue*, **driver_status=***driverStatsValue*.

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1518 **RESERVE ioctl failed with rc=***returnCode***. Related values are SCSI status=***scsiStatusValue***, host_status=***hostStatusValue***, driver_status=***driverStatsValue***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1519 **INQUIRY ioctl failed with rc=***returnCode***. Related values are SCSI status=***scsiStatusValue***, host_status=***hostStatusValue***, driver_status=***driverStatsValue***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1520 **PREEMPT ABORT ioctl failed with rc=***returnCode***. Related values are SCSI status=***scsiStatusValue***, host_status=***hostStatusValue***, driver_status=***driverStatsValue***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1521 **Can not find register key** *registerKeyValue* **at device** *diskName***.**

Explanation: Unable to find given register key at the disk.

User response: Correct the problem and reissue the command.

6027-1522 **CLEAR ioctl failed with rc=***returnCode***. Related values are SCSI status=***scsiStatusValue***, host_status=***hostStatusValue***, driver_status=***driverStatsValue***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-1523 **Disk name longer than** *value* **is not allowed.**

Explanation: The specified disk name is too long.

User response: Reissue the command with a valid disk name.

6027-1524 **The READ_KEYS ioctl data does not contain the key that was passed as input.**

Explanation: A REGISTER **ioctl** call apparently succeeded, but when the device was queried for the key, the key was not found.

User response: Check the device subsystem and try to correct the problem. If the problem persists, contact the IBM Support Center.

6027-1525 **Invalid `minReleaseLevel` parameter:** *value*

Explanation: The second argument to the **`mmcrfsc`** command is **`minReleaseLevel`** and should be greater than 0.

User response: **`minReleaseLevel`** should be greater than 0. The **`mmcrfs`** command should never call the **`mmcrfsc`** command without a valid **`minReleaseLevel`** argument. Contact the IBM Support Center.

6027-1530 **Attention:** *parameter* **is set to** *value***.**

Explanation: A configuration parameter is temporarily assigned a new value.

User response: Check the **mmfs.cfg** file. Use the **mmchconfig** command to set a valid value for the parameter.

6027-1531 *parameter value*

Explanation: The configuration parameter was changed from its default value.

User response: Check the **mmfs.cfg** file.

6027-1532 **Attention:** *parameter* **(***value***) is not valid in conjunction with** *parameter* **(***value***).**

Explanation: A configuration parameter has a value that is not valid in relation to some other parameter. This can also happen when the default value for some parameter is not sufficiently large for the new, user set value of a related parameter.

User response: Check the **mmfs.cfg** file.

6027-1533 *parameter* **cannot be set dynamically.**

Explanation: The **mmchconfig** command encountered a configuration parameter that cannot be set dynamically.

User response: Check the **mmchconfig** command arguments. If the parameter must be changed, use the **mmshutdown**, **mmchconfig**, and **mmstartup** sequence of commands.

6027-1534 *parameter* **must have a value.**

Explanation: The **tsctl** command encountered a configuration parameter that did not have a specified value.

User response: Check the **mmchconfig** command arguments.

6027-1535 **Unknown config name:** *parameter*

Explanation: The **tsctl** command encountered an unknown configuration parameter.

User response: Check the **mmchconfig** command arguments.

6027-1536 *parameter* **must be set using the tschpool command.**

Explanation: The **tsctl** command encountered a configuration parameter that must be set using the **tschpool** command.

User response: Check the **mmchconfig** command arguments.

6027-1537 [E] Connect failed to *ipAddress*: *reason*

Explanation: An attempt to connect sockets between nodes failed.

User response: Check the reason listed and the connection to the indicated IP address.

6027-1538 [I] Connect in progress to *ipAddress*

Explanation: Connecting sockets between nodes.

User response: None. Information message only.

6027-1539 [E] Connect progress select failed to *ipAddress*: *reason*

Explanation: An attempt to connect sockets between nodes failed.

User response: Check the reason listed and the connection to the indicated IP address.

6027-1540 [A] Try and buy license has expired!

Explanation: Self explanatory.

User response: Purchase a GPFS license to continue using GPFS.

6027-1541 [N] Try and buy license expires in *number* **days.**

Explanation: Self-explanatory.

User response: When the **Try and Buy** license expires, you will need to purchase a GPFS license to continue using GPFS.

6027-1542 [A] Old shared memory exists but it is not valid nor cleanable.

Explanation: A new GPFS daemon started and found existing shared segments. The contents were not recognizable, so the GPFS daemon could not clean them up.

User response:

1. Stop the GPFS daemon from trying to start by issuing the **mmshutdown** command for the nodes having the problem.
2. Find the owner of the shared segments with keys from `0x9283a0ca` through `0x9283a0d1`. If a non-GPFS program owns these segments, GPFS cannot run on this node.
3. If these segments are left over from a previous GPFS daemon:
   a. Remove them by issuing:
      `ipcrm -m shared_memory_id`
   b. Restart GPFS by issuing the **mmstartup** command on the affected nodes.

6027-1543 **error propagating** *parameter*.

Explanation: **mmfsd** could not propagate a configuration parameter value to one or more nodes in the cluster.

User response: Contact the IBM Support Center.

6027-1544 [W] Sum of prefetchthreads(*value***), worker1threads(***value***) and nsdMaxWorkerThreads (***value***) exceeds** *value***. Reducing them to** *value*, *value* **and** *value*.

Explanation: The sum of **prefetchthreads**, **worker1threads**, and **nsdMaxWorkerThreads** exceeds the permitted value.

User response: Accept the calculated values or reduce the individual settings using **mmchconfig prefetchthreads=newvalue** or **mmchconfig worker1threads=newvalue**. or **mmchconfig**

nsdMaxWorkerThreads=newvalue. After using **mmchconfig**, the new settings will not take affect until the GPFS daemon is restarted.

6027-1545 [A] The GPFS product that you are attempting to run is not a fully functioning version. This probably means that this is an update version and not the full product version. Install the GPFS full product version first, then apply any applicable update version before attempting to start GPFS.

Explanation: GPFS requires a fully licensed GPFS installation.

User response: Verify installation of licensed GPFS, or purchase and install a licensed version of GPFS.

6027-1546 [W] Attention: *parameter* size of *value* is too small. New value is *value*.

Explanation: A configuration parameter is temporarily assigned a new value.

User response: Check the **mmfs.cfg** file. Use the **mmchconfig** command to set a valid value for the parameter.

6027-1547 [A] Error initializing daemon: performing shutdown

Explanation: GPFS kernel extensions are not loaded, and the daemon cannot initialize. GPFS may have been started incorrectly.

User response: Check GPFS log for errors resulting from kernel extension loading. Ensure that GPFS is started with the **mmstartup** command.

6027-1548 [A] Error: daemon and kernel extension do not match.

Explanation: The GPFS kernel extension loaded in memory and the daemon currently starting do not appear to have come from the same build.

User response: Ensure that the kernel extension was reloaded after upgrading GPFS. See "GPFS modules cannot be loaded on Linux" on page 301 for details.

6027-1549 [A] Attention: custom-built kernel extension; the daemon and kernel extension do not match.

Explanation: The GPFS kernel extension loaded in memory does not come from the same build as the starting daemon. The kernel extension appears to have been built from the kernel open source package.

User response: None.

6027-1550 [W] Error: Unable to establish a session with an Active Directory server. ID remapping via Microsoft Identity Management for Unix will be unavailable.

Explanation: GPFS tried to establish an LDAP session with an Active Directory server (normally the domain controller host), and has been unable to do so.

User response: Ensure the domain controller is available.

6027-1555 Mount point and device name cannot be equal: *name*

Explanation: The specified mount point is the same as the absolute device name.

User response: Enter a new device name or absolute mount point path name.

6027-1556 Interrupt received.

Explanation: A GPFS administration command received an interrupt.

User response: None. Informational message only.

6027-1557 You must first generate an authentication key file. Run: mmauth genkey new.

Explanation: Before setting a cipher list, you must generate an authentication key file.

User response: Run the specified command to establish an authentication key for the nodes in the cluster.

6027-1559 The -i option failed. Changes will take effect after GPFS is restarted.

Explanation: The **-i** option on the **mmchconfig** command failed. The changes were processed successfully, but will take effect only after the GPFS daemons are restarted.

User response: Check for additional error messages. Correct the problem and reissue the command.

6027-1560 This GPFS cluster contains file systems. You cannot delete the last node.

Explanation: An attempt has been made to delete a GPFS cluster that still has one or more file systems associated with it.

User response: Before deleting the last node of a GPFS cluster, delete all file systems that are associated with it. This applies to both local and remote file systems.

6027-1561 Attention: Failed to remove node-specific changes.

Explanation: The internal **mmfixcfg** routine failed to remove node-specific configuration settings, if any, for one or more of the nodes being deleted. This is of consequence only if the **mmchconfig** command was indeed used to establish node specific settings and these nodes are later added back into the cluster.

User response: If you add the nodes back later, ensure that the configuration parameters for the nodes are set as desired.

6027-1562 *command* **command cannot be executed. Either none of the nodes in the cluster are reachable, or GPFS is down on all of the nodes.**

Explanation: The command that was issued needed to perform an operation on a remote node, but none of the nodes in the cluster were reachable, or GPFS was not accepting commands on any of the nodes.

User response: Ensure that the affected nodes are available and all authorization requirements are met. Correct any problems and reissue the command.

6027-1563 Attention: The file system may no longer be properly balanced.

Explanation: The restripe phase of the **mmadddisk** or **mmdeldisk** command failed.

User response: Determine the cause of the failure and run the **mmrestripefs -b** command.

6027-1564 To change the authentication key for the local cluster, run: mmauth genkey.

Explanation: The authentication keys for the local cluster must be created only with the specified command.

User response: Run the specified command to establish a new authentication key for the nodes in the cluster.

6027-1565 *disk* **not found in file system** *fileSystem***.**

Explanation: A disk specified for deletion or replacement does not exist.

User response: Specify existing disks for the indicated file system.

6027-1566 Remote cluster *clusterName* **is already defined.**

Explanation: A request was made to add the cited cluster, but the cluster is already known to GPFS.

User response: None. The cluster is already known to GPFS.

6027-1567 *fileSystem* **from cluster** *clusterName* **is already defined.**

Explanation: A request was made to add the cited file system from the cited cluster, but the file system is already known to GPFS.

User response: None. The file system is already known to GPFS.

6027-1568 *command* **command failed. Only** *parameterList* **changed.**

Explanation: The **mmchfs** command failed while making the requested changes. Any changes to the attributes in the indicated parameter list were successfully completed. No other file system attributes were changed.

User response: Reissue the command if you want to change additional attributes of the file system. Changes can be undone by issuing the **mmchfs** command with the original value for the affected attribute.

6027-1570 virtual shared disk support is not installed.

Explanation: The command detected that IBM Virtual Shared Disk support is not installed on the node on which it is running.

User response: Install IBM Virtual Shared Disk support.

6027-1571 *commandName* **does not exist or failed; automount mounting may not work.**

Explanation: One or more of the GPFS file systems were defined with the automount attribute but the requisite automount command is missing or failed.

User response: Correct the problem and restart GPFS. Or use the **mount** command to explicitly mount the file system.

6027-1572 The command must run on a node that is part of the cluster.

Explanation: The node running the **mmcrcluster** command (this node) must be a member of the GPFS cluster.

User response: Issue the command from a node that will belong to the cluster.

6027-1573 Command completed: No changes made.

Explanation: Informational message.

User response: Check the preceding messages, correct any problems, and reissue the command.

6027-1574 Permission failure. The command requires root authority to execute.

Explanation: The command, or the specified command option, requires root authority.

User response: Log on as **root** and reissue the command.

6027-1578 File *fileName* **does not contain node names.**

Explanation: The specified file does not contain valid node names.

User response: Node names must be specified one per line. The name **localhost** and lines that start with '#' character are ignored.

6027-1579 File *fileName* **does not contain data.**

Explanation: The specified file does not contain data.

User response: Verify that you are specifying the correct file name and reissue the command.

6027-1587 Unable to determine the local device name for disk *nsdName* **on node** *nodeName***.**

Explanation: GPFS was unable to determine the local device name for the specified GPFS disk.

User response: Determine why the specified disk on the specified node could not be accessed and correct the problem. Possible reasons include: connectivity problems, authorization problems, fenced disk, and so forth.

6027-1588 Unknown GPFS execution environment: *value*

Explanation: A GPFS administration command (prefixed by **mm**) was asked to operate on an unknown GPFS cluster type. The only supported GPFS cluster type is **lc**. This message may also be generated if there is corruption in the GPFS system files.

User response: Verify that the correct level of GPFS is installed on the node. If this is a cluster environment, make sure the node has been defined as a member of the GPFS cluster with the help of the **mmcrcluster** or the **mmaddnode** command. If the problem persists, contact the IBM Support Center.

6027-1590 *nodeName* **cannot be reached.**

Explanation: A command needs to issue a remote function on a particular node but the node is not reachable.

User response: Determine why the node is unreachable, correct the problem, and reissue the command.

6027-1591 Attention: Unable to retrieve GPFS cluster files from node *nodeName*

Explanation: A command could not retrieve the GPFS cluster files from a particular node. An attempt will be made to retrieve the GPFS cluster files from a backup node.

User response: None. Informational message only.

6027-1592 Unable to retrieve GPFS cluster files from node *nodeName*

Explanation: A command could not retrieve the GPFS cluster files from a particular node.

User response: Correct the problem and reissue the command.

6027-1594 Run the *command* **command until successful.**

Explanation: The command could not complete normally. The GPFS cluster data may be left in a state that precludes normal operation until the problem is corrected.

User response: Check the preceding messages, correct the problems, and issue the specified command until it completes successfully.

6027-1595 No nodes were found that matched the input specification.

Explanation: No nodes were found in the GPFS cluster that matched those specified as input to a GPFS command.

User response: Determine why the specified nodes were not valid, correct the problem, and reissue the GPFS command.

6027-1596 The same node was specified for both the primary and the secondary server.

Explanation: A command would have caused the primary and secondary GPFS cluster configuration server nodes to be the same.

User response: Specify a different primary or secondary node.

6027-1597 **Node** *node* **is specified more than once.**

Explanation: The same node appears more than once on the command line or in the input file for the command.

User response: All specified nodes must be unique. Note that even though two node identifiers may appear different on the command line or in the input file, they may still refer to the same node.

6027-1598 **Node** *nodeName* **was not added to the cluster. The node appears to already belong to a GPFS cluster.**

Explanation: A GPFS cluster command found that a node to be added to a cluster already has GPFS cluster files on it.

User response: Use the **mmlscluster** command to verify that the node is in the correct cluster. If it is not, follow the procedure in "Node cannot be added to the GPFS cluster" on page 295.

6027-1599 **The level of GPFS on node** *nodeName* **does not support the requested action.**

Explanation: A GPFS command found that the level of the GPFS code on the specified node is not sufficient for the requested action.

User response: Install the correct level of GPFS.

6027-1600 **Make sure that the following nodes are available:** *nodeList*

Explanation: A GPFS command was unable to complete because nodes critical for the success of the operation were not reachable or the command was interrupted.

User response: This message will normally be followed by a message telling you which command to issue as soon as the problem is corrected and the specified nodes become available.

6027-1602 *nodeName* **is not a member of this cluster.**

Explanation: A command found that the specified node is not a member of the GPFS cluster.

User response: Correct the input or add the node to the GPFS cluster and reissue the command.

6027-1603 **The following nodes could not be added to the GPFS cluster:** *nodeList***. Correct the problems and use the mmaddnode command to add these nodes to the cluster.**

Explanation: The **mmcrcluster** or the **mmaddnode** command was unable to add the listed nodes to a GPFS cluster.

User response: Correct the problems and add the nodes to the cluster using the **mmaddnode** command.

6027-1604 **Information cannot be displayed. Either none of the nodes in the cluster are reachable, or GPFS is down on all of the nodes.**

Explanation: The command needed to perform an operation on a remote node, but none of the nodes in the cluster were reachable, or GPFS was not accepting commands on any of the nodes.

User response: Ensure that the affected nodes are available and all authorization requirements are met. Correct any problems and reissue the command.

6027-1610 **Disk** *diskName* **is the only disk in file system** *fileSystem***. You cannot replace a disk when it is the only remaining disk in the file system.**

Explanation: The **mmrpldisk** command was issued, but there is only one disk in the file system.

User response: Add a second disk and reissue the command.

6027-1613 **WCOLL (working collective) environment variable not set.**

Explanation: The **mmdsh** command was invoked without explicitly specifying the nodes on which the command is to run by means of the **-F** or **-L** options, and the WCOLL environment variable has not been set.

User response: Change the invocation of the **mmdsh** command to use the **-F** or **-L** options, or set the WCOLL environment variable before invoking the **mmdsh** command.

6027-1614 **Cannot open file** *fileName***. Error string was:** *errorString***.**

Explanation: The **mmdsh** command was unable to successfully open a file.

User response: Determine why the file could not be opened and correct the problem.

6027-1615 *nodeName* **remote shell process had return code** *value***.**

Explanation: A child remote shell process completed with a nonzero return code.

User response: Determine why the child remote shell process failed and correct the problem.

6027-1616 **Caught SIG** *signal* **- terminating the child processes.**

Explanation: The **mmdsh** command has received a signal causing it to terminate.

User response: Determine what caused the signal and correct the problem.

6027-1617 **There are no available nodes on which to run the command.**

Explanation: The **mmdsh** command found that there are no available nodes on which to run the specified command. Although nodes were specified, none of the nodes were reachable.

User response: Determine why the specified nodes were not available and correct the problem.

6027-1618 **Unable to pipe. Error string was:** *errorString***.**

Explanation: The **mmdsh** command attempted to open a pipe, but the pipe command failed.

User response: Determine why the call to pipe failed and correct the problem.

6027-1619 **Unable to redirect** *outputStream***. Error string was:** *string***.**

Explanation: The **mmdsh** command attempted to redirect an output stream using open, but the open command failed.

User response: Determine why the call to open failed and correct the problem.

6027-1623 *command***: Mounting file systems ...**

Explanation: This message contains progress information about the **mmmount** command.

User response: None. Informational message only.

6027-1625 *option* **cannot be used with attribute** *name***.**

Explanation: An attempt was made to change a configuration attribute and requested the change to take effect immediately (**-i** or **-I** option). However, the specified attribute does not allow the operation.

User response: If the change must be made now, leave off the **-i** or **-I** option. Then recycle the nodes to pick up the new value.

6027-1626 **Command is not supported in the** *type* **environment.**

Explanation: A GPFS administration command (**mm...**) is not supported in the specified environment.

User response: Verify if the task is needed in this environment, and if it is, use a different command.

6027-1627 **The following nodes are not aware of the configuration server change:** *nodeList***. Do not start GPFS on the above nodes until the problem is resolved.**

Explanation: The **mmchcluster** command could not propagate the new cluster configuration servers to the specified nodes.

User response: Correct the problems and run the **mmchcluster -p LATEST** command before starting GPFS on the specified nodes.

6027-1628 **Cannot determine basic environment information. Not enough nodes are available.**

Explanation: The **mmchcluster** command was unable to retrieve the GPFS cluster data files. Usually, this is due to too few nodes being available.

User response: Correct any problems and ensure that as many of the nodes in the cluster are available as possible. Reissue the command. If the problem persists, record the above information and contact the IBM Support Center.

6027-1629 **Error found while checking node descriptor** *descriptor*

Explanation: A node descriptor was found to be unsatisfactory in some way.

User response: Check the preceding messages, if any, and correct the condition that caused the disk descriptor to be rejected.

6027-1630 **The GPFS cluster data on** *nodeName* **is back level.**

Explanation: A GPFS command attempted to commit changes to the GPFS cluster configuration data, but the data on the server is already at a higher level. This can happen if the GPFS cluster configuration files were altered outside the GPFS environment, or if the **mmchcluster** command did not complete successfully.

User response: Correct any problems and reissue the command. If the problem persists, issue the **mmrefresh -f -a** command.

6027-1631 The commit process failed.

Explanation: A GPFS administration command (**mm...**) cannot commit its changes to the GPFS cluster configuration data.

User response: Examine the preceding messages, correct the problem, and reissue the command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-1632 The GPFS cluster configuration data on *nodeName* is different than the data on *nodeName*.

Explanation: The GPFS cluster configuration data on the primary cluster configuration server node is different than the data on the secondary cluster configuration server node. This can happen if the GPFS cluster configuration files were altered outside the GPFS environment or if the **mmchcluster** command did not complete successfully.

User response: Correct any problems and issue the **mmrefresh -f -a** command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-1633 Failed to create a backup copy of the GPFS cluster data on *nodeName*.

Explanation: Commit could not create a correct copy of the GPFS cluster configuration data.

User response: Check the preceding messages, correct any problems, and reissue the command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-1634 The GPFS cluster configuration server node *nodeName* cannot be removed.

Explanation: An attempt was made to delete a GPFS cluster configuration server node.

User response: You cannot remove a cluster configuration server node unless all nodes in the GPFS cluster are being deleted. Before deleting a cluster configuration server node, you must use the **mmchcluster** command to transfer its function to another node in the GPFS cluster.

6027-1636 Error found while checking disk descriptor *descriptor*

Explanation: A disk descriptor was found to be unsatisfactory in some way.

User response: Check the preceding messages, if any, and correct the condition that caused the disk descriptor to be rejected.

6027-1637 *command* quitting. None of the specified nodes are valid.

Explanation: A GPFS command found that none of the specified nodes passed the required tests.

User response: Determine why the nodes were not accepted, fix the problems, and reissue the command.

6027-1638 *Command*: There are no unassigned nodes in the cluster.

Explanation: A GPFS command in a cluster environment needs unassigned nodes, but found there are none.

User response: Verify whether there are any unassigned nodes in the cluster. If there are none, either add more nodes to the cluster using the **mmaddnode** command, or delete some nodes from the cluster using the **mmdelnode** command, and then reissue the command.

6027-1639 Command failed. Examine previous error messages to determine cause.

Explanation: A GPFS command failed due to previously-reported errors.

User response: Check the previous error messages, fix the problems, and then reissue the command. If no other messages are shown, examine the GPFS log files in the **/var/adm/ras** directory on each node.

6027-1642 *command*: Starting GPFS ...

Explanation: Progress information for the **mmstartup** command.

User response: None. Informational message only.

6027-1643 The number of quorum nodes exceeds the maximum (*number*) allowed.

Explanation: An attempt was made to add more quorum nodes to a cluster than the maximum number allowed.

User response: Reduce the number of quorum nodes, and reissue the command.

6027-1644 Attention: The number of quorum nodes exceeds the suggested maximum (*number*).

Explanation: The number of quorum nodes in the cluster exceeds the maximum suggested number of quorum nodes.

User response: Informational message. Consider reducing the number of quorum nodes to the maximum suggested number of quorum nodes for improved performance.

6027-1645 **Node** *nodeName* **is fenced out from disk** *diskName***.**

Explanation: A GPFS command attempted to access the specified disk, but found that the node attempting the operation was fenced out from the disk.

User response: Check whether there is a valid reason why the node should be fenced out from the disk. If there is no such reason, unfence the disk and reissue the command.

6027-1647 **Unable to find disk with NSD volume id** *NSD volume id***.**

Explanation: A disk with the specified NSD volume id cannot be found.

User response: Specify a correct disk NSD volume id.

6027-1648 **GPFS was unable to obtain a lock from node** *nodeName***.**

Explanation: GPFS failed in its attempt to get a lock from another node in the cluster.

User response: Verify that the reported node is reachable. Examine previous error messages, if any. Fix the problems and then reissue the command.

6027-1661 **Failed while processing disk descriptor** *descriptor* **on node** *nodeName***.**

Explanation: A disk descriptor was found to be unsatisfactory in some way.

User response: Check the preceding messages, if any, and correct the condition that caused the disk descriptor to be rejected.

6027-1662 **Disk device** *deviceName* **refers to an existing NSD** *name*

Explanation: The specified disk device refers to an existing NSD.

User response: Specify another disk that is not an existing NSD.

6027-1663 **Disk descriptor** *descriptor* **should refer to an existing NSD. Use mmcrnsd to create the NSD.**

Explanation: An NSD disk given as input is not known to GPFS.

User response: Create the NSD. Then rerun the command.

6027-1664 *command***: Processing node** *nodeName*

Explanation: Progress information.

User response: None. Informational message only.

6027-1665 **Issue the command from a node that remains in the cluster.**

Explanation: The nature of the requested change requires the command be issued from a node that will remain in the cluster.

User response: Run the command from a node that will remain in the cluster.

6027-1666 [I] **No disks were found.**

Explanation: A command searched for disks but found none.

User response: If disks are desired, create some using the **mmcrnsd** command.

6027-1670 **Incorrect or missing remote shell command:** *name*

Explanation: The specified remote command does not exist or is not executable.

User response: Specify a valid command.

6027-1671 **Incorrect or missing remote file copy command:** *name*

Explanation: The specified remote command does not exist or is not executable.

User response: Specify a valid command.

6027-1672 *option value* **parameter must be an absolute path name.**

Explanation: The mount point does not begin with '/'.

User response: Specify the full path for the mount point.

6027-1674 *command***: Unmounting file systems ...**

Explanation: This message contains progress information about the **mmumount** command.

User response: None. Informational message only.

6027-1677 **Disk** *diskName* **is of an unknown type.**

Explanation: The specified disk is of an unknown type.

User response: Specify a disk whose type is recognized by GPFS.

6027-1680 **Disk name** *diskName* **is already registered for use by GPFS.**

Explanation: The cited disk name was specified for use by GPFS, but there is already a disk by that name registered for use by GPFS.

User response: Specify a different disk name for use by GPFS and reissue the command.

6027-1681 **Node** *nodeName* **is being used as an NSD server.**

Explanation: The specified node is defined as a server node for some disk.

User response: If you are trying to delete the node from the GPFS cluster, you must either delete the disk or define another node as its server.

6027-1685 **Processing continues without lock protection.**

Explanation: The command will continue processing although it was not able to obtain the lock that prevents other GPFS commands from running simultaneously.

User response: Ensure that no other GPFS command is running. See the command documentation for additional details.

6027-1688 **Command was unable to obtain the lock for the GPFS system data. Unable to reach the holder of the lock** *nodeName***. Check the preceding messages, if any. Follow the procedure outlined in the GPFS: Problem Determination Guide.**

Explanation: A command requires the lock for the GPFS system data but was not able to obtain it.

User response: Check the preceding messages, if any. Follow the procedure in the *IBM Spectrum Scale: Problem Determination Guide* for what to do when the GPFS system data is locked. Then reissue the command.

6027-1689 **vpath disk** *diskName* **is not recognized as an IBM SDD device.**

Explanation: The **mmvsdhelper** command found that the specified disk is a vpath disk, but it is not recognized as an IBM SDD device.

User response: Ensure the disk is configured as an IBM SDD device. Then reissue the command.

6027-1699 **Remount failed for file system** *fileSystem***. Error code** *errorCode***.**

Explanation: The specified file system was internally unmounted. An attempt to remount the file system failed with the specified error code.

User response: Check the daemon log for additional error messages. Ensure that all file system disks are available and reissue the **mount** command.

6027-1700 **Failed to load LAPI library.** *functionName* **not found. Changing communication protocol to TCP.**

Explanation: The GPFS daemon failed to load **liblapi_r.a** dynamically.

User response: Verify installation of **liblapi_r.a**.

6027-1701 **mmfsd waiting to connect to mmspsecserver. Setting up to retry every** *number* **seconds for** *number* **minutes.**

Explanation: The GPFS daemon failed to establish a connection with the **mmspsecserver** process.

User response: None. Informational message only.

6027-1702 **Process** *pid* **failed at** *functionName* **call, socket** *socketName***, errno** *value*

Explanation: Either The **mmfsd** daemon or the **mmspsecserver** process failed to create or set up the communication socket between them.

User response: Determine the reason for the error.

6027-1703 **The** *processName* **process encountered error:** *errorString***.**

Explanation: Either the **mmfsd** daemon or the **mmspsecserver** process called the error log routine to log an incident.

User response: None. Informational message only.

6027-1704 **mmspsecserver (pid** *number***) ready for service.**

Explanation: The **mmspsecserver** process has created all the service threads necessary for **mmfsd**.

User response: None. Informational message only.

6027-1705 *command***: incorrect number of connections (***number***), exiting...**

Explanation: The **mmspsecserver** process was called with an incorrect number of connections. This will happen only when the **mmspsecserver** process is run as an independent program.

User response: Retry with a valid number of connections.

6027-1706 mmspsecserver: parent program is not "mmfsd", exiting...

Explanation: The **mmspsecserver** process was invoked from a program other than **mmfsd**.

User response: None. Informational message only.

6027-1707 mmfsd connected to mmspsecserver

Explanation: The **mmfsd** daemon has successfully connected to the **mmspsecserver** process through the communication socket.

User response: None. Informational message only.

6027-1708 The mmfsd daemon failed to fork mmspsecserver. Failure reason *explanation*

Explanation: The **mmfsd** daemon failed to fork a child process.

User response: Check the GPFS installation.

6027-1709 [I] Accepted and connected to *ipAddress*

Explanation: The local **mmfsd** daemon has successfully accepted and connected to a remote daemon.

User response: None. Informational message only.

6027-1710 [N] Connecting to *ipAddress*

Explanation: The local **mmfsd** daemon has started a connection request to a remote daemon.

User response: None. Informational message only.

6027-1711 [I] Connected to *ipAddress*

Explanation: The local **mmfsd** daemon has successfully connected to a remote daemon.

User response: None. Informational message only.

6027-1712 Unexpected zero bytes received from *name***. Continuing.**

Explanation: This is an informational message. A socket read resulted in zero bytes being read.

User response: If this happens frequently, check IP connections.

6027-1715 EINVAL trap from connect call to *ipAddress* **(socket** *name***)**

Explanation: The connect call back to the requesting node failed.

User response: This is caused by a bug in AIX socket support. Upgrade AIX kernel and TCP client support.

6027-1716 [N] Close connection to *ipAddress*

Explanation: Connection socket closed.

User response: None. Informational message only.

6027-1717 [E] Error initializing the configuration server, err *value*

Explanation: The configuration server module could not be initialized due to lack of system resources.

User response: Check system memory.

6027-1718 [E] Could not run command *name***, err** *value*

Explanation: The GPFS daemon failed to run the specified command.

User response: Verify correct installation.

6027-1724 [E] The key used by the cluster named *clusterName* **has changed. Contact the administrator to obtain the new key and register it using "mmremotecluster update".**

Explanation: The administrator of the cluster has changed the key used for authentication.

User response: Contact the administrator to obtain the new key and register it using **mmremotecluster update**.

6027-1725 [E] [E] The key used by the contact node named *contactNodeName* **has changed. Contact the administrator to obtain the new key and register it using `mmauth update`.**

Explanation: The administrator of the cluster has changed the key used for authentication.

User response: Contact the administrator to obtain the new key and register it using **`mmauth update`**.

6027-1726 [E] The administrator of the cluster named *clusterName* **requires authentication. Contact the administrator to obtain the clusters key and register the key using "mmremotecluster update".**

Explanation: The administrator of the cluster requires authentication.

User response: Contact the administrator to obtain the

cluster's key and register it using: **mmremotecluster update**.

6027-1727 [E] The administrator of the cluster named *clusterName* **does not require authentication. Unregister the clusters key using "mmremotecluster update".**

Explanation: The administrator of the cluster does not require authentication.

User response: Unregister the clusters key using: **mmremotecluster update**.

6027-1728 [E] Remote mounts are not enabled within the cluster named *clusterName***. Contact the administrator and request that they enable remote mounts.**

Explanation: The administrator of the cluster has not enabled remote mounts.

User response: Contact the administrator and request remote mount access.

6027-1729 [E] The cluster named *clusterName* **has not authorized this cluster to mount file systems. Contact the cluster administrator and request access.**

Explanation: The administrator of the cluster has not authorized this cluster to mount file systems.

User response: Contact the administrator and request access.

6027-1730 [E] Unsupported cipherList *cipherList* **requested.**

Explanation: The target cluster requested a **cipherList** not supported by the installed version of OpenSSL.

User response: Install a version of OpenSSL that supports the required **cipherList** or contact the administrator of the target cluster and request that a supported **cipherList** be assigned to this remote cluster.

6027-1731 [E] Unsupported cipherList *cipherList* **requested.**

Explanation: The target cluster requested a **cipherList** that is not supported by the installed version of OpenSSL.

User response: Either install a version of OpenSSL that supports the required **cipherList** or contact the administrator of the target cluster and request that a supported **cipherList** be assigned to this remote cluster.

6027-1732 [X] Remote mounts are not enabled within this cluster.

Explanation: Remote mounts cannot be performed in this cluster.

User response: See the *IBM Spectrum Scale: Administration Guide* for instructions about enabling remote mounts. In particular, make sure the keys have been generated and a **cipherlist** has been set.

6027-1733 OpenSSL dynamic lock support could not be loaded.

Explanation: One of the functions required for dynamic lock support was not included in the version of the OpenSSL library that GPFS is configured to use.

User response: If this functionality is required, shut down the daemon, install a version of OpenSSL with the desired functionality, and configure GPFS to use it. Then restart the daemon.

6027-1734 [E] OpenSSL engine support could not be loaded.

Explanation: One of the functions required for engine support was not included in the version of the OpenSSL library that GPFS is configured to use.

User response: If this functionality is required, shut down the daemon, install a version of OpenSSL with the desired functionality, and configure GPFS to use it. Then restart the daemon.

6027-1735 [E] Close connection to *ipAddress***. Attempting reconnect.**

Explanation: Connection socket closed. The GPFS daemon will attempt to reestablish the connection.

User response: None. Informational message only.

6027-1736 [N] Reconnected to *ipAddress*

Explanation: The local **mmfsd** daemon has successfully reconnected to a remote daemon following an unexpected connection break.

User response: None. Informational message only.

6027-1737 [N] Close connection to *ipAddress* **(***errorString***).**

Explanation: Connection socket closed.

User response: None. Informational message only.

6027-1738 [E] Close connection to *ipAddress* **(***errorString***). Attempting reconnect.**

Explanation: Connection socket closed.

User response: None. Informational message only.

6027-1739 [X] Accept socket connection failed: err *value***.**

Explanation: The Accept socket connection received an unexpected error.

User response: None. Informational message only.

6027-1740 [E] Timed out waiting for a reply from node *ipAddress***.**

Explanation: A message that was sent to the specified node did not receive a response within the expected time limit.

User response: None. Informational message only.

6027-1741 [E] Error code *value* **received from node** *ipAddress***.**

Explanation: When a message was sent to the specified node to check its status, an error occurred and the node could not handle the message.

User response: None. Informational message only.

6027-1742 [E] Message ID *value* **was lost by node** *ipAddress***.**

Explanation: During a periodic check of outstanding messages, a problem was detected where the destination node no longer has any knowledge of a particular message.

User response: None. Informational message only.

6027-1743 [W] Failed to load GSKit library *path***: (dlerror)** *errorMessage*

Explanation: The GPFS daemon could not load the library required to secure the node-to-node communications.

User response: Verify that the `gpfs.gskit` package was properly installed.

6027-1744 [I] GSKit library loaded and initialized.

Explanation: The GPFS daemon successfully loaded the library required to secure the node-to-node communications.

User response: None. Informational message only.

6027-1745 [E] Unable to resolve symbol for routine: *functionName* **(dlerror)** *errorMessage*

Explanation: An error occurred while resolving a symbol required for transport-level security.

User response: Verify that the `gpfs.gskit` package was properly installed.

6027-1746 [E] Failed to load or initialize GSKit library: error *value*

Explanation: An error occurred during the initialization of the transport-security code.

User response: Verify that the `gpfs.gskit` package was properly installed.

6027-1747 [W] The TLS handshake with node *ipAddress* **failed with error** *value* **(***handshakeType***).**

Explanation: An error occurred while trying to establish a secure connection with another GPFS node.

User response: Examine the error messages to obtain information about the error. Under normal circumstances, the retry logic will ensure that the connection is re-established. If this error persists, record the error code and contact the IBM Support Center.

6027-1748 [W] A secure receive from node *ipAddress* **failed with error** *value***.**

Explanation: An error occurred while receiving an encrypted message from another GPFS node.

User response: Examine the error messages to obtain information about the error. Under normal circumstances, the retry logic will ensure that the connection is re-established and the message is received. If this error persists, record the error code and contact the IBM Support Center.

6027-1749 [W] A secure send to node *ipAddress* **failed with error** *value***.**

Explanation: An error occurred while sending an encrypted message to another GPFS node.

User response: Examine the error messages to obtain information about the error. Under normal circumstances, the retry logic will ensure that the connection is re-established and the message is sent. If this error persists, record the error code and contact the IBM Support Center.

6027-1750 [N] The *handshakeType* **TLS handshake with node** *ipAddress* **was cancelled: connection reset by peer (return code** *value***).**

Explanation: A secure connection could not be

established because the remote GPFS node closed the connection.

User response: None. Informational message only.

6027-1751 [N] A secure send to node *ipAddress* **was cancelled: connection reset by peer (return code** *value***).**

Explanation: Securely sending a message failed because the remote GPFS node closed the connection.

User response: None. Informational message only.

6027-1752 [N] A secure receive to node *ipAddress* **was cancelled: connection reset by peer (return code** *value***).**

Explanation: Securely receiving a message failed because the remote GPFS node closed the connection.

User response: None. Informational message only.

6027-1753 [E] The crypto library with FIPS support is not available for this architecture. Disable FIPS mode and reattempt the operation.

Explanation: GPFS is operating in FIPS mode, but the initialization of the cryptographic library failed because FIPS mode is not yet supported on this architecture.

User response: Disable FIPS mode and attempt the operation again.

6027-1754 [E] Failed to initialize the crypto library in FIPS mode. Ensure that the crypto library package was correctly installed.

Explanation: GPFS is operating in FIPS mode, but the initialization of the cryptographic library failed.

User response: Ensure that the packages required for encryption are properly installed on each node in the cluster.

6027-1755 [W] The certificate for '*canonicalName***' is expired. Validity period is from** *begDate* **to** *endDate***.**

Explanation: The validity period of the certificate used by a remote node is expired.

User response: Contact the administrator of the remote cluster and instruct them to use the **mmauth** command to generate a new certificate.

6027-1756 [E] The TCP connection to IP address *ipAddress* **(socket** *socketNum***) state is unexpected: ca_state=***caStateValue* **unacked=***unackedCount* **rto=***rtoValue***.**

Explanation: An unexpected TCP socket state has been detected, which may lead to data no longer flowing over the connection. The socket state includes a nonzero *tcpi_ca_state* value, a larger than default retransmission timeout (tcpi_rto) and a nonzero number of currently unacknowledged segments (tcpi_unacked), or a larger than default tcpi_unacked value. All these cases indicate an unexpected TCP socket state, possibly triggered by an outage in the network.

User response: Check network connectivity and whether network packets may have been lost or delayed. Check network interface packet drop statistics.

6027-1757 [E] The TLS handshake with node *ipAddress* **failed with error** *value***: Certificate Signature Algorithm is not supported by the SSL/TLS Handshake (***handshakeType***).**

Explanation: A secure connection could not be established because the signature algorithm of one of the certificates used in the TLS handshake was not supported.

User response: Use the **mmauth** command to generate a new certificate.

6027-1803 [E] Global NSD disk, *name***, not found.**

Explanation: A client tried to open a globally-attached NSD disk, but a scan of all disks failed to find that NSD.

User response: Ensure that the globally-attached disk is available on every node that references it.

6027-1804 [E] I/O to NSD disk, *name***, fails. No such NSD locally found.**

Explanation: A server tried to perform I/O on an NSD disk, but a scan of all disks failed to find that NSD.

User response: Make sure that the NSD disk is accessible to the client. If necessary, break a reservation.

6027-1805 [N] Rediscovered nsd server access to *name***.**

Explanation: A server rediscovered access to the specified disk.

User response: None.

6027-1806 [X] A Persistent Reserve could not be established on device name (*deviceName***):** *errorLine***.**

Explanation: GPFS is using Persistent Reserve on this disk, but was unable to establish a reserve for this node.

User response: Perform disk diagnostics.

6027-1807 [E] NSD *nsdName* **is using Persistent Reserve, this will require an NSD server on an** *osName* **node.**

Explanation: A client tried to open a globally-attached NSD disk, but the disk is using Persistent Reserve. An *osName* NSD server is needed. GPFS only supports Persistent Reserve on certain operating systems.

User response: Use the **mmchnsd** command to add an *osName* NSD server for the NSD.

6027-1808 [A] Unable to reserve space for NSD buffers. Increase pagepool size to at least *requiredPagePoolSize* **MB. Refer to the** *IBM Spectrum Scale: Administration Guide* **for more information on selecting an appropriate pagepool size.**

Explanation: The **pagepool** usage for an NSD buffer (4***maxblocksize**) is limited by factor **nsdBufSpace**. The value of **nsdBufSpace** can be in the range of 10-70. The default value is 30.

User response: Use the **mmchconfig** command to decrease the value of **maxblocksize** or to increase the value of **pagepool** or **nsdBufSpace**.

6027-1809 [E] The defined server *serverName* **for NSD** *NsdName* **couldn't be resolved.**

Explanation: The host name of the NSD server could not be resolved by **gethostbyName()**.

User response: Fix the host name resolution.

6027-1810 [I] Vdisk server recovery: delay *number* **sec. for safe recovery.**

Explanation: Wait for the existing disk lease to expire before performing vdisk server recovery.

User response: None.

6027-1811 [I] Vdisk server recovery: delay complete.

Explanation: Done waiting for existing disk lease to expire before performing vdisk server recovery.

User response: None.

6027-1812 [E] Rediscovery failed for *name***.**

Explanation: A server failed to rediscover access to the specified disk.

User response: Check the disk access issues and run the command again.

6027-1813 [A] Error reading volume identifier (for *objectName name***) from configuration file.**

Explanation: The volume identifier for the named recovery group or vdisk could not be read from the **mmsdrfs** file. This should never occur.

User response: Check for damage to the **mmsdrfs** file.

6027-1814 [E] Vdisk *vdiskName* **cannot be associated with its recovery group** *recoveryGroupName***. This vdisk will be ignored.**

Explanation: The named vdisk cannot be associated with its recovery group.

User response: Check for damage to the **mmsdrfs** file.

6027-1815 [A] Error reading volume identifier (for NSD *name***) from configuration file.**

Explanation: The volume identifier for the named NSD could not be read from the **mmsdrfs** file. This should never occur.

User response: Check for damage to the **mmsdrfs** file.

6027-1816 [E] The defined server *serverName* **for recovery group** *recoveryGroupName* **could not be resolved.**

Explanation: The hostname of the NSD server could not be resolved by **gethostbyName()**.

User response: Fix hostname resolution.

6027-1817 [E] Vdisks are defined, but no recovery groups are defined.

Explanation: There are vdisks defined in the **mmsdrfs** file, but no recovery groups are defined. This should never occur.

User response: Check for damage to the **mmsdrfs** file.

6027-1818 [I] Relinquished recovery group *recoveryGroupName* **(err** *errorCode***).**

Explanation: This node has relinquished serving the named recovery group.

User response: None.

6027-1819 Disk descriptor for *name* **refers to an existing pdisk.**

Explanation: The **mmcrrecoverygroup** command or **mmaddpdisk** command found an existing pdisk.

User response: Correct the input file, or use the **-v** option.

6027-1820 Disk descriptor for *name* refers to an existing NSD.

Explanation: The **mmcrrecoverygroup** command or **mmaddpdisk** command found an existing NSD.

User response: Correct the input file, or use the **-v** option.

6027-1821 Error *errno* writing disk descriptor on *name*.

Explanation: The **mmcrrecoverygroup** command or **mmaddpdisk** command got an error writing the disk descriptor.

User response: Perform disk diagnostics.

6027-1822 Error *errno* reading disk descriptor on *name*.

Explanation: The **tspreparedpdisk** command got an error reading the disk descriptor.

User response: Perform disk diagnostics.

6027-1823 Path error, *name* and *name* are the same disk.

Explanation: The **tspreparedpdisk** command got an error during path verification. The pdisk descriptor file is miscoded.

User response: Correct the pdisk descriptor file and reissue the command.

6027-1824 [X] An unexpected Device Mapper path *dmDevice* (*nsdId*) has been detected. The new path does not have a Persistent Reserve set up. Server disk *diskName* will be put offline

Explanation: A new device mapper path is detected or a previously failed path is activated after the local device discovery has finished. This path lacks a Persistent Reserve, and cannot be used. All device paths must be active at mount time.

User response: Check the paths to all disks making up the file system. Repair any paths to disks which have failed. Rediscover the paths for the NSD.

6027-1825 [A] Unrecoverable NSD checksum error on I/O to NSD disk *nsdName*, using server *serverName*. Exceeds retry limit *number*.

Explanation: The allowed number of retries was exceeded when encountering an NSD checksum error on I/O to the indicated disk, using the indicated server.

User response: There may be network issues that require investigation.

6027-1826 [W] The host name of the server *serverName* that is defined for NSD local cache *NsdName* could not be resolved.

Explanation: The host name of NSD server could not be resolved by **gethostbyName()**.

User response: Fix host name resolution.

6027-1900 Failed to stat *pathName*.

Explanation: A **stat()** call failed for the specified object.

User response: Correct the problem and reissue the command.

6027-1901 *pathName* is not a GPFS file system object.

Explanation: The specified path name does not resolve to an object within a mounted GPFS file system.

User response: Correct the problem and reissue the command.

6027-1902 The policy file cannot be determined.

Explanation: The command was not able to retrieve the policy rules associated with the file system.

User response: Examine the preceding messages and correct the reported problems. Establish a valid policy file with the **mmchpolicy** command or specify a valid policy file on the command line.

6027-1903 *path* must be an absolute path name.

Explanation: The path name did not begin with a **/**.

User response: Specify the absolute path name for the object.

6027-1904 Device with major/minor numbers *number* and *number* already exists.

Explanation: A device with the cited major and minor numbers already exists.

User response: Check the preceding messages for detailed information.

6027-1905 *name* was not created by GPFS or could not be refreshed.

Explanation: The attributes (device type, major/minor number) of the specified file system device name are not as expected.

User response: Check the preceding messages for detailed information on the current and expected values. These errors are most frequently caused by the presence of **/dev** entries that were created outside the GPFS environment. Resolve the conflict by renaming or

deleting the offending entries. Reissue the command letting GPFS create the **/dev** entry with the appropriate parameters.

6027-1906 **There is no file system with drive letter** *driveLetter***.**

Explanation: No file system in the GPFS cluster has the specified drive letter.

User response: Reissue the command with a valid file system.

6027-1908 **The** *option* **option is not allowed for remote file systems.**

Explanation: The specified option can be used only for locally-owned file systems.

User response: Correct the command line and reissue the command.

6027-1909 **There are no available free disks. Disks must be prepared prior to invoking** *command***. Define the disks using the** *command* **command.**

Explanation: The currently executing command (**mmcrfs**, **mmadddisk**, **mmrpldisk**) requires disks to be defined for use by GPFS using one of the GPFS disk creation commands: **mmcrnsd**, **mmcrvsd**.

User response: Create disks and reissue the failing command.

6027-1910 **Node** *nodeName* **is not a quorum node.**

Explanation: The **mmchmgr** command was asked to move the cluster manager to a nonquorum node. Only one of the quorum nodes can be a cluster manager.

User response: Designate the node to be a quorum node, specify a different node on the command line, or allow GPFS to choose the new cluster manager node.

6027-1911 **File system** *fileSystem* **belongs to cluster** *clusterName***. The** *option* **option is not allowed for remote file systems.**

Explanation: The specified option can be used only for locally-owned file systems.

User response: Correct the command line and reissue the command.

6027-1922 **IP aliasing is not supported (***node***). Specify the main device.**

Explanation: IP aliasing is not supported.

User response: Specify a node identifier that resolves to the IP address of a main device for the node.

6027-1927 **The requested disks are not known to GPFS.**

Explanation: GPFS could not find the requested NSDs in the cluster.

User response: Reissue the command, specifying known disks.

6027-1929 *cipherlist* **is not a valid cipher list.**

Explanation: The cipher list must be set to a value supported by GPFS. All nodes in the cluster must support a common cipher.

User response: Use **mmauth show ciphers** to display a list of the supported ciphers.

6027-1930 **Disk** *diskName* **belongs to file system** *fileSystem***.**

Explanation: A GPFS administration command (**mm...**) found that the requested disk to be deleted still belongs to a file system.

User response: Check that the correct disk was requested. If so, delete the disk from the file system before proceeding.

6027-1931 **The following disks are not known to GPFS:** *diskNames***.**

Explanation: A GPFS administration command (**mm...**) found that the specified disks are not known to GPFS.

User response: Verify that the correct disks were requested.

6027-1932 **No disks were specified that could be deleted.**

Explanation: A GPFS administration command (**mm...**) determined that no disks were specified that could be deleted.

User response: Examine the preceding messages, correct the problems, and reissue the command.

6027-1933 **Disk** *diskName* **has been removed from the GPFS cluster configuration data but the NSD volume id was not erased from the disk. To remove the NSD volume id, issue mmdelnsd -p** *NSDvolumeid***.**

Explanation: A GPFS administration command (**mm...**) successfully removed the specified disk from the GPFS cluster configuration data, but was unable to erase the NSD volume id from the disk.

User response: Issue the specified command to remove the NSD volume id from the disk.

6027-1934 **Disk** *diskName* **has been removed from the GPFS cluster configuration data but the NSD volume id was not erased from the disk. To remove the NSD volume id, issue: mmdelnsd -p** *NSDvolumeid* **-N** *nodeList***.**

Explanation: A GPFS administration command (**mm...**) successfully removed the specified disk from the GPFS cluster configuration data but was unable to erase the NSD volume id from the disk.

User response: Issue the specified command to remove the NSD volume id from the disk.

6027-1936 **Node** *nodeName* **cannot support Persistent Reserve on disk** *diskName* **because it is not an AIX node. The disk will be used as a non-PR disk.**

Explanation: A non-AIX node was specified as an NSD server for the disk. The disk will be used as a non-PR disk.

User response: None. Informational message only.

6027-1937 **A node was specified more than once as an NSD server in disk descriptor** *descriptor***.**

Explanation: A node was specified more than once as an NSD server in the disk descriptor shown.

User response: Change the disk descriptor to eliminate any redundancies in the list of NSD servers.

6027-1938 *configParameter* **is an incorrect parameter. Line in error:** *configLine***. The line is ignored; processing continues.**

Explanation: The specified parameter is not valid and will be ignored.

User response: None. Informational message only.

6027-1939 **Line in error:** *line***.**

Explanation: The specified line from a user-provided input file contains errors.

User response: Check the preceding messages for more information. Correct the problems and reissue the command.

6027-1940 **Unable to set reserve policy** *policy* **on disk** *diskName* **on node** *nodeName***.**

Explanation: The specified disk should be able to support Persistent Reserve, but an attempt to set up the registration key failed.

User response: Correct the problem and reissue the command.

6027-1941 **Cannot handle multiple interfaces for host** *hostName***.**

Explanation: Multiple entries were found for the given hostname or IP address either in **/etc/hosts** or by the **host** command.

User response: Make corrections to **/etc/hosts** and reissue the command.

6027-1942 **Unexpected output from the 'host -t a** *name***' command:**

Explanation: A GPFS administration command (**mm...**) received unexpected output from the **host -t a** command for the given host.

User response: Issue the **host -t a** command interactively and carefully review the output, as well as any error messages.

6027-1943 **Host** *name* **not found.**

Explanation: A GPFS administration command (**mm...**) could not resolve a host from **/etc/hosts** or by using the **host** command.

User response: Make corrections to **/etc/hosts** and reissue the command.

6027-1945 **Disk name** *diskName* **is not allowed. Names beginning with gpfs are reserved for use by GPFS.**

Explanation: The cited disk name is not allowed because it begins with **gpfs**.

User response: Specify a disk name that does not begin with **gpfs** and reissue the command.

6027-1947 **Use mmauth genkey to recover the file** *fileName***, or to generate and commit a new key.**

Explanation: The specified file was not found.

User response: Recover the file, or generate a new key by running: **mmauth genkey propagate** or generate a new key by running **mmauth genkey new**, followed by the **mmauth genkey commit** command.

6027-1948 **Disk** *diskName* **is too large.**

Explanation: The specified disk is too large.

User response: Specify a smaller disk and reissue the command.

6027-1949 Propagating the cluster configuration data to all affected nodes.

Explanation: The cluster configuration data is being sent to the rest of the nodes in the cluster.

User response: This is an informational message.

6027-1950 Local update lock is busy.

Explanation: More than one process is attempting to update the GPFS environment at the same time.

User response: Repeat the command. If the problem persists, verify that there are no blocked processes.

6027-1951 Failed to obtain the local environment update lock.

Explanation: GPFS was unable to obtain the local environment update lock for more than 30 seconds.

User response: Examine previous error messages, if any. Correct any problems and reissue the command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-1962 Permission denied for disk *diskName*

Explanation: The user does not have permission to access disk *diskName*.

User response: Correct the permissions and reissue the command.

6027-1963 Disk *diskName* **was not found.**

Explanation: The specified disk was not found.

User response: Specify an existing disk and reissue the command.

6027-1964 I/O error on *diskName*

Explanation: An I/O error occurred on the specified disk.

User response: Check for additional error messages. Check the error log for disk hardware problems.

6027-1967 Disk *diskName* **belongs to back-level file system** *fileSystem* **or the state of the disk is not ready. Use mmchfs -V to convert the file system to the latest format. Use mmchdisk to change the state of a disk.**

Explanation: The specified disk cannot be initialized for use as a tiebreaker disk. Possible reasons are suggested in the message text.

User response: Use the **mmlsfs** and **mmlsdisk** commands to determine what action is needed to correct the problem.

6027-1968 Failed while processing disk *diskName***.**

Explanation: An error was detected while processing the specified disk.

User response: Examine prior messages to determine the reason for the failure. Correct the problem and reissue the command.

6027-1969 Device *device* **already exists on node** *nodeName*

Explanation: This device already exists on the specified node.

User response: None.

6027-1970 Disk *diskName* **has no space for the quorum data structures. Specify a different disk as tiebreaker disk.**

Explanation: There is not enough free space in the file system descriptor for the tiebreaker disk data structures.

User response: Specify a different disk as a tiebreaker disk.

6027-1974 None of the quorum nodes can be reached.

Explanation: Ensure that the quorum nodes in the cluster can be reached. At least one of these nodes is required for the command to succeed.

User response: Ensure that the quorum nodes are available and reissue the command.

6027-1975 The descriptor file contains more than one descriptor.

Explanation: The descriptor file must contain only one descriptor.

User response: Correct the descriptor file.

6027-1976 The descriptor file contains no descriptor.

Explanation: The descriptor file must contain only one descriptor.

User response: Correct the descriptor file.

6027-1977 Failed validating disk *diskName***. Error code** *errorCode***.**

Explanation: GPFS control structures are not as expected.

User response: Contact the IBM Support Center.

6027-1984 **Name** *name* **is not allowed. It is longer than the maximum allowable length (***length***).**

Explanation: The cited name is not allowed because it is longer than the cited maximum allowable length.

User response: Specify a name whose length does not exceed the maximum allowable length, and reissue the command.

6027-1985 **mmfskxload: The format of the GPFS kernel extension is not correct for this version of AIX.**

Explanation: This version of AIX is incompatible with the current format of the GPFS kernel extension.

User response: Contact your system administrator to check the AIX version and GPFS kernel extension.

6027-1986 *junctionName* **does not resolve to a directory in** *deviceName***. The junction must be within the specified file system.**

Explanation: The cited junction path name does not belong to the specified file system.

User response: Correct the junction path name and reissue the command.

6027-1987 **Name** *name* **is not allowed.**

Explanation: The cited name is not allowed because it is a reserved word or a prohibited character.

User response: Specify a different name and reissue the command.

6027-1988 **File system** *fileSystem* **is not mounted.**

Explanation: The cited file system is not currently mounted on this node.

User response: Ensure that the file system is mounted and reissue the command.

6027-1993 **File** *fileName* **either does not exist or has an incorrect format.**

Explanation: The specified file does not exist or has an incorrect format.

User response: Check whether the input file specified actually exists.

6027-1994 **Did not find any match with the input disk address.**

Explanation: The **mmfileid** command returned without finding any disk addresses that match the given input.

User response: None. Informational message only.

6027-1995 **Device** *deviceName* **is not mounted on node** *nodeName***.**

Explanation: The specified device is not mounted on the specified node.

User response: Mount the specified device on the specified node and reissue the command.

6027-1996 **Command was unable to determine whether file system** *fileSystem* **is mounted.**

Explanation: The command was unable to determine whether the cited file system is mounted.

User response: Examine any prior error messages to determine why the command could not determine whether the file system was mounted, resolve the problem if possible, and then reissue the command. If you cannot resolve the problem, reissue the command with the daemon down on all nodes of the cluster. This will ensure that the file system is not mounted, which may allow the command to proceed.

6027-1998 **Line** *lineNumber* **of file** *fileName* **is incorrect:**

Explanation: A line in the specified file passed to the command had incorrect syntax. The line with the incorrect syntax is displayed next, followed by a description of the correct syntax for the line.

User response: Correct the syntax of the line and reissue the command.

6027-1999 **Syntax error. The correct syntax is:** *string***.**

Explanation: The specified input passed to the command has incorrect syntax.

User response: Correct the syntax and reissue the command.

6027-2000 **Could not clear fencing for disk** *physicalDiskName***.**

Explanation: The fencing information on the disk could not be cleared.

User response: Make sure the disk is accessible by this node and retry.

6027-2002 **Disk** *physicalDiskName* **of type** *diskType* **is not supported for fencing.**

Explanation: This disk is not a type that supports fencing.

User response: None.

6027-2004 **None of the specified nodes belong to this GPFS cluster.**

Explanation: The nodes specified do not belong to the GPFS cluster.

User response: Choose nodes that belong to the cluster and try the command again.

6027-2007 **Unable to display fencing for disk** *physicalDiskName***.**

Explanation: Cannot retrieve fencing information for this disk.

User response: Make sure that this node has access to the disk before retrying.

6027-2008 **For the logical volume specification -l** *lvName* **to be valid** *lvName* **must be the only logical volume in the volume group. However, volume group** *vgName* **contains logical volumes.**

Explanation: The command is being run on a logical volume that belongs to a volume group that has more than one logical volume.

User response: Run this command only on a logical volume where it is the only logical volume in the corresponding volume group.

6027-2009 *logicalVolume* **is not a valid logical volume.**

Explanation: *logicalVolume* does not exist in the ODM, implying that logical name does not exist.

User response: Run the command on a valid logical volume.

6027-2010 *vgName* **is not a valid volume group name.**

Explanation: *vgName* passed to the command is not found in the ODM, implying that *vgName* does not exist.

User response: Run the command on a valid volume group name.

6027-2011 **For the hdisk specification -h** *physicalDiskName* **to be valid** *physicalDiskName* **must be the only disk in the volume group. However, volume group** *vgName* **contains disks.**

Explanation: The hdisk specified belongs to a volume group that contains other disks.

User response: Pass an hdisk that belongs to a volume group that contains only this disk.

6027-2012 *physicalDiskName* **is not a valid physical volume name.**

Explanation: The specified name is not a valid physical disk name.

User response: Choose a correct physical disk name and retry the command.

6027-2013 *pvid* **is not a valid physical volume id.**

Explanation: The specified value is not a valid physical volume ID.

User response: Choose a correct physical volume ID and retry the command.

6027-2014 **Node** *node* **does not have access to disk** *physicalDiskName***.**

Explanation: The specified node is not able to access the specified disk.

User response: Choose a different node or disk (or both), and retry the command. If both the node and disk name are correct, make sure that the node has access to the disk.

6027-2015 **Node** *node* **does not hold a reservation for disk** *physicalDiskName***.**

Explanation: The node on which this command is run does not have access to the disk.

User response: Run this command from another node that has access to the disk.

6027-2016 **SSA fencing support is not present on this node.**

Explanation: This node does not support SSA fencing.

User response: None.

6027-2017 **Node ID** *nodeId* **is not a valid SSA node ID. SSA node IDs must be a number in the range of 1 to 128.**

Explanation: You specified a node ID outside of the acceptable range.

User response: Choose a correct node ID and retry the command.

6027-2018 **The SSA node id is not set.**

Explanation: The SSA node ID has not been set.

User response: Set the SSA node ID.

6027-2019 Unable to retrieve the SSA node id.

Explanation: A failure occurred while trying to retrieve the SSA node ID.

User response: None.

6027-2020 Unable to set fencing for disk *physicalDiskName***.**

Explanation: A failure occurred while trying to set fencing for the specified disk.

User response: None.

6027-2021 Unable to clear PR reservations for disk *physicalDiskNam***.**

Explanation: Failed to clear Persistent Reserve information on the disk.

User response: Make sure the disk is accessible by this node before retrying.

6027-2022 Could not open disk *physicalDiskName***, errno** *value***.**

Explanation: The specified disk cannot be opened.

User response: Examine the **errno** value and other messages to determine the reason for the failure. Correct the problem and reissue the command.

6027-2023 retVal = *value***, errno =** *value* **for key** *value***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-2024 ioctl failed with rc=*returnCode***, errno=***errnoValue***. Related values are scsi_status=***scsiStatusValue***, sense_key=***senseKeyValue***, scsi_asc=***scsiAscValue***, scsi_ascq=***scsiAscqValue***.**

Explanation: An **ioctl** call failed with stated return code, **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-2025 READ_KEYS ioctl failed with errno=*returnCode***, tried** *timesTried* **times. Related values are scsi_status=***scsiStatusValue***, sense_key=***senseKeyValue***, scsi_asc=***scsiAscValue***, scsi_ascq=***scsiAscqValue***.**

Explanation: A **READ_KEYS ioctl** call failed with stated **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-2026 READRES ioctl failed with errno=*returnCode***, tried** *timesTried* **times. Related values are: scsi_status=***scsiStatusValue***, sense_key=***senseKeyValue***, scsi_asc=***scsiAscValue***, scsi_ascq=***scsiAscqValue***.**

Explanation: A **REGISTER ioctl** call failed with stated **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-2027 READRES ioctl failed with errno=*returnCode***, tried** *timesTried* **times. Related values are: scsi_status=***scsiStatusValue***, sense_key=***senseKeyValue***, scsi_asc=***scsiAscValue***, scsi_ascq=***scsiAscqValue***.**

Explanation: A **READRES ioctl** call failed with stated **errno** value, and related values.

User response: Check the reported **errno** and correct the problem if possible. Otherwise, contact the IBM Support Center.

6027-2028 could not open disk device *diskDeviceName*

Explanation: A problem occurred on a disk open.

User response: Ensure the disk is accessible and not fenced out, and then reissue the command.

6027-2029 could not close disk device *diskDeviceName*

Explanation: A problem occurred on a disk close.

User response: None.

6027-2030 ioctl failed with DSB=*value* **and result=***value* **reason:** *explanation*

Explanation: An **ioctl** call failed with stated return code, *errno* value, and related values.

User response: Check the reported *errno* and correct the problem, if possible. Otherwise, contact the IBM Support Center.

6027-2031 ioctl failed with non-zero return code

Explanation: An ioctl failed with a non-zero return code.

User response: Correct the problem, if possible. Otherwise, contact the IBM Support Center.

6027-2049 [X] Cannot pin a page pool of size *value* **bytes.**

Explanation: A GPFS page pool cannot be pinned into memory on this machine.

User response: Increase the physical memory size of the machine.

6027-2050 [E] Pagepool has size *actualValue* **bytes instead of the requested** *requestedValue* **bytes.**

Explanation: The configured GPFS page pool is too large to be allocated or pinned into memory on this machine. GPFS will work properly, but with reduced capacity for caching user data.

User response: To prevent this message from being generated when the GPFS daemon starts, reduce the page pool size using the **mmchconfig** command.

6027-2100 Incorrect range *value*-*value* **specified.**

Explanation: The range specified to the command is incorrect. The first parameter value must be less than or equal to the second parameter value.

User response: Correct the address range and reissue the command.

6027-2101 Insufficient free space in *fileSystem* **(***storage* **minimum required).**

Explanation: There is not enough free space in the specified file system or directory for the command to successfully complete.

User response: Correct the problem and reissue the command.

6027-2102 Node *nodeName* **is not mmremotefs to run the command.**

Explanation: The specified node is not available to run a command. Depending on the command, a different node may be tried.

User response: Determine why the specified node is not available and correct the problem.

6027-2103 Directory *dirName* **does not exist**

Explanation: The specified directory does not exist.

User response: Reissue the command specifying an existing directory.

6027-2104 The GPFS release level could not be determined on nodes: *nodeList***.**

Explanation: The command was not able to determine the level of the installed GPFS code on the specified nodes.

User response: Reissue the command after correcting the problem.

6027-2105 The following nodes must be upgraded to GPFS release *productVersion* **or higher:** *nodeList*

Explanation: The command requires that all nodes be at the specified GPFS release level.

User response: Correct the problem and reissue the command.

6027-2106 Ensure the nodes are available and run: *command***.**

Explanation: The command could not complete normally.

User response: Check the preceding messages, correct the problems, and issue the specified command until it completes successfully.

6027-2107 Upgrade the lower release level nodes and run: *command***.**

Explanation: The command could not complete normally.

User response: Check the preceding messages, correct the problems, and issue the specified command until it completes successfully.

6027-2108 Error found while processing stanza

Explanation: A stanza was found to be unsatisfactory in some way.

User response: Check the preceding messages, if any, and correct the condition that caused the stanza to be rejected.

6027-2109 Failed while processing disk stanza on node *nodeName***.**

Explanation: A disk stanza was found to be unsatisfactory in some way.

User response: Check the preceding messages, if any,

and correct the condition that caused the stanza to be rejected.

6027-2110 **Missing required parameter** *parameter*

Explanation: The specified parameter is required for this command.

User response: Specify the missing information and reissue the command.

6027-2111 **The following disks were not deleted:** *diskList*

Explanation: The command could not delete the specified disks. Check the preceding messages for error information.

User response: Correct the problems and reissue the command.

6027-2112 **Permission failure. Option** *option* **requires root authority to run.**

Explanation: The specified command option requires root authority.

User response: Log on as **root** and reissue the command.

6027-2113 **Not able to associate** *diskName* **on node** *nodeName* **with any known GPFS disk.**

Explanation: A command could not find a GPFS disk that matched the specified disk and node values passed as input.

User response: Correct the disk and node values passed as input and reissue the command.

6027-2114 **The** *subsystem* **subsystem is already active.**

Explanation: The user attempted to start a subsystem that was already active.

User response: None. Informational message only.

6027-2115 **Unable to resolve address range for disk** *diskName* **on node** *nodeName***.**

Explanation: A command could not perform address range resolution for the specified disk and node values passed as input.

User response: Correct the disk and node values passed as input and reissue the command.

6027-2116 [E] The GPFS daemon must be active on the recovery group server nodes.

Explanation: The command requires that the GPFS daemon be active on the recovery group server nodes.

User response: Ensure GPFS is running on the recovery group server nodes and reissue the command.

6027-2117 [E] *object name* **already exists.**

Explanation: The user attempted to create an object with a name that already exists.

User response: Correct the name and reissue the command.

6027-2118 [E] The *parameter* **is invalid or missing in the pdisk descriptor.**

Explanation: The pdisk descriptor is not valid. The bad descriptor is displayed following this message.

User response: Correct the input and reissue the command.

6027-2119 [E] Recovery group *name* **not found.**

Explanation: The specified recovery group was not found.

User response: Correct the input and reissue the command.

6027-2120 [E] Unable to delete recovery group *name* **on nodes** *nodeNames***.**

Explanation: The recovery group could not be deleted on the specified nodes.

User response: Perform problem determination.

6027-2121 [I] Recovery group *name* **deleted on node** *nodeName***.**

Explanation: The recovery group has been deleted.

User response: This is an informational message.

6027-2122 [E] The number of spares (*numberOfSpares***) must be less than the number of pdisks (***numberOfpdisks***) being created.**

Explanation: The number of spares specified must be less than the number of pdisks that are being created.

User response: Correct the input and reissue the command.

6027-2123 [E] The GPFS daemon is down on the *vdiskName* servers.

Explanation: The GPFS daemon was down on the vdisk servers when **mmdelvdisk** was issued.

User response: Start the GPFS daemon on the specified nodes and issue the specified **mmdelvdisk** command.

6027-2124 [E] Vdisk *vdiskName* is still NSD *nsdName*. Use the mmdelnsd command.

Explanation: The specified vdisk is still an NSD.

User response: Use the **mmdelnsd** command.

6027-2125 [E] *nsdName* is a vdisk-based NSD and cannot be used as a tiebreaker disk.

Explanation: Vdisk-based NSDs cannot be specified as tiebreaker disks.

User response: Correct the input and reissue the command.

6027-2126 [I] No recovery groups were found.

Explanation: A command searched for recovery groups but found none.

User response: None. Informational message only.

6027-2127 [E] Disk descriptor *descriptor* refers to an existing pdisk.

Explanation: The specified disk descriptor refers to an existing pdisk.

User response: Specify another disk that is not an existing pdisk.

6027-2128 [E] The *attribute* attribute must be configured to use *hostname* as a recovery group server.

Explanation: The specified GPFS configuration attributes must be configured to use the node as a recovery group server.

User response: Use the **mmchconfig** command to set the attributes, then reissue the command.

6027-2129 [E] Vdisk block size (*blockSize*) must match the file system block size (*blockSize*).

Explanation: The specified NSD is a vdisk with a block size that does not match the block size of the file system.

User response: Reissue the command using block sizes that match.

6027-2130 [E] Could not find an active server for recovery group *name*.

Explanation: A command was issued that acts on a recovery group, but no active server was found for the specified recovery group.

User response: Perform problem determination.

6027-2131 [E] Cannot create an NSD on a log vdisk.

Explanation: The specified disk is a log vdisk; it cannot be used for an NSD.

User response: Specify another disk that is not a log vdisk.

6027-2132 [E] Log vdisk *vdiskName* cannot be deleted while there are other vdisks in recovery group *name*.

Explanation: The specified disk is a log vdisk; it must be the last vdisk deleted from the recovery group.

User response: Delete the other vdisks first.

6027-2133 [E] Unable to delete recovery group *name*; vdisks are still defined.

Explanation: Cannot delete a recovery group while there are still vdisks defined.

User response: Delete all the vdisks first.

6027-2134 Node *nodeName* cannot be used as an NSD server for Persistent Reserve disk *diskName* because it is not a Linux node.

Explanation: There was an attempt to enable Persistent Reserve for a disk, but not all of the NSD server nodes are running Linux.

User response: Correct the configuration and enter the command again.

6027-2135 All nodes in the cluster must be running AIX to enable Persistent Reserve for SAN attached disk *diskName*.

Explanation: There was an attempt to enable Persistent Reserve for a SAN-attached disk, but not all nodes in the cluster are running AIX.

User response: Correct the configuration and run the command again.

6027-2136 All NSD server nodes must be running AIX to enable Persistent Reserve for disk *diskName*.

Explanation: There was an attempt to enable Persistent Reserve for the specified disk, but not all NSD servers are running AIX.

User response: Correct the configuration and enter the command again.

6027-2137 An attempt to clear the Persistent Reserve reservations on disk *diskName* **failed.**

Explanation: You are importing a disk into a cluster in which Persistent Reserve is disabled. An attempt to clear the Persistent Reserve reservations on the disk failed.

User response: Correct the configuration and enter the command again.

6027-2138 The cluster must be running either all AIX or all Linux nodes to change Persistent Reserve disk *diskName* **to a SAN-attached disk.**

Explanation: There was an attempt to redefine a Persistent Reserve disk as a SAN attached disk, but not all nodes in the cluster were running either all AIX or all Linux nodes.

User response: Correct the configuration and enter the command again.

6027-2139 NSD server nodes must be running either all AIX or all Linux to enable Persistent Reserve for disk *diskName***.**

Explanation: There was an attempt to enable Persistent Reserve for a disk, but not all NSD server nodes were running all AIX or all Linux nodes.

User response: Correct the configuration and enter the command again.

6027-2140 All NSD server nodes must be running AIX or all running Linux to enable Persistent Reserve for disk *diskName***.**

Explanation: Attempt to enable Persistent Reserve for a disk while not all NSD server nodes are running AIX or all running Linux.

User response: Correct the configuration first.

6027-2141 Disk *diskName* **is not configured as a regular hdisk.**

Explanation: In an AIX only cluster, Persistent Reserve is supported for regular hdisks only.

User response: Correct the configuration and enter the command again.

6027-2142 Disk *diskName* **is not configured as a regular generic disk.**

Explanation: In a Linux only cluster, Persistent Reserve is supported for regular generic or device mapper virtual disks only.

User response: Correct the configuration and enter the command again.

6027-2143 Mount point *mountPoint* **can not be part of automount directory** *automountDir***.**

Explanation: The mount point cannot be the parent directory of the automount directory.

User response: Specify a mount point that is not the parent of the automount directory.

6027-2144 [E] The *lockName* **lock for file system** *fileSystem* **is busy.**

Explanation: More than one process is attempting to obtain the specified lock.

User response: Repeat the command. If the problem persists, verify that there are no blocked processes.

6027-2145 [E] Internal remote command 'mmremote *command*' **no longer supported.**

Explanation: A GPFS administration command invoked an internal remote command which is no longer supported. Backward compatibility for remote commands are only supported for release 3.4 and newer.

User response: All nodes within the cluster must be at release 3.4 or newer. If all the cluster nodes meet this requirement, contact the IBM Support Center.

6027-2147 [E] BlockSize must be specified in disk descriptor.

Explanation: The blockSize positional parameter in a vdisk descriptor was empty. The bad disk descriptor is displayed following this message.

User response: Correct the input and reissue the command.

6027-2148 [E] *nodeName* **is not a valid recovery group server for** *recoveryGroupName***.**

Explanation: The server name specified is not one of the defined recovery group servers.

User response: Correct the input and reissue the command.

6027-2149 [E] Could not get recovery group information from an active server.

Explanation: A command that needed recovery group information failed; the GPFS daemons may have become inactive or the recovery group is temporarily unavailable.

User response: Reissue the command.

6027-2150 The archive system client *backupProgram* could not be found or is not executable.

Explanation: TSM **dsmc** or other specified backup or archive system client could not be found.

User response: Verify that TSM is installed, **dsmc** can be found in the installation location or that the archiver client specified is executable.

6027-2151 The path *directoryPath* is not contained in the snapshot *snapshotName*.

Explanation: The directory path supplied is not contained in the snapshot named with the **-S** parameter.

User response: Correct the directory path or snapshot name supplied, or omit **-S** and the snapshot name in the command.

6027-2152 The path *directoryPath* containing image archives was not found.

Explanation: The directory path supplied does not contain the expected image files to archive into TSM.

User response: Correct the directory path name supplied.

6027-2153 The archiving system *backupProgram* exited with status *return code*. Image backup files have been preserved in *globalWorkDir*

Explanation: Archiving system executed and returned a non-zero exit status due to some error.

User response: Examine archiver log files to discern the cause of the archiver's failure. Archive the preserved image files from the indicated path.

6027-2154 Unable to create a policy file for image backup in *policyFilePath*.

Explanation: A temporary file could not be created in the global shared directory path.

User response: Check or correct the directory path name supplied.

6027-2155 File system *fileSystem* must be mounted read only for restore.

Explanation: The empty file system targeted for restoration must be mounted in read only mode during restoration.

User response: Unmount the file system on all nodes and remount it read only, then try the command again.

6027-2156 The image archive index *ImagePath* could not be found.

Explanation: The archive image index could be found in the specified path

User response: Check command arguments for correct specification of image path, then try the command again.

6027-2157 The image archive index *ImagePath* is corrupt or incomplete.

Explanation: The archive image index specified is damaged.

User response: Check the archive image index file for corruption and remedy.

6027-2158 Disk usage must be dataOnly, metadataOnly, descOnly, dataAndMetadata, vdiskLog, vdiskLogTip, vdiskLogTipBackup, or vdiskLogReserved.

Explanation: The disk usage positional parameter in a vdisk descriptor has a value that is not valid. The bad disk descriptor is displayed following this message.

User response: Correct the input and reissue the command.

6027-2159 [E] *parameter* is not valid or missing in the vdisk descriptor.

Explanation: The vdisk descriptor is not valid. The bad descriptor is displayed following this message.

User response: Correct the input and reissue the command.

6027-2160 [E] Vdisk *vdiskName* is already mapped to NSD *nsdName*.

Explanation: The command cannot create the specified NSD because the underlying vdisk is already mapped to a different NSD.

User response: Correct the input and reissue the command.

6027-2161 [E] NAS servers cannot be specified when creating an NSD on a vdisk.

Explanation: The command cannot create the specified NSD because servers were specified and the underlying disk is a vdisk.

User response: Correct the input and reissue the command.

6027-2162 [E] Cannot set nsdRAIDTracks to zero; *nodeName* is a recovery group server.

Explanation: **nsdRAIDTracks** cannot be set to zero while the node is still a recovery group server.

User response: Modify or delete the recovery group and reissue the command.

6027-2163 [E] Vdisk *name* not found in the daemon. Recovery may be occurring. The disk will not be deleted.

Explanation: GPFS cannot find the specified vdisk. This can happen if recovery is taking place and the recovery group is temporarily inactive.

User response: Reissue the command. If the recovery group is damaged, specify the **-p** option.

6027-2164 [E] Disk descriptor for *name* refers to an existing pdisk.

Explanation: The specified pdisk already exists.

User response: Correct the command invocation and try again.

6027-2165 [E] Node *nodeName* cannot be used as a server of both vdisks and non-vdisk NSDs.

Explanation: The command specified an action that would have caused vdisks and non-vdisk NSDs to be defined on the same server. This is not a supported configuration.

User response: Correct the command invocation and try again.

6027-2166 [E] IBM Spectrum Scale RAID is not configured.

Explanation: IBM Spectrum Scale RAID is not configured on this node.

User response: Reissue the command on the appropriate node.

6027-2167 [E] Device *deviceName* does not exist or is not active on this node.

Explanation: The specified device does not exist or is not active on the node.

User response: Reissue the command on the appropriate node.

6027-2168 [E] The GPFS cluster must be shut down before downloading firmware to port cards.

Explanation: The GPFS daemon must be down on all nodes in the cluster before attempting to download firmware to a port card.

User response: Stop GPFS on all nodes and reissue the command.

6027-2169 Unable to disable Persistent Reserve on the following disks: *diskList*

Explanation: The command was unable to disable Persistent Reserve on the specified disks.

User response: Examine the disks and additional error information to determine if the disks should support Persistent Reserve. Correct the problem and reissue the command.

6027-2170 [E] Recovery group *recoveryGroupName* does not exist or is not active.

Explanation: A command was issued to a recovery group that does not exist or is not in the active state.

User response: Reissue the command with a valid recovery group name or wait for the recovery group to become active.

6027-2171 [E] *objectType objectName* already exists in the cluster.

Explanation: The file system being imported contains an object with a name that conflicts with the name of an existing object in the cluster.

User response: If possible, remove the object with the conflicting name.

6027-2172 [E] Errors encountered while importing IBM Spectrum Scale RAID objects.

Explanation: Errors were encountered while trying to import a IBM Spectrum Scale RAID based file system. No file systems will be imported.

User response: Check the previous error messages and if possible, correct the problems.

6027-2173 [I] Use mmchrecoverygroup to assign and activate servers for the following recovery groups (automatically assigns NSD servers as well): *recoveryGroupList*

Explanation: The **mmimportfs** command imported the specified recovery groups. These must have servers assigned and activated.

User response: After the **mmimportfs** command finishes, use the **mmchrecoverygroup** command to assign NSD server nodes as needed.

6027-2174 Option *option* **can be specified only in conjunction with** *option***.**

Explanation: The cited option cannot be specified by itself.

User response: Correct the input and reissue the command.

6027-2175 [E] Exported path *exportPath* **does not exist**

Explanation: The directory or one of the components in the directory path to be exported does not exist.

User response: Correct the input and reissue the command.

6027-2176 [E] mmchattr for *fileName* **failed.**

Explanation: The command to change the attributes of the file failed.

User response: Check the previous error messages and correct the problems.

6027-2177 [E] Cannot create file *fileName***.**

Explanation: The command to create the specified file failed.

User response: Check the previous error messages and correct the problems.

6027-2178 File *fileName* **does not contain any NSD descriptors or stanzas.**

Explanation: The input file should contain at least one NSD descriptor or stanza.

User response: Correct the input file and reissue the command.

6027-2181 [E] Failover is allowed only for single-writer, independent-writer filesets.

Explanation: The fileset AFM mode is not compatible with the requested operation.

User response: Check the previous error messages and correct the problems.

6027-2182 [E] Resync is allowed only for single-writer filesets.

Explanation: The fileset AFM mode is not compatible with the requested operation.

User response: Check the previous error messages and correct the problems.

6027-2183 [E] Peer snapshots using mmpsnap are allowed only for single-writer or primary filesets.

Explanation: The fileset AFM mode is not compatible with the requested operation.

User response: Check the previous error messages and correct the problems.

6027-2184 [E] If the recovery group is damaged, issue mmdelrecoverygroup *name* **-p.**

Explanation: No active servers were found for the recovery group that is being deleted. If the recovery group is damaged the **-p** option is needed.

User response: Perform diagnosis and reissue the command.

6027-2185 [E] There are no pdisk stanzas in the input file *fileName***.**

Explanation: The **mmcrrecoverygroup** input stanza file has no pdisk stanzas.

User response: Correct the input file and reissue the command.

6027-2186 [E] There were no valid vdisk stanzas in the input file *fileName***.**

Explanation: The **mmcrvdisk** input stanza file has no valid vdisk stanzas.

User response: Correct the input file and reissue the command.

6027-2187 [E] Could not get pdisk information for the following recovery groups: *recoveryGroupList*

Explanation: An **mmlspdisk all** command could not query all of the recovery groups because some nodes could not be reached.

User response: None.

6027-2188 Unable to determine the local node identity.

Explanation: The command is not able to determine the identity of the local node. This can be the result of

a disruption in the network over which the GPFS daemons communicate.

User response: Ensure the GPFS daemon network (as identified in the output of the **mmlscluster** command on a good node) is fully operational and reissue the command.

6027-2189 [E] Action *action* is allowed only for read-only filesets.

Explanation: The specified action is only allowed for read-only filesets.

User response: None.

6027-2190 [E] Cannot prefetch file *fileName*. The file does not belong to fileset *fileset*.

Explanation: The requested file does not belong to the fileset.

User response: None.

6027-2191 [E] Vdisk *vdiskName* not found in recovery group *recoveryGroupName*.

Explanation: The **mmdelvdisk** command was invoked with the **--recovery-group** option to delete one or more vdisks from a specific recovery group. The specified vdisk does not exist in this recovery group.

User response: Correct the input and reissue the command.

6027-2193 [E] Recovery group *recoveryGroupName* must be active on the primary server *serverName*.

Explanation: The recovery group must be active on the specified node.

User response: Use the **mmchrecoverygroup** command to activate the group and reissue the command.

6027-2194 [E] The state of fileset *filesetName* is Expired; prefetch cannot be performed.

Explanation: The prefetch operation cannot be performed on filesets that are in the Expired state.

User response: None.

6027-2195 [E] Error getting snapshot ID for *snapshotName*.

Explanation: The command was unable to obtain the resync snapshot ID.

User response: Examine the preceding messages, correct the problem, and reissue the command. If the problem persists, perform problem determination and contact the IBM Support Center.

6027-2196 [E] Resync is allowed only when the fileset queue is in active state.

Explanation: This operation is allowed only when the fileset queue is in active state.

User response: None.

6027-2197 [E] Empty file encountered when running the mmafmctl flushPending command.

Explanation: The **mmafmctl flushPending** command did not find any entries in the file specified with the **--list-file** option.

User response: Correct the input file and reissue the command.

6027-2198 [E] Cannot run the mmafmctl flushPending command on directory *dirName*.

Explanation: The **mmafmctl flushPending** command cannot be issued on this directory.

User response: Correct the input and reissue the command.

6027-2199 [E] No enclosures were found.

Explanation: A command searched for disk enclosures but none were found.

User response: None.

6027-2200 [E] Cannot have multiple nodes updating firmware for the same enclosure. Enclosure *serialNumber* is already being updated by node *nodeName*.

Explanation: The **mmchenclosure** command was called with multiple nodes updating the same firmware.

User response: Correct the node list and reissue the command.

6027-2201 [E] The mmafmctl flushPending command completed with errors.

Explanation: An error occurred while flushing the queue.

User response: Examine the GPFS log to identify the cause.

6027-2202 [E] There is a SCSI-3 PR reservation on disk *diskname*. mmcrnsd cannot format the disk because the cluster is not configured as PR enabled.

Explanation: The specified disk has a SCSI-3 PR reservation, which prevents the **mmcrnsd** command from formatting it.

User response: Clear the PR reservation by following the instructions in "Clearing a leftover Persistent Reserve reservation" on page 362.

6027-2203 Node *nodeName* is not a gateway node.

Explanation: The specified node is not a gateway node.

User response: Designate the node as a gateway node or specify a different node on the command line.

6027-2204 AFM target map *mapName* is already defined.

Explanation: A request was made to create an AFM target map with the cited name, but that map name is already defined.

User response: Specify a different name for the new AFM target map or first delete the current map definition and then recreate it.

6027-2205 There are no AFM target map definitions.

Explanation: A command searched for AFM target map definitions but found none.

User response: None. Informational message only.

6027-2206 AFM target map *mapName* is not defined.

Explanation: The cited AFM target map name is not known to GPFS.

User response: Specify an AFM target map known to GPFS.

6027-2207 Node *nodeName* is being used as a gateway node for the AFM cluster *clusterName*.

Explanation: The specified node is defined as a gateway node for the specified AFM cluster.

User response: If you are trying to delete the node from the GPFS cluster or delete the gateway node role, you must remove it from the export server map.

6027-2208 [E] *commandName* is already running in the cluster.

Explanation: Only one instance of the specified command is allowed to run.

User response: None.

6027-2209 [E] Unable to list *objectName* on node *nodeName*.

Explanation: A command was unable to list the specific object that was requested.

User response: None.

6027-2210 [E] Unable to build a storage enclosure inventory file on node *nodeName*.

Explanation: A command was unable to build a storage enclosure inventory file. This is a temporary file that is required to complete the requested command.

User response: None.

6027-2211 [E] Error collecting firmware information on node *nodeName*.

Explanation: A command was unable to gather firmware information from the specified node.

User response: Ensure the node is active and retry the command.

6027-2212 [E] Firmware update file *updateFile* was not found.

Explanation: The **mmchfirmware** command could not find the specified firmware update file to load.

User response: Locate the firmware update file and retry the command.

6027-2213 [E] Pdisk path redundancy was lost while updating enclosure firmware.

Explanation: The **mmchfirmware** command lost paths after loading firmware and rebooting the Enclosure Services Module.

User response: Wait a few minutes and then retry the command. GPFS might need to be shut down to finish updating the enclosure firmware.

6027-2214 [E] Timeout waiting for firmware to load.

Explanation: A storage enclosure firmware update was in progress, but the update did not complete within the expected time frame.

User response: Wait a few minutes, and then use the **mmlsfirmware** command to ensure the operation completed.

6027-2215 [E] Storage enclosure *serialNumber* not found.

Explanation: The specified storage enclosure was not found.

User response: None.

6027-2216 Quota management is disabled for file system *fileSystem***.**

Explanation: Quota management is disabled for the specified file system.

User response: Enable quota management for the file system.

6027-2217 [E] Error *errno* **updating firmware for drives** *driveList***.**

Explanation: The firmware load failed for the specified drives. Some of the drives may have been updated.

User response: None.

6027-2218 [E] Storage enclosure *serialNumber* **component** *componentType* **component ID** *componentId* **not found.**

Explanation: The **mmchenclosure** command could not find the component specified for replacement.

User response: Use the **mmlsenclosure** command to determine valid input and then retry the command.

6027-2219 [E] Storage enclosure *serialNumber* **component** *componentType* **component ID** *componentId* **did not fail. Service is not required.**

Explanation: The component specified for the **mmchenclosure** command does not need service.

User response: Use the **mmlsenclosure** command to determine valid input and then retry the command.

6027-2220 [E] Recovery group *name* **has pdisks with missing paths. Consider using the -v no option of the mmchrecoverygroup command.**

Explanation: The **mmchrecoverygroup** command failed because all the servers could not see all the disks, and the primary server is missing paths to disks.

User response: If the disks are cabled correctly, use the **-v no** option of the **mmchrecoverygroup** command.

6027-2221 [E] Error determining redundancy of enclosure *serialNumber* **ESM** *esmName***.**

Explanation: The **mmchrecoverygroup** command failed. Check the following error messages.

User response: Correct the problem and retry the command.

6027-2222 [E] Storage enclosure *serialNumber* **already has a newer firmware version:** *firmwareLevel***.**

Explanation: The **mmchfirmware** command found a newer level of firmware on the specified storage enclosure.

User response: If the intent is to force on the older firmware version, use the **-v no** option.

6027-2223 [E] Storage enclosure *serialNumber* **is not redundant. Shutdown GPFS in the cluster and retry the mmchfirmware command.**

Explanation: The **mmchfirmware** command found a non-redundant storage enclosure. Proceeding could cause loss of data access.

User response: Shutdown GPFS in the cluster and retry the **mmchfirmware** command.

6027-2224 [E] Peer snapshot creation failed. Error code *errorCode***.**

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure there is at least one gateway node configured for the cluster. Examine the preceding messages and the GPFS log for additional details.

User response: Correct the problems and reissue the command.

6027-2225 [E] Peer snapshot successfully deleted at cache. The delete snapshot operation failed at home. Error code *errorCode***.**

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure there is at least one gateway node configured for the cluster. Examine the preceding messages and the GPFS log for additional details.

User response: Correct the problems and reissue the command.

6027-2226 [E] Invalid firmware update file.

Explanation: An invalid firmware update file was specified for the **mmchfirmware** command.

User response: Reissue the command with a valid update file.

6027-2227 [E] Failback is allowed only for independent-writer filesets.

Explanation: Failback operation is allowed only for independent-writer filesets.

User response: Check the fileset mode.

6027-2228 [E] The daemon version (*daemonVersion*) on node *nodeName* is lower than the daemon version (*daemonVersion*) on node *nodeName*.

Explanation: A command was issued that requires nodes to be at specific levels, but the affected GPFS servers are not at compatible levels to support this operation.

User response: Update the GPFS code on the specified servers and retry the command.

6027-2229 [E] Cache Eviction/Prefetch is not allowed for Primary and Secondary mode filesets.

Explanation: Cache eviction/prefetch is not allowed for primary and secondary mode filesets.

User response: None.

6027-2230 [E] afmTarget=*newTargetString* is not allowed. To change the AFM target, use mmafmctl failover with the --target-only option. For primary filesets, use mmafmctl changeSecondary.

Explanation: The **mmchfileset** command cannot be used to change the NFS server or IP address of the home cluster.

User response: To change the AFM target, use the **mmafmctl failover** command and specify the **--target-only** option. To change the AFM target for primary filesets, use the **mmafmctl changeSecondary** command.

6027-2231 [E] The specified block size *blockSize* is smaller than the system page size *pageSize*.

Explanation: The file system block size cannot be smaller than the system memory page size.

User response: Specify a block size greater than or equal to the system memory page size.

6027-2232 [E] Peer snapshots are allowed only for targets using the NFS protocol.

Explanation: The **mmpsnap** command can be used to create snapshots only for filesets that are configured to use the NFS protocol.

User response: Specify a valid fileset target.

6027-2233 [E] Fileset *filesetName* in file system *filesystemName* does not contain peer snapshot *snapshotName*. The delete snapshot operation failed at cache. Error code *errorCode*.

Explanation: The specified snapshot name was not found. The command expects the name of an existing peer snapshot of the active fileset in the specified file system.

User response: Reissue the command with a valid peer snapshot name.

6027-2234 [E] Use the mmafmctl converttoprimary command for converting to primary fileset.

Explanation: Converting to a primary fileset is not allowed directly.

User response: Check the previous error messages and correct the problems.

6027-2235 [E] Only independent filesets can be converted to secondary filesets.

Explanation: Converting to secondary filesets is allowed only for independent filesets.

User response: None.

6027-2236 [E] The CPU architecture on this node does not support tracing in *traceMode* mode. Switching to *traceMode* mode.

Explanation: The CPU does not have constant time stamp counter capability, which is required for overwrite trace mode. The trace has been enabled in blocking mode.

User response: Update the configuration parameters to use the trace facility in blocking mode or replace this node with modern CPU architecture.

6027-2237 [W] An image backup made from the live file system may not be usable for image restore. Specify a valid global snapshot for image backup.

Explanation: The **mmimgbackup** command should always be used with a global snapshot to make a consistent image backup of the file system.

User response: Correct the command invocation to include the **-S** option to specify either a global snapshot name or a directory path that includes the snapshot root directory for the file system and a valid global snapshot name.

6027-2238 [E] Use the mmafmctl convertToSecondary command for converting to secondary.

Explanation: Converting to secondary is allowed by using the **mmafmctl convertToSecondary** command.

User response: None.

6027-2239 [E] Drive serialNumber *serialNumber* **is being managed by server** *nodeName***. Reissue the mmchfirmware command for server** *nodeName***.**

Explanation: The **mmchfirmware** command was issued to update a specific disk drive which is not currently being managed by this node.

User response: Reissue the command specifying the active server.

6027-2240 [E] Option is not supported for a secondary fileset.

Explanation: This option cannot be set for a secondary fileset.

User response: None.

6027-2241 [E] Node *nodeName* **is not a CES node.**

Explanation: A Cluster Export Service command specified a node that is not defined as a CES node.

User response: Reissue the command specifying a CES node.

6027-2242 [E] Error in configuration file.

Explanation: The **mmnfs export load loadCfgFile** command found an error in the NFS configuration files.

User response: Correct the configuration file error.

6027-2245 [E] To change the AFM target, use mmafmctl changeSecondary for the primary.

Explanation: Failover with the **targetonly** option can be run on a primary fileset.

User response: None.

6027-2246 [E] Timeout executing function: *functionName* **(return code=***returnCode***).**

Explanation: The **executeCommandWithTimeout** function was called but it timed out.

User response: Correct the problem and issue the command again.

6027-2247 [E] Creation of *exchangeDir* **failed.**

Explanation: A Cluster Export Service command was unable to create the CCR exchange directory.

User response: Correct the problem and issue the command again.

6027-2248 [E] CCR command failed: *command*

Explanation: A CCR update command failed.

User response: Correct the problem and issue the command again.

6027-2249 [E] Error getting next *nextName* **from CCR.**

Explanation: An expected value from CCR was not obtained.

User response: Issue the command again.

6027-2250 [E] Error putting next *nextName* **to CCR, new ID:** *newExpid* **version:** *version*

Explanation: A CCR value update failed.

User response: Issue the command again.

6027-2251 [E] Error retrieving configuration file: *configFile*

Explanation: Error retrieving configuration file from CCR.

User response: Issue the command again.

6027-2252 [E] Error reading export configuration file (return code: *returnCode***).**

Explanation: A CES command was unable to read the export configuration file.

User response: Correct the problem and issue the command again.

6027-2253 [E] Error creating the internal export data objects (return code *returnCode***).**

Explanation: A CES command was unable to create an export data object.

User response: Correct the problem and issue the command again.

6027-2254 [E] Error creating single export output, export *exportPath* **not found (return code** *returnCode***).**

Explanation: A CES command was unable to create a single export print output.

User response: Correct the problem and reissue the command.

6027-2255 [E] Error creating export output (return code: *returnCode***).**

Explanation: A CES command was unable to create the export print output.

User response: Correct the problem and issue the command again.

6027-2256 [E] Error creating the internal export output file string array (return code: *returnCode*).

Explanation: A CES command was unable to create the array for print output.

User response: Correct the problem and issue the command again.

6027-2257 [E] Error deleting export, export *exportPath* not found (return code: *returnCode*).

Explanation: A CES command was unable to delete an export. The *exportPath* was not found.

User response: Correct the problem and issue the command again.

6027-2258 [E] Error writing export configuration file to CCR (return code: *returnCode*).

Explanation: A CES command was unable to write configuration file to CCR.

User response: Correct the problem and issue the command again.

6027-2259 [E] The path *exportPath* to create the export does not exist (return code:*returnCode*).

Explanation: A CES command was unable to create an export because the path does not exist.

User response: Correct the problem and issue the command again.

6027-2260 [E] The path *exportPath* to create the export is invalid (return code: *returnCode*).

Explanation: A CES command was unable to create an export because the path is invalid.

User response: Correct the problem and issue the command again.

6027-2261 [E] Error creating new export object, invalid data entered (return code: *returnCode*).

Explanation: A CES command was unable to add an export because the input data is invalid.

User response: Correct the problem and issue the command again.

6027-2262 [E] Error creating new export object; getting new export ID (return code: *returnCode*).

Explanation: A CES command was unable to add an export. A new export ID was not obtained.

User response: Correct the problem and issue the command again.

6027-2263 [E] Error adding export; new export path *exportPath* already exists.

Explanation: A CES command was unable to add an export because the path already exists.

User response: Correct the problem and issue the command again.

6027-2264 [E] The --servers option is only used to provide names for primary and backup server configurations. Provide a maximum of two server names.

Explanation: An input node list has too many nodes specified.

User response: Verify the list of nodes and shorten the list to the supported number.

6027-2265 [E] Cannot convert fileset to secondary fileset.

Explanation: Fileset cannot be converted to a secondary fileset.

User response: None.

6027-2266 [E] The snapshot names that start with psnap-rpo or psnap0-rpo are reserved for RPO.

Explanation: The specified snapshot name starts with **psnap-rpo** or **psnap0-rpo**, which are reserved for RPO snapshots.

User response: Use a different snapshot name for the **mmcrsnapshot** command.

6027-2267 [I] Fileset *filesetName* in file system *fileSystem* is either unlinked or being deleted. Home delete-snapshot operation was not queued.

Explanation: The command expects that the peer snapshot at home is not deleted because the fileset at cache is either unlinked or being deleted.

User response: Delete the snapshot at home manually.

6027-2268 [E] This is already a secondary fileset.

Explanation: The fileset is already a secondary fileset.

User response: None.

6027-2269 [E] Adapter *adapterIdentifier* **was not found.**

Explanation: The specified adapter was not found.

User response: Specify an existing adapter and reissue the command.

6027-2270 [E] Error *errno* **updating firmware for adapter** *adapterIdentifier***.**

Explanation: The firmware load failed for the specified adapter.

User response: None.

6027-2271 [E] Error locating the reference client in *inputStringContainingClient* **(return code:** *returnCode***).**

Explanation: The reference client for reordering a client could not be found for the given export path.

User response: Correct the problem and try again.

6027-2272 [E] Error removing the requested client in *inputStringContainingClient* **from a client declaration, return code:** *returnCode*

Explanation: One of the specified clients to remove could not be found in any client declaration for the given export path.

User response: Correct the problem and try again.

6027-2273 [E] Error adding the requested client in *inputStringContainingClient* **to a client declaration, return code:** *returnCode*

Explanation: One of the specified clients to add could not be applied for the given export path.

User response: Correct the problem and try again.

6027-2274 [E] Error locating the reference client in *inputStringContainingClient* **(return code:** *returnCode***).**

Explanation: The reference client for reordering a client could not be applied for the given export path.

User response: Correct the problem and try again.

6027-2275 [E] Unable to determine the status of DASD device *dasdDevice*

Explanation: The **dasdview** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2276 [E] The specified DASD device *dasdDevice* **is not properly formatted. It is not an ECKD-type device, or it has a format other then CDL or LDL, or it has a block size other then 4096.**

Explanation: The specified device is not properly formatted.

User response: Correct the problem and reissue the command.

6027-2277 [E] Unable to determine if DASD device *dasdDevice* **is partitioned.**

Explanation: The **fdasd** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2278 [E] Cannot partition DASD device *dasdDevice***; it is already partitioned.**

Explanation: The specified DASD device is already partitioned.

User response: Remove the existing partitions, or reissue the command using the desired partition name.

6027-2279 [E] Unable to partition DASD device *dasdDevice*

Explanation: The **fdasd** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2280 [E] The DASD device with bus ID *busID* **cannot be found or it is in use.**

Explanation: The **chccwdev** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2281 [E] Error *errno* **updating firmware for enclosure** *enclosureIdentifier***.**

Explanation: The firmware load failed for the specified enclosure.

User response: None.

6027-2282 [E] Action *action* **is not allowed for secondary filesets.**

Explanation: The specified action is not allowed for secondary filesets.

User response: None.

6027-2283 [E] Node *nodeName* **is already a CES node.**

Explanation: An **mmchnode** command attempted to enable CES services on a node that is already part of the CES cluster.

User response: Reissue the command specifying a node that is not a CES node.

6027-2284 [E] The fileset afmshowhomesnapshot value is 'yes'. The fileset mode cannot be changed.

Explanation: The fileset **afmshowhomesnapshot** attribute value is **yes**. The fileset mode change is not allowed.

User response: First change the attribute **afmshowhomesnapshot** value to **no**, and then issue the command again to change the mode.

6027-2285 [E] Deletion of initial snapshot *snapshotName* **of fileset** *filesetName* **in file system** *fileSystem* **failed. The delete fileset operation failed at cache. Error code** *errorCode***.**

Explanation: The deletion of the initial snapshot **psnap0** of *filesetName* failed. The primary and secondary filesets cannot be deleted without deleting the initial snapshot.

User response: None.

6027-2286 [E] RPO peer snapshots using mmpsnap are allowed only for primary filesets.

Explanation: RPO snapshots can be created only for primary filesets.

User response: Reissue the command with a valid primary fileset or without the **--rpo** option.

6027-2287 The fileset needs to be linked to change afmShowHomeSnapshot to 'no'.

Explanation: The **afmShowHomeSnapshot** value cannot be changed to **no** if the fileset is unlinked.

User response: Link the fileset and reissue the command.

6027-2288 [E] Option *optionName* **is not supported for AFM filesets.**

Explanation: IAM modes are not supported for AFM filesets.

User response: None.

6027-2289 [E] Peer snapshot creation failed while running *subCommand***. Error code** *errorCode*

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure there is at least one gateway node configured for the cluster. Examine the preceding messages and the GPFS log for additional details.

User response: Correct the problems and reissue the command.

6027-2290 [E] The comment string should be less than 50 characters long.

Explanation: The comment/prefix string of the snapshot is longer than 50 characters.

User response: Reduce the comment string size and reissue the command.

6027-2291 [E] Peer snapshot creation failed while generating snapshot name. Error code *errorCode*

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure there is at least one gateway node configured for the cluster. Examine the preceding messages and the GPFS log for additional details.

User response: Correct the problems and reissue the command.

6027-2292 [E] The initial snapshot *psnap0Name* **does not exist. The peer snapshot creation failed. Error code** *errorCode*

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure the initial peer snapshot exists for the fileset. Examine the preceding messages and the GPFS log for additional details.

User response: Verify that the fileset is a primary fileset and that it has **psnap0** created and try again.

6027-2293 [E] The peer snapshot creation failed because fileset *filesetName* **is in** *filesetState* **state.**

Explanation: For an active fileset, check the AFM target configuration for peer snapshots. Ensure there is at least one gateway node configured for the cluster. Examine the preceding messages and the GPFS log for additional details.

User response: None. The fileset needs to be in active or dirty state.

6027-2294 [E] Removing older peer snapshots failed while obtaining snap IDs. Error code *errorCode*

Explanation: Ensure the fileset exists. Examine the preceding messages and the GPFS log for additional details.

User response: Verify that snapshots exist for the given fileset.

6027-2295 [E] Removing older peer snapshots failed while obtaining old snap IDs. Error code *errorCode*

Explanation: Ensure the fileset exists. Examine the preceding messages and the GPFS log for additional details.

User response: Verify that snapshots exist for the given fileset.

6027-2296 [E] Need a target to convert to the primary fileset.

Explanation: Need a target to convert to the primary fileset.

User response: Specify a target to convert to the primary fileset.

6027-2297 [E] The check-metadata and nocheck-metadata options are not supported for a non-AFM fileset.

Explanation: The **check-metadata** and **nocheck-metadata** options are not supported for a non-AFM fileset.

User response: None.

6027-2298 [E] Only independent filesets can be converted to primary or secondary.

Explanation: Only independent filesets can be converted to primary or secondary.

User response: Specify an independent fileset.

6027-2299 [E] Issue the `mmafmctl getstate` command to check fileset state and if required issue `mmafmctl convertToPrimary`.

Explanation: Issue the **`mmafmctl getstate`** command to check fileset state and if required issue **`mmafmctl convertToPrimary`**.

User response: Issue the **`mmafmctl getstate`** command to check fileset state and if required issue **`mmafmctl convertToPrimary`**.

6027-2300 [E] The check-metadata and nocheck-metadata options are not supported for the primary fileset.

Explanation: The **check-metadata** and **nocheck-metadata** options are not supported for the primary fileset.

User response: None.

6027-2301 [E] The inband option is not supported for the primary fileset.

Explanation: The inband option is not supported for the primary fileset.

User response: None.

6027-2302 [E] AFM target cannot be changed for the primary fileset.

Explanation: AFM target cannot be changed for the primary fileset.

User response: None.

6027-2303 [E] The inband option is not supported for an AFM fileset.

Explanation: The inband option is not supported for an AFM fileset.

User response: None.

6027-2304 [E] Target cannot be changed for an AFM fileset.

Explanation: Target cannot be changed for an AFM fileset.

User response: None.

6027-2305 [E] The mmafmctl convertToPrimary command is not allowed for this primary fileset.

Explanation: The **mmafmctl convertToPrimary** command is not allowed for the primary fileset because it is not in **PrimInitFail** state.

User response: None.

6027-2306 [E] Failed to check for cached files while doing primary conversion from *filesetMode* **mode.**

Explanation: Failed to check for cached files while doing primary conversion.

User response: None.

6027-2307 [E] Uncached files present, run prefetch first.

Explanation: Uncached files present.

User response: Run prefetch and then do the conversion.

6027-2308 [E] Uncached files present, run prefetch first using policy output: *nodeDirFileOut***.**

Explanation: Uncached files present.

User response: Run prefetch first using policy output.

6027-2309 [E] Conversion to primary not allowed for *filesetMode* **mode.**

Explanation: Conversion to primary not allowed for this mode.

User response: None.

6027-2310 [E] This option is available only for a primary fileset.

Explanation: This option is available only for a primary fileset.

User response: None.

6027-2311 [E] The target-only option is not allowed for a promoted primary without a target.

Explanation: The **target-only** option is not allowed for a promoted primary without a target.

User response: None.

6027-2312 [E] Need a target to setup the new secondary.

Explanation: Target is required to setup the new secondary.

User response: None.

6027-2313 [E] The target-only and inband options are not allowed together.

Explanation: The **target-only** and **inband** options are not allowed together.

User response: None.

6027-2314 [E] Could not run *commandName***. Verify that the Object protocol was installed.**

Explanation: The **mmcesobjlscfg** command cannot find a prerequisite command on the system.

User response: Install the missing command and try again.

6027-2315 [E] Could not determine CCR file for service *serviceName*

Explanation: For the given service name, there is not a corresponding file in the CCR.

User response: None.

6027-2316 [E] Unable to retrieve file *fileName* **from CCR using** *command* **command. Verify that the Object protocol is correctly installed.**

Explanation: There was an error downloading a file from the CCR repository.

User response: Correct the error and try again.

6027-2317 [E] Unable to parse version number of file *fileName* **from mmccr output**

Explanation: The current version should be printed by **mmccr** when a file is extracted. The command could not read the version number from the output and failed.

User response: Investigate the failure in the CCR and fix the problem.

6027-2318 [E] Could not put *localFilePath* **into the CCR as** *ccrName*

Explanation: There was an error when trying to do an **fput** of a file into the CCR.

User response: Investigate the error and fix the problem.

6027-2319 [I] Version mismatch during upload of *fileName* **(***version***). Retrying.**

Explanation: The file could not be uploaded to the CCR because another process updated it in the meantime. The file will be downloaded, modified, and uploaded again.

User response: None. The upload will automatically be tried again.

6027-2320 *directoryName* **does not resolve to a directory in** *deviceName***. The directory must be within the specified file system.**

Explanation: The cited directory does not belong to the specified file system.

User response: Correct the directory name and reissue the command.

6027-2321 [E] AFM primary or secondary filesets cannot be created for file system *fileSystem* because version is less than *supportedVersion*.

Explanation: The AFM primary or secondary filesets are not supported for a file system version that is less than 14.20.

User response: Upgrade the file system and reissue the command.

6027-2322 [E] The OBJ service cannot be enabled because it is not installed. The file *fileName* was not found.

Explanation: The node could not enable the CES OBJ service because of a missing binary or configuration file.

User response: Install the required software and retry the command.

6027-2323 [E] The OBJ service cannot be enabled because the number of CES IPs below the minimum of *minValue* expected.

Explanation: The value of CES IPs was below the minimum.

User response: Add at least *minValue* CES IPs to the cluster.

6027-2324 [E] The object store for *serviceName* is either not a GPFS type or *mountPoint* does not exist.

Explanation: The object store is not available at this time.

User response: Verify that *serviceName* is a GPFS type. Verify that the *mountPoint* exists, the file system is mounted, or the fileset is linked.

6027-2325 [E] File *fileName* does not exist in CCR. Verify that the Object protocol is correctly installed.

Explanation: There was an error verifying Object config and ring files in the CCR repository.

User response: Correct the error and try again.

6027-2326 [E] The OBJ service cannot be enabled because attribute *attributeName* for a CES IP has not been defined. Verify that the Object protocol is correctly installed.

Explanation: There was an error verifying *attributeName* on CES IPs.

User response: Correct the error and try again.

6027-2327 The snapshot *snapshotName* is the wrong scope for use in *targetType* backup

Explanation: The snapshot specified is the wrong scope.

User response: Please provide a valid snapshot name for this backup type.

6027-2329 [E] The fileset attributes cannot be set for the primary fileset with caching disabled.

Explanation: The fileset attributes cannot be set for the primary fileset with caching disabled.

User response: None.

6027-2330 [E] The outband option is not supported for AFM filesets.

Explanation: The outband option is not supported for AFM filesets.

User response: None.

6027-2331 [E] CCR value *ccrValue* not defined. The OBJ service cannot be enabled if identity authentication is not configured.

Explanation: Object authentication type was not found.

User response: Configure identity authentication and try again.

6027-2332 [E] Only regular independent filesets are converted to secondary filesets.

Explanation: Only regular independent filesets can be converted to secondary filesets.

User response: Specify a regular independent fileset and run the command again.

6027-2333 [E] Failed to disable *serviceName* service. Ensure *authType* authentication is removed.

Explanation: Disable CES service failed because authentication was not removed.

User response: Remove authentication and retry.

6027-2334 [E] Fileset *indFileset* cannot be changed because it has a dependent fileset *depFileset*

Explanation: Filesets with dependent filesets cannot be converted to primary or secondary.

User response: This operation cannot proceed until all the dependent filesets are unlinked.

6027-2335 [E] Failed to convert fileset, because the policy to detect special files is failing.

Explanation: The policy to detect special files is failing.

User response: Retry the command later.

6027-2336 [E] Immutable/append-only files or clones copied from a snapshot are present, hence conversion is disallowed

Explanation: Conversion is disallowed if immutable/append-only files or clones copied from a snapshot are present.

User response: Files should not be immutable/append-only.

6027-2337 [E] Conversion to primary is not allowed at this time. Retry the command later.

Explanation: Conversion to primary is not allowed at this time.

User response: Retry the command later.

6027-2338 [E] Conversion to primary is not allowed because the state of the fileset is *filesetState***.**

Explanation: Conversion to primary is not allowed with the current state of the fileset.

User response: Retry the command later.

6027-2339 [E] Orphans are present, run prefetch first.

Explanation: Orphans are present.

User response: Run prefetch on the fileset and then do the conversion.

6027-2340 [E] Fileset was left in PrimInitFail state. Take the necessary actions.

Explanation: The fileset was left in **PrimInitFail** state.

User response: Take the necessary actions.

6027-2341 [E] This operation can be done only on a primary fileset

Explanation: This is not a primary fileset.

User response: None.

6027-2342 [E] Failover/resync is currently running so conversion is not allowed

Explanation: Failover/resync is currently running so conversion is not allowed.

User response: Retry the command later after failover/resync completes.

6027-2343 [E] DR Setup cannot be done on a fileset with mode *filesetMode***.**

Explanation: Setup cannot be done on a fileset with this mode.

User response: None.

6027-2344 [E] The GPFS daemon must be active on the node from which the *mmcmd* **is executed with option --inode-criteria or -o.**

Explanation: The GPFS daemon needs to be active on the node where the command is issued with **--inode-criteria** or **-o** options.

User response: Run the command where the daemon is active.

6027-2345 [E] The provided snapshot name must be unique to list filesets in a specific snapshot

Explanation: The **mmlsfileset** command received a snapshot name that is not unique.

User response: Correct the command invocation or remove the duplicate named snapshots and try again.

6027-2346 [E] The local node is not a CES node.

Explanation: A local Cluster Export Service command was invoked on a node that is not defined as a Cluster Export Service node.

User response: Reissue the command on a CES node.

6027-2347 [E] Error changing export, export *exportPath* **not found.**

Explanation: A CES command was unable to change an export. The *exportPath* was not found.

User response: Correct problem and issue the command again.

6027-2348 [E] A device for *directoryName* **does not exist or is not active on this node.**

Explanation: The device containing the specified directory does not exist or is not active on the node.

User response: Reissue the command with a correct directory or on an appropriate node.

6027-2349 [E] The fileset for *junctionName* does not exist in the *targetType* specified.

Explanation: The fileset to back up cannot be found in the file system or snapshot specified.

User response: Reissue the command with a correct name for the fileset, snapshot, or file system.

6027-2350 [E] The fileset for *junctionName* is not linked in the *targetType* specified.

Explanation: The fileset to back up is not linked in the file system or snapshot specified.

User response: Relink the fileset in the file system. Optionally create a snapshot and reissue the command with a correct name for the fileset, snapshot, and file system.

6027-2351 [E] One or more unlinked filesets (*filesetNames*) exist in the *targetType* specified. Check your filesets and try again.

Explanation: The file system to back up contains one or more filesets that are unlinked in the file system or snapshot specified.

User response: Relink the fileset in the file system. Optionally create a snapshot and reissue the command with a correct name for the fileset, snapshot, and file system.

6027-2352 The snapshot *snapshotName* could not be found for use by *commandName*

Explanation: The snapshot specified could not be located.

User response: Please provide a valid snapshot name.

6027-2353 [E] The snapshot name cannot be generated.

Explanation: The snapshot name cannot be generated.

User response: None.

6027-2354 Node *nodeName* must be disabled as a CES node before trying to remove it from the GPFS cluster.

Explanation: The specified node is defined as a CES node.

User response: Disable the CES node and try again.

6027-2355 [E] Unable to reload *moduleName*. Node *hostname* should be rebooted.

Explanation: Host adapter firmware was updated so the specified module needs to be unloaded and reloaded. Linux does not display the new firmware level until the module is reloaded.

User response: Reboot the node.

6027-2356 [E] Node *nodeName* is being used as a recovery group server.

Explanation: The specified node is defined as a server node for some disk.

User response: If you are trying to delete the node from the GPFS cluster, you must either delete the disk or define another node as its server.

6027-2357 [E] Root fileset cannot be converted to primary fileset.

Explanation: Root fileset cannot be converted to the primary fileset.

User response: None.

6027-2358 [E] Root fileset cannot be converted to secondary fileset.

Explanation: Root fileset cannot be converted to the secondary fileset.

User response: None.

6027-2359 [I] Attention: *command* is now enabled. This attribute can no longer be modified.

Explanation: Indefinite retention protection is enabled. This value can not be changed in the future.

User response: None.

6027-2360 [E] The current value of *command* is *attrName*. This value cannot be changed.

Explanation: Indefinite retention protection is enabled for this cluster and this attribute cannot be changed.

User response: None.

6027-2361 [E] *command* is enabled. File systems cannot be deleted.

Explanation: When indefinite retention protection is enabled the file systems cannot be deleted.

User response: None.

6027-2362 [E] The current value of *command* is *attrName*. No changes made.

Explanation: The current value and the request value are the same. No changes made.

User response: None.

6027-2363 [E] Operation is not permitted as state of the fileset is *filesetState*.

Explanation: This operation is not allowed with the current state of the fileset.

User response: Retry the command later.

6027-2364 [E] Fileset name is missing.

Explanation: This operation needs to be run for a particular fileset.

User response: Retry the command with a fileset name.

6027-2365 [E] Firmware loader *filename* not executable.

Explanation: The listed firmware loader is not executable.

User response: Make the firmware loader executable and retry the command.

6027-2366 Node *nodeName* is being used as an NSD server. This may include Local Read Only Cache (LROC) storage. Review these details and determine the NSD type by running the `mmlsnsd` command. For standard NSDs, you must either delete the disk or define another node as its server. For nodes that include LROC NSDs (local cache) must have all the LROC NSDs removed before the node can be deleted. Fully review the `mmdelnsd` command documentation before making any changes.

Explanation: The specified node is defined as a server node for some disk.

User response: If you are trying to delete the node from the GPFS cluster, you must either delete the disk or define another node as its server.

6027-2367 [E] Fileset having *iammode* mode cannot be converted to primary fileset.

Explanation: Fileset with Integrated Archive Manager (IAM) mode cannot be converted to primary fileset.

User response: None.

6027-2368 [E] Unable to find information for Hypervisor.

Explanation: The **`lscpu`** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2369 [E] Unable to list DASD devices

Explanation: The **`lsdasd`** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2370 [E] Unable to flush buffer for DASD device *name1*

Explanation: The **`blockdev --flushbufs`** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2371 [E] Unable to read the partition table for DASD device *dasdDevice*.

Explanation: The **`blockdev --rereadpt`** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2372 [E] Unable to find information to DASD device *dasdDevice*.

Explanation: The **`dasdinfo`** command failed.

User response: Examine the preceding messages, correct the problem, and reissue the command.

6027-2373 *feature* is only available in the IBM Spectrum Scale Advanced Edition.

Explanation: The specified function or feature is only part of the IBM Spectrum Scale Advanced Edition.

User response: Install the IBM Spectrum Scale Advanced Edition on all nodes in the cluster, and then reissue the command.

6027-2374 [E] Unable to delete recovery group *name*; as the associated VDisk sets are still defined.

Explanation: Cannot delete a recovery group when vdisk sets are still associated with it.

User response: Delete all the associated vdisk sets before deleting the recovery group.

6027-2376 [E] Node class *nodeclass* cannot be *action*. It is marked for use by Transparent Cloud Tiering. To remove this node class, first disable all the nodes with `mmchnode --cloud-gateway-disable`.

Explanation: Cannot delete a node class that has cloud gateway enabled.

User response: Disable the nodes first with **`mmchnode --cloud-gateway-disable`**.

6027-2377 [E] Node *nodeclass* **cannot be deleted. It is marked for use by Transparent Cloud Tiering. To remove this node, first disable it with `mmchnode --cloud-gateway-disable`.**

Explanation: Cannot delete a node that has cloud gateway enabled.

User response: Disable the node first with **`mmchnode --cloud-gateway-disable`**.

6027-2378 [E] To enable Transparent Cloud Tiering nodes, you must first enable the Transparent\n\ Cloud Tiering feature. This feature provides a new level of storage tiering capability to the IBM Spectrum Scale customer. Please contact your IBM Client Technical Specialist (or send an email to scale@us.ibm.com) to review your use case of the Transparent Cloud Tiering feature and to obtain the instructions to enable the feature in your environment.

Explanation: The Transparent Cloud Tiering feature must be enabled with assistance from IBM.

User response: Contact IBM support for more information.

6027-2379 [E] The FBA-type DASD device *dasdDevice* **is not a partition.**

Explanation: The FBA-type DASD device has to be a partition.

User response: Reissue the command using the desired partition name.

6027-2380 [E] Support for FBA-type DASD device is not enabled. Run `mmchconfig release=LATEST` to activate the new function.

Explanation: FBA-type DASD must be supported in the entire cluster.

User response: Verify the IBM Spectrum Scale level on all nodes, update to the required level to support FBA by using the **`mmchconfig release=LATEST`** command, and reissue the command.

6027-2381 [E] Missing argument *missingArg*

Explanation: An IBM Spectrum Scale administration command received an insufficient number of arguments.

User response: Correct the command line and reissue the command.

6027-2382 [E] Conversion is not allowed for filesets with active clone files.

Explanation: Conversion is disallowed if clones are present.

User response: Remove the clones and try again.

6027-2383 [E] Conversion to secondary fileset has failed.

Explanation: Fileset could not be converted to secondary.

User response: Run the **`mmafmctl convertToSecondary`** command again.

6027-2384 [E] No object storage policy found.

Explanation: Error while retrieving object storage policies.

User response: Verify if object protocol is enabled on all nodes, and reissue the command.

6027-2385 [E] Failed to create soft link between directories: *directoryName1*, *directoryName2***.**

Explanation: Error while creating soft link between provided fileset path and container path.

User response: Examine the command output to determine the root cause.

6027-2386 [E] Provided fileset path *filesetPath* **is already enabled for objectization.**

Explanation: The provided fileset path is already enabled for objectization.

User response: Retry using different fileset path.

6027-2387 [E] Provided container *containerName* **is already enabled for objectization.**

Explanation: The provided container is already enabled for objectization.

User response: Retry using a different container name.

6027-2388 [E] Given fileset: *filesetName* **is not part of object file system:** *fileSystemName***.**

Explanation: Provided fileset is derived from a non object file system.

User response: Retry using the fileset which is derived from object file system.

6027-2389 [E] Fileset path is already used by object protocol. It cannot be selected for objectization.

Explanation: The provided fileset path is already in use by the object protocol.

User response: Retry using a different fileset path.

6027-2390 [E] SELinux needs to be in either disabled or permissive mode.

Explanation: The command validates SELinux state.

User response: Retry with SELinux in disabled mode.

6027-2391 [E] The configuration of SED based encryption for the drive '*name1*' is failed.

Explanation: The enrollment of SED drive for SED based encryption is failed.

User response: Rerun the command after fixing the drive.

6027-2392 [E] Found pdisk *serialNumber* in recovery group *recoverygroupName* has *pdiskName* paths.

Explanation: The **mmchfirmware** command found a non-redundant pdisk. Proceeding could cause loss of data access.

User response: Shutdown GPFS in the cluster and retry the **mmchfirmware** command.

6027-2393 [E] Use the -N parameter to specify the nodes that have access to the hardware to be updated.

Explanation: The **mmchfirmware** command was issued to update firmware, but no devices were found on the specified nodes.

User response: Reissue the command with the **-N** parameter.

6027-2394 [E] No drive serial number was found for *driveName*.

Explanation: The **mmchcarrier** command was unable to determine the drive serial number for the replacement drive.

User response: Contact the IBM Support Center.

6027-2395 The *feature* is only available in the IBM Spectrum Scale Advanced Edition or the Data Management Edition.

Explanation: The specified function or feature is only part of the IBM Spectrum Scale Advanced Edition or the Data Management Edition.

User response: Install the IBM Spectrum Scale Advanced Edition or the Data Management Edition on all nodes in the cluster, and then reissue the command.

6027-2396 [E] Failed to verify the presence of uncached files while disabling AFM.

Explanation: Failed to verify uncached files because the **mmapplypolicy** command that is run internally as part of disabling AFM is failing.

User response: Rerun the command.

6027-2397 [E] Orphans are present, run prefetch first by using the policy output: *nodeDirFileOut*

Explanation: Orphans are present.

User response: Run prefetch first by using the policy output.

6027-2398 [E] Fileset is unlinked. Link the fileset and then rerun the command, so that the uncached files and orphans can be verified.

Explanation: Fileset is unlinked and policy cannot be run on it to verify the uncached files and orphans.

User response: Link the fileset first and then retry the command.

6027-2399 [E] The mmchfirmware command cannot be completed due to mixed code levels on the recovery group servers.

Explanation: The **mmchfirmware** command discovered incompatible code levels on the recovery group servers.

User response: Update the code levels on the recovery group servers and try again.

6027-2400 At least one node in the cluster must be defined as an admin node.

Explanation: All nodes were explicitly designated or allowed to default to non-admin. At least one node must be designated as admin nodes.

User response: Specify which of the nodes must be considered as an admin node and then reissue the command.

6027-2401 [E] This cluster node is not designated as an admin node. Commands are allowed to only run on the admin nodes.

Explanation: Only nodes that are designated as admin nodes are allowed to run commands. The node where the command was attempted run is not admin node.

User response: Use the **mmlscluster** command to

| identify the admin nodes where commands can run.
| Use the **mmchnode** command to designate admin nodes
| in the cluster.

6027-2402 Missing option: *MissingOption***.**

Explanation: A GPFS administrative command is missing a required option.

User response: Correct the command line and reissue the command.

6027-2403 Invalid argument for *MissingOption1* **option:** *MissingOption2*

Explanation: A GPFS administrative command option is invalid.

User response: Correct the command line and reissue the command.

6027-2404 [E] No NVMe devices are in use by GNR.

Explanation: Either no NVMe devices exist or GNR is not using any NVMe devices.

User response: Verify whether recovery groups exist and configured to use NVMe devices.

6027-2405 [E] The NVMe device *NVMeDeviceName* **is not in use by GNR.**

Explanation: Either the specified NVMe device does not exist or GNR is not configured to use the specified NVMe device.

User response: Verify whether recovery groups exist and configured to use the specified NVMe device.

6027-2406 [E] The recovery group servers are not found.

Explanation: There are no recovery groups configured.

User response: Use **-N** option to specify node names or node class.

6027-2407 [E] Vdisk (*vdiskName***) with block size (***blockSize***) cannot be allocated in storage pool (***storagePoolName***) with block size (***blockSize***).**

Explanation: The **mmcrfs** command specified an NSD but its underlying vdisk block size does not match the storage pool block size.

User response: Reissue the command using block sizes that match.

6027-2500 mmsanrepairfs already in progress for "*name***"**

Explanation: This is an output from **mmsanrepairfs** when another **mmsanrepairfs** command is already running.

User response: Wait for the currently running command to complete and reissue the command.

6027-2501 Could not allocate storage.

Explanation: Sufficient memory could not be allocated to run the **mmsanrepairfs** command.

User response: Increase the amount of memory available.

6027-2503 "*name***" is not SANergy enabled.**

Explanation: The file system is not SANergy enabled, there is nothing to repair on this file system.

User response: None. **mmsanrepairfs** cannot be run against this file system.

6027-2504 Waiting *number* **seconds for SANergy data structure cleanup.**

Explanation: This is an output from **mmsanrepairfs** reporting a delay in the command completion because it must wait until internal SANergy cleanup occurs.

User response: None. Information message only.

6027-2576 [E] Error: Daemon *value* **kernel** *value* **PAGE_SIZE mismatch.**

Explanation: The GPFS kernel extension loaded in memory does not have the same **PAGE_SIZE** value as the GPFS daemon **PAGE_SIZE** value that was returned from the POSIX **sysconf** API.

User response: Verify that the kernel header files used to build the GPFS portability layer are the same kernel header files used to build the running kernel.

6027-2600 Cannot create a new snapshot until an existing one is deleted. File system *fileSystem* **has a limit of** *number* **online snapshots.**

Explanation: The file system has reached its limit of online snapshots

User response: Delete an existing snapshot, then issue the create snapshot command again.

6027-2601 **Snapshot name** *dirName* **already exists.**

Explanation: by the **tscrsnapshot** command.

User response: Delete existing file/directory and reissue the command.

6027-2602 **Unable to delete snapshot** *snapshotName* **from file system** *fileSystem***. rc=***returnCode***.**

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Delete the snapshot using the **tsdelsnapshot** command.

6027-2603 **Unable to get permission to create snapshot, rc=***returnCode***.**

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Reissue the command.

6027-2604 **Unable to quiesce all nodes, rc=***returnCode***.**

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Restart failing nodes or switches and reissue the command.

6027-2605 **Unable to resume all nodes, rc=***returnCode***.**

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Restart failing nodes or switches.

6027-2606 **Unable to sync all nodes, rc=***returnCode***.**

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Restart failing nodes or switches and reissue the command.

6027-2607 **Cannot create new snapshot until an existing one is deleted. Fileset** *filesetName* **has a limit of** *number* **snapshots.**

Explanation: The fileset has reached its limit of snapshots.

User response: Delete an existing snapshot, then issue the create snapshot command again.

6027-2608 **Cannot create new snapshot: state of fileset** *filesetName* **is inconsistent (***badState***).**

Explanation: An operation on the cited fileset is incomplete.

User response: Complete pending fileset actions, then issue the create snapshot command again.

6027-2609 **Fileset named** *filesetName* **does not exist.**

Explanation: One of the filesets listed does not exist.

User response: Specify only existing fileset names.

6027-2610 **File system** *fileSystem* **does not contain snapshot** *snapshotName* **err =** *number***.**

Explanation: An incorrect snapshot name was specified.

User response: Select a valid snapshot and issue the command again.

6027-2611 **Cannot delete snapshot** *snapshotName* **which is in state** *snapshotState***.**

Explanation: The snapshot cannot be deleted while it is in the cited transition state because of an in-progress snapshot operation.

User response: Wait for the in-progress operation to complete and then reissue the command.

6027-2612 **Snapshot named** *snapshotName* **does not exist.**

Explanation: A snapshot to be listed does not exist.

User response: Specify only existing snapshot names.

6027-2613 **Cannot restore snapshot.** *fileSystem* **is mounted on** *number* **node(s) and in use on** *number* **node(s).**

Explanation: This message is issued by the **tsressnapshot** command.

User response: Unmount the file system and reissue the restore command.

6027-2614 **File system** *fileSystem* **does not contain snapshot** *snapshotName* **err =** *number***.**

Explanation: An incorrect snapshot name was specified.

User response: Specify a valid snapshot and issue the command again.

6027-2615 **Cannot restore snapshot** *snapshotName* **which is** *snapshotState*, **err =** *number*.

Explanation: The specified snapshot is not in a valid state.

User response: Specify a snapshot that is in a valid state and issue the command again.

6027-2616 **Restoring snapshot** *snapshotName* **requires** *quotaTypes* **quotas to be enabled.**

Explanation: The snapshot being restored requires quotas to be enabled, since they were enabled when the snapshot was created.

User response: Issue the recommended **mmchfs** command to enable quotas.

6027-2617 **You must run: mmchfs** *fileSystem* **-Q yes.**

Explanation: The snapshot being restored requires quotas to be enabled, since they were enabled when the snapshot was created.

User response: Issue the cited **mmchfs** command to enable quotas.

6027-2618 [N] **Restoring snapshot** *snapshotName* **in file system** *fileSystem* **requires** *quotaTypes* **quotas to be enabled.**

Explanation: The snapshot being restored in the cited file system requires quotas to be enabled, since they were enabled when the snapshot was created.

User response: Issue the **mmchfs** command to enable quotas.

6027-2619 **Restoring snapshot** *snapshotName* **requires** *quotaTypes* **quotas to be disabled.**

Explanation: The snapshot being restored requires quotas to be disabled, since they were not enabled when the snapshot was created.

User response: Issue the cited **mmchfs** command to disable quotas.

6027-2620 **You must run: mmchfs** *fileSystem* **-Q no.**

Explanation: The snapshot being restored requires quotas to be disabled, since they were not enabled when the snapshot was created.

User response: Issue the cited **mmchfs** command to disable quotas.

6027-2621 [N] **Restoring snapshot** *snapshotName* **in file system** *fileSystem* **requires** *quotaTypes* **quotas to be disabled.**

Explanation: The snapshot being restored in the cited file system requires quotas to be disabled, since they were disabled when the snapshot was created.

User response: Issue the **mmchfs** command to disable quotas.

6027-2623 [E] **Error deleting snapshot** *snapshotName* **in file system** *fileSystem* **err** *number*

Explanation: The cited snapshot could not be deleted during file system recovery.

User response: Run the **mmfsck** command to recover any lost data blocks.

6027-2624 **Previous snapshot** *snapshotName* **is not valid and must be deleted before a new snapshot may be created.**

Explanation: The cited previous snapshot is not valid and must be deleted before a new snapshot may be created.

User response: Delete the previous snapshot using the **mmdelsnapshot** command, and then reissue the original snapshot command.

6027-2625 **Previous snapshot** *snapshotName* **must be restored before a new snapshot may be created.**

Explanation: The cited previous snapshot must be restored before a new snapshot may be created.

User response: Run **mmrestorefs** on the previous snapshot, and then reissue the original snapshot command.

6027-2626 **Previous snapshot** *snapshotName* **is not valid and must be deleted before another snapshot may be deleted.**

Explanation: The cited previous snapshot is not valid and must be deleted before another snapshot may be deleted.

User response: Delete the previous snapshot using the **mmdelsnapshot** command, and then reissue the original snapshot command.

6027-2627 **Previous snapshot** *snapshotName* **is not valid and must be deleted before another snapshot may be restored.**

Explanation: The cited previous snapshot is not valid and must be deleted before another snapshot may be restored.

User response: Delete the previous snapshot using the **mmdelsnapshot** command, and then reissue the original snapshot command.

6027-2628 **More than one snapshot is marked for restore.**

Explanation: More than one snapshot is marked for restore.

User response: Restore the previous snapshot and then reissue the original snapshot command.

6027-2629 **Offline snapshot being restored.**

Explanation: An offline snapshot is being restored.

User response: When the restore of the offline snapshot completes, reissue the original snapshot command.

6027-2630 *Program* **failed, error** *number***.**

Explanation: The **tssnaplatest** command encountered an error and **printErrnoMsg** failed.

User response: Correct the problem shown and reissue the command.

6027-2631 **Attention: Snapshot** *snapshotName* **was being restored to** *fileSystem***.**

Explanation: A file system in the process of a snapshot restore cannot be mounted except under a restricted mount.

User response: None. Informational message only.

6027-2633 **Attention: Disk configuration for** *fileSystem* **has changed while tsdf was running.**

Explanation: The disk configuration for the cited file system changed while the **tsdf** command was running.

User response: Reissue the **mmdf** command.

6027-2634 **Attention:** *number* **of** *number* **regions in** *fileSystem* **were unavailable for free space.**

Explanation: Some regions could not be accessed during the **tsdf** run. Typically, this is due to utilities such **mmdefragfs** or **mmfsck** running concurrently.

User response: Reissue the **mmdf** command.

6027-2635 **The free space data is not available. Reissue the command without the -q option to collect it.**

Explanation: The existing free space information for the file system is currently unavailable.

User response: Reissue the **mmdf** command.

6027-2636 **Disks in storage pool** *storagePool* **must have disk usage type dataOnly.**

Explanation: A non-system storage pool cannot hold metadata or descriptors.

User response: Modify the command's disk descriptors and reissue the command.

6027-2637 **The file system must contain at least one disk for metadata.**

Explanation: The disk descriptors for this command must include one and only one storage pool that is allowed to contain metadata.

User response: Modify the command's disk descriptors and reissue the command.

6027-2638 **Maximum of** *number* **storage pools allowed.**

Explanation: The cited limit on the number of storage pools that may be defined has been exceeded.

User response: Modify the command's disk descriptors and reissue the command.

6027-2639 **Incorrect fileset name** *filesetName***.**

Explanation: The fileset name provided in the command invocation is incorrect.

User response: Correct the fileset name and reissue the command.

6027-2640 **Incorrect path to fileset junction** *filesetJunction***.**

Explanation: The path to the cited fileset junction is incorrect.

User response: Correct the junction path and reissue the command.

6027-2641 **Incorrect fileset junction name** *filesetJunction***.**

Explanation: The cited junction name is incorrect.

User response: Correct the junction name and reissue the command.

6027-2642 **Specify one and only one of FilesetName or -J JunctionPath.**

Explanation: The change fileset and unlink fileset commands accept either a fileset name or the fileset's junction path to uniquely identify the fileset. The user failed to provide either of these, or has tried to provide both.

User response: Correct the command invocation and reissue the command.

6027-2643 **Cannot create a new fileset until an existing one is deleted. File system** *fileSystem* **has a limit of** *maxNumber* **filesets.**

Explanation: An attempt to create a fileset for the cited file system failed because it would exceed the cited limit.

User response: Remove unneeded filesets and reissue the command.

6027-2644 **Comment exceeds maximum length of** *maxNumber* **characters.**

Explanation: The user-provided comment for the new fileset exceeds the maximum allowed length.

User response: Shorten the comment and reissue the command.

6027-2645 **Fileset** *filesetName* **already exists.**

Explanation: An attempt to create a fileset failed because the specified fileset name already exists.

User response: Select a unique name for the fileset and reissue the command.

6027-2646 **Unable to sync all nodes while quiesced, rc=***returnCode*

Explanation: This message is issued by the **tscrsnapshot** command.

User response: Restart failing nodes or switches and reissue the command.

6027-2647 **Fileset** *filesetName* **must be unlinked to be deleted.**

Explanation: The cited fileset must be unlinked before it can be deleted.

User response: Unlink the fileset, and then reissue the delete command.

6027-2648 **Filesets have not been enabled for file system** *fileSystem***.**

Explanation: The current file system format version does not support filesets.

User response: Change the file system format version by issuing **mmchfs -V**.

6027-2649 **Fileset** *filesetName* **contains user files and cannot be deleted unless the -f option is specified.**

Explanation: An attempt was made to delete a non-empty fileset.

User response: Remove all files and directories from the fileset, or specify the **-f** option to the **mmdelfileset** command.

6027-2650 **Fileset information is not available.**

Explanation: A fileset command failed to read file system metadata file. The file system may be corrupted.

User response: Run the **mmfsck** command to recover the file system.

6027-2651 **Fileset** *filesetName* **cannot be unlinked.**

Explanation: The user tried to unlink the root fileset, or is not authorized to unlink the selected fileset.

User response: None. The fileset cannot be unlinked.

6027-2652 **Fileset at** *junctionPath* **cannot be unlinked.**

Explanation: The user tried to unlink the root fileset, or is not authorized to unlink the selected fileset.

User response: None. The fileset cannot be unlinked.

6027-2653 **Failed to unlink fileset** *filesetName* **from** *filesetName***.**

Explanation: An attempt was made to unlink a fileset that is linked to a parent fileset that is being deleted.

User response: Delete or unlink the children, and then delete the parent fileset.

6027-2654 **Fileset** *filesetName* **cannot be deleted while other filesets are linked to it.**

Explanation: The fileset to be deleted has other filesets linked to it, and cannot be deleted without using the **-f** flag, or unlinking the child filesets.

User response: Delete or unlink the children, and then delete the parent fileset.

6027-2655 **Fileset** *filesetName* **cannot be deleted.**

Explanation: The user is not allowed to delete the root fileset.

User response: None. The fileset cannot be deleted.

6027-2656 Unable to quiesce fileset at all nodes.

Explanation: An attempt to quiesce the fileset at all nodes failed.

User response: Check communication hardware and reissue the command.

6027-2657 Fileset *filesetName* has open files. Specify -f to force unlink.

Explanation: An attempt was made to unlink a fileset that has open files.

User response: Close the open files and then reissue command, or use the **-f** option on the unlink command to force the open files to close.

6027-2658 Fileset *filesetName* cannot be linked into a snapshot at *pathName*.

Explanation: The user specified a directory within a snapshot for the junction to a fileset, but snapshots cannot be modified.

User response: Select a directory within the active file system, and reissue the command.

6027-2659 Fileset *filesetName* is already linked.

Explanation: The user specified a fileset that was already linked.

User response: Unlink the fileset and then reissue the link command.

6027-2660 Fileset *filesetName* cannot be linked.

Explanation: The fileset could not be linked. This typically happens when the fileset is in the process of being deleted.

User response: None.

6027-2661 Fileset junction *pathName* already exists.

Explanation: A file or directory already exists at the specified junction.

User response: Select a new junction name or a new directory for the link and reissue the link command.

6027-2662 Directory *pathName* for junction has too many links.

Explanation: The directory specified for the junction has too many links.

User response: Select a new directory for the link and reissue the command.

6027-2663 Fileset *filesetName* cannot be changed.

Explanation: The user specified a fileset to **tschfileset** that cannot be changed.

User response: None. You cannot change the attributes of the root fileset.

6027-2664 Fileset at *pathName* cannot be changed.

Explanation: The user specified a fileset to **tschfileset** that cannot be changed.

User response: None. You cannot change the attributes of the root fileset.

6027-2665 mmfileid already in progress for *name*.

Explanation: An **mmfileid** command is already running.

User response: Wait for the currently running command to complete, and issue the new command again.

6027-2666 mmfileid can only handle a maximum of *diskAddresses* disk addresses.

Explanation: Too many disk addresses specified.

User response: Provide less than 256 disk addresses to the command.

6027-2667 [I] Allowing block allocation for file system *fileSystem* that makes a file ill-replicated due to insufficient *resource* and puts data at risk.

Explanation: The **partialReplicaAllocation** file system option allows allocation to succeed even when all replica blocks cannot be allocated. The file was marked as not replicated correctly and the data may be at risk if one of the remaining disks fails.

User response: None. Informational message only.

6027-2670 Fileset name *filesetName* not found.

Explanation: The fileset name that was specified with the command invocation was not found.

User response: Correct the fileset name and reissue the command.

6027-2671 Fileset command on *fileSystem* failed; snapshot *snapshotName* must be restored first.

Explanation: The file system is being restored either from an offline backup or a snapshot, and the restore operation has not finished. Fileset commands cannot be run.

User response: Run the **mmrestorefs** command to

complete the snapshot restore operation or to finish the offline restore, then reissue the fileset command.

6027-2672 Junction parent directory inode number *inodeNumber* **is not valid.**

Explanation: An inode number passed to **tslinkfileset** is not valid.

User response: Check the **mmlinkfileset** command arguments for correctness. If a valid junction path was provided, contact the IBM Support Center.

6027-2673 [X] Duplicate owners of an allocation region (index *indexNumber*, **region** *regionNumber*, **pool** *poolNumber*) **were detected for file system** *fileSystem*: **nodes** *nodeName* **and** *nodeName*.

Explanation: The allocation region should not have duplicate owners.

User response: Contact the IBM Support Center.

6027-2674 [X] The owner of an allocation region (index *indexNumber*, **region** *regionNumber*, **pool** *poolNumber*) **that was detected for file system** *fileSystem*: **node** *nodeName* **is not valid.**

Explanation: The file system had detected a problem with the ownership of an allocation region. This may result in a corrupted file system and loss of data. One or more nodes may be terminated to prevent any further damage to the file system.

User response: Unmount the file system and run the **kwdmmfsck** command to repair the file system.

6027-2675 Only file systems with NFSv4 ACL semantics enabled can be mounted on this platform.

Explanation: A user is trying to mount a file system on Microsoft Windows, but the ACL semantics disallow NFSv4 ACLs.

User response: Enable NFSv4 ACL semantics using the **mmchfs** command (**-k** option)

6027-2676 Only file systems with NFSv4 locking semantics enabled can be mounted on this platform.

Explanation: A user is trying to mount a file system on Microsoft Windows, but the POSIX locking semantics are in effect.

User response: Enable NFSv4 locking semantics using the **mmchfs** command (**-D** option).

6027-2677 Fileset *filesetName* **has pending changes that need to be synced.**

Explanation: A user is trying to change a caching option for a fileset while it has local changes that are not yet synced with the home server.

User response: Perform AFM recovery before reissuing the command.

6027-2678 File system *fileSystem* **is mounted on** *nodes* **nodes or fileset** *filesetName* **is not unlinked.**

Explanation: A user is trying to change a caching feature for a fileset while the file system is still mounted or the fileset is still linked.

User response: Unmount the file system from all nodes or unlink the fileset before reissuing the command.

6027-2679 Mount of *fileSystem* **failed because mount event not handled by any data management application.**

Explanation: The mount failed because the file system is enabled for DMAPI events (**-z yes**), but there was no data management application running to handle the event.

User response: Make sure the DM application (for example HSM or HPSS) is running before the file system is mounted.

6027-2680 AFM filesets cannot be created for file system *fileSystem*.

Explanation: The current file system format version does not support AFM-enabled filesets; the **-p** option cannot be used.

User response: Change the file system format version by issuing **mmchfs -V**.

6027-2681 Snapshot *snapshotName* **has linked independent filesets**

Explanation: The specified snapshot is not in a valid state.

User response: Correct the problem and reissue the command.

6027-2682 [E] Set quota file attribute error (*reasonCode***)***explanation*

Explanation: While mounting a file system a new quota file failed to be created due to inconsistency with the current degree of replication or the number of failure groups.

User response: Disable quotas. Check and correct the

degree of replication and the number of failure groups. Re-enable quotas.

6027-2683 **Fileset** *filesetName* **in file system** *fileSystem* **does not contain snapshot** *snapshotName*, **err =** *number*

Explanation: An incorrect snapshot name was specified.

User response: Select a valid snapshot and issue the command again.

6027-2684 **File system** *fileSystem* **does not contain global snapshot** *snapshotName*, **err =** *number*

Explanation: An incorrect snapshot name was specified.

User response: Select a valid snapshot and issue the command again.

6027-2685 **Total file system capacity allows** *minMaxInodes* **inodes in** *fileSystem*. **Currently the total inode limits used by all the inode spaces in** *inodeSpace* **is** *inodeSpaceLimit*. **There must be at least** *number* **inodes available to create a new inode space. Use the mmlsfileset -L command to show the maximum inode limits of each fileset. Try reducing the maximum inode limits for some of the inode spaces in** *fileSystem*.

Explanation: The number of inodes available is too small to create a new inode space.

User response: Reduce the maximum inode limits and issue the command again.

6027-2688 **Only independent filesets can be configured as AFM filesets. The --inode-space=new option is required.**

Explanation: Only independent filesets can be configured for caching.

User response: Specify the **--inode-space=new** option.

6027-2689 **The value for --block-size must be the keyword auto or the value must be of the form [n]K, [n]M, [n]G or [n]T, where n is an optional integer in the range 1 to 1023.**

Explanation: An invalid value was specified with the **--block-size** option.

User response: Reissue the command with a valid option.

6027-2690 **Fileset** *filesetName* **can only be linked within its own inode space.**

Explanation: A dependent fileset can only be linked within its own inode space.

User response: Correct the junction path and reissue the command.

6027-2691 **The fastea feature needs to be enabled for file system** *fileSystem* **before creating AFM filesets.**

Explanation: The current file system on-disk format does not support storing of extended attributes in the file's inode. This is required for AFM-enabled filesets.

User response: Use the **mmmigratefs** command to enable the fast extended-attributes feature.

6027-2692 **Error encountered while processing the input file.**

Explanation: The **tscrsnapshot** command encountered an error while processing the input file.

User response: Check and validate the fileset names listed in the input file.

6027-2693 **Fileset junction name** *junctionName* **conflicts with the current setting of mmsnapdir.**

Explanation: The fileset junction name conflicts with the current setting of **mmsnapdir**.

User response: Select a new junction name or a new directory for the link and reissue the **mmlinkfileset** command.

6027-2694 [I] **The requested maximum number of inodes is already at** *number*.

Explanation: The specified number of nodes is already in effect.

User response: This is an informational message.

6027-2695 [E] **The number of inodes to preallocate cannot be higher than the maximum number of inodes.**

Explanation: The specified number of nodes to preallocate is not valid.

User response: Correct the **--inode-limit** argument then retry the command.

6027-2696 [E] The number of inodes to preallocate cannot be lower than the *number* inodes already allocated.

Explanation: The specified number of nodes to preallocate is not valid.

User response: Correct the **--inode-limit** argument then retry the command.

6027-2697 Fileset at *junctionPath* has pending changes that need to be synced.

Explanation: A user is trying to change a caching option for a fileset while it has local changes that are not yet synced with the home server.

User response: Perform AFM recovery before reissuing the command.

6027-2698 File system *fileSystem* is mounted on *nodes* nodes or fileset at *junctionPath* is not unlinked.

Explanation: A user is trying to change a caching feature for a fileset while the file system is still mounted or the fileset is still linked.

User response: Unmount the file system from all nodes or unlink the fileset before reissuing the command.

6027-2699 Cannot create a new independent fileset until an existing one is deleted. File system *fileSystem* has a limit of *maxNumber* independent filesets.

Explanation: An attempt to create an independent fileset for the cited file system failed because it would exceed the cited limit.

User response: Remove unneeded independent filesets and reissue the command.

6027-2700 [E] A node join was rejected. This could be due to incompatible daemon versions, failure to find the node in the configuration database, or no configuration manager found.

Explanation: A request to join nodes was explicitly rejected.

User response: Verify that compatible versions of GPFS are installed on all nodes. Also, verify that the joining node is in the configuration database.

6027-2701 The mmpmon command file is empty.

Explanation: The **mmpmon** command file is empty.

User response: Check file size, existence, and access permissions.

6027-2702 Unexpected mmpmon response from file system daemon.

Explanation: An unexpected response was received to an **mmpmon** request.

User response: Ensure that the **mmfsd** daemon is running. Check the error log. Ensure that all GPFS software components are at the same version.

6027-2703 Unknown mmpmon command *command*.

Explanation: An unknown **mmpmon** command was read from the input file.

User response: Correct the command and rerun.

6027-2704 Permission failure. The command requires root authority to execute.

Explanation: The **mmpmon** command was issued with a nonzero UID.

User response: Log on as root and reissue the command.

6027-2705 Could not establish connection to file system daemon.

Explanation: The connection between a GPFS command and the **mmfsd** daemon could not be established. The daemon may have crashed, or never been started, or (for **mmpmon**) the allowed number of simultaneous connections has been exceeded.

User response: Ensure that the **mmfsd** daemon is running. Check the error log. For **mmpmon**, ensure that the allowed number of simultaneous connections has not been exceeded.

6027-2706 [I] Recovered *number* nodes.

Explanation: The asynchronous part (phase 2) of node failure recovery has completed.

User response: None. Informational message only.

6027-2707 [I] Node join protocol waiting *value* seconds for node recovery

Explanation: Node join protocol is delayed until phase 2 of previous node failure recovery protocol is complete.

User response: None. Informational message only.

6027-2708 [E] Rejected node join protocol. Phase two of node failure recovery appears to still be in progress.

Explanation: Node join protocol is rejected after a number of internal delays and phase two node failure protocol is still in progress.

User response: None. Informational message only.

6027-2709 Configuration manager node *nodeName* **not found in the node list.**

Explanation: The specified node was not found in the node list.

User response: Add the specified node to the node list and reissue the command.

6027-2710 [E] Node *nodeName* **is being expelled due to expired lease.**

Explanation: The nodes listed did not renew their lease in a timely fashion and will be expelled from the cluster.

User response: Check the network connection between this node and the node specified above.

6027-2711 [E] File system table full.

Explanation: The **mmfsd** daemon cannot add any more file systems to the table because it is full.

User response: None. Informational message only.

6027-2712 Option '*optionName***' has been deprecated.**

Explanation: The option that was specified with the command is no longer supported. A warning message is generated to indicate that the option has no effect.

User response: Correct the command line and then reissue the command.

6027-2713 Permission failure. The command requires *SuperuserName* **authority to execute.**

Explanation: The command, or the specified command option, requires administrative authority.

User response: Log on as a user with administrative privileges and reissue the command.

6027-2714 Could not appoint node *nodeName* **as cluster manager.** *errorString*

Explanation: The **mmchmgr -c** command generates this message if the specified node cannot be appointed as a new cluster manager.

User response: Make sure that the specified node is a quorum node and that GPFS is running on that node.

6027-2715 Could not appoint a new cluster manager. *errorString*

Explanation: The **mmchmgr -c** command generates this message when a node is not available as a cluster manager.

User response: Make sure that GPFS is running on a sufficient number of quorum nodes.

6027-2716 [I] Challenge response received; canceling disk election.

Explanation: The node has challenged another node, which won the previous election, and detected a response to the challenge.

User response: None. Informational message only.

6027-2717 Node *nodeName* **is already a cluster manager or another node is taking over as the cluster manager.**

Explanation: The **mmchmgr -c** command generates this message if the specified node is already the cluster manager.

User response: None. Informational message only.

6027-2718 Incorrect port range: GPFSCMDPORTRANGE='*range***'. Using default.**

Explanation: The GPFS command port range format is *lllll*[*-hhhhh*], where *lllll* is the low port value and *hhhhh* is the high port value. The valid range is 1 to 65535.

User response: None. Informational message only.

6027-2719 The files provided do not contain valid quota entries.

Explanation: The quota file provided does not have valid quota entries.

User response: Check that the file being restored is a valid GPFS quota file.

6027-2722 [E] Node limit of *number* **has been reached. Ignoring** *nodeName***.**

Explanation: The number of nodes that have been added to the cluster is greater than some cluster members can handle.

User response: Delete some nodes from the cluster using the **mmdelnode** command, or shut down GPFS on nodes that are running older versions of the code with lower limits.

6027-2723 [N] This node (*nodeName*) is now Cluster Manager for *clusterName*.

Explanation: This is an informational message when a new cluster manager takes over.

User response: None. Informational message only.

6027-2724 [I] *reasonString*. **Probing cluster** *clusterName*

Explanation: This is an informational message when a lease request has not been renewed.

User response: None. Informational message only.

6027-2725 [N] Node *nodeName* **lease renewal is overdue. Pinging to check if it is alive**

Explanation: This is an informational message on the cluster manager when a lease request has not been renewed.

User response: None. Informational message only.

6027-2726 [I] Recovered *number* **nodes for file system** *fileSystem*.

Explanation: The asynchronous part (phase 2) of node failure recovery has completed.

User response: None. Informational message only.

6027-2727 *fileSystem*: **quota manager is not available.**

Explanation: An attempt was made to perform a quota command without a quota manager running. This could be caused by a conflicting offline **mmfsck** command.

User response: Reissue the command once the conflicting program has ended.

6027-2728 [N] Connection from *node* **rejected because it does not support IPv6**

Explanation: A connection request was received from a node that does not support Internet Protocol Version 6 (IPv6), and at least one node in the cluster is configured with an IPv6 address (not an IPv4-mapped one) as its primary address. Since the connecting node will not be able to communicate with the IPv6 node, it is not permitted to join the cluster.

User response: Upgrade the connecting node to a version of GPFS that supports IPv6, or delete all nodes with IPv6-only addresses from the cluster.

6027-2729 Value *value* **for option** *optionName* **is out of range. Valid values are** *value* **through** *value*.

Explanation: An out of range value was specified for the specified option.

User response: Correct the command line.

6027-2730 [E] Node *nodeName* **failed to take over as cluster manager.**

Explanation: An attempt to takeover as cluster manager failed.

User response: Make sure that GPFS is running on a sufficient number of quorum nodes.

6027-2731 Failed to locate a working cluster manager.

Explanation: The cluster manager has failed or changed. The new cluster manager has not been appointed.

User response: Check the internode communication configuration and ensure enough GPFS nodes are up to make a quorum.

6027-2732 Attention: No data disks remain in the system pool. Use mmapplypolicy to migrate all data left in the system pool to other storage pool.

Explanation: The **mmchdisk** command has been issued but no data disks remain in the system pool. Warn user to use **mmapplypolicy** to move data to other storage pool.

User response: None. Informational message only.

6027-2733 The file system name (*fsname*) is longer than the maximum allowable length (*maxLength*).

Explanation: The file system name is invalid because it is longer than the maximum allowed length of 255 characters.

User response: Specify a file system name whose length is 255 characters or less and reissue the command.

6027-2734 [E] Disk failure from node *nodeName* **Volume** *name*. **Physical volume** *name*.

Explanation: An I/O request to a disk or a request to fence a disk has failed in such a manner that GPFS can no longer use the disk.

User response: Check the disk hardware and the software subsystems in the path to the disk.

6027-2735 [E] Not a manager

Explanation: This node is not a manager or no longer a manager of the type required to proceed with the operation. This could be caused by the change of manager in the middle of the operation.

User response: Retry the operation.

6027-2736 The value for --block-size must be the keyword auto or the value must be of the form *n*K, *n*M, *n*G or *n*T, where n is an optional integer in the range 1 to 1023.

Explanation: An invalid value was specified with the **--block-size** option.

User response: Reissue the command with a valid option.

6027-2737 Editing quota limits for root fileset is not permitted.

Explanation: The root fileset was specified for quota limits editing in the **mmedquota** command.

User response: Specify a non-root fileset in the **mmedquota** command. Editing quota limits for the root fileset is prohibited.

6027-2738 Editing quota limits for the root user is not permitted

Explanation: The **root** user was specified for quota limits editing in the **mmedquota** command.

User response: Specify a valid user or group in the **mmedquota** command. Editing quota limits for the **root** user or **system** group is prohibited.

6027-2739 Editing quota limits for *groupName* group not permitted.

Explanation: The **system** group was specified for quota limits editing in the **mmedquota** command.

User response: Specify a valid user or group in the **mmedquota** command. Editing quota limits for the **root** user or **system** group is prohibited.

6027-2740 [I] Starting new election as previous clmgr is expelled

Explanation: This node is taking over as clmgr without challenge as the old clmgr is being expelled.

User response: None. Informational message only.

6027-2741 [W] This node can not continue to be cluster manager

Explanation: This node invoked the user-specified callback handler for event **tiebreakerCheck** and it returned a non-zero value. This node cannot continue to be the cluster manager.

User response: None. Informational message only.

6027-2742 [I] CallExitScript: exit script *exitScript* on event *eventName* returned code *returnCode*, quorumloss.

Explanation: This node invoked the user-specified callback handler for the **tiebreakerCheck** event and it returned a non-zero value. The user-specified action with the error is **quorumloss**.

User response: None. Informational message only.

6027-2743 Permission denied.

Explanation: The command is invoked by an unauthorized user.

User response: Retry the command with an authorized user.

6027-2744 [D] Invoking tiebreaker callback script

Explanation: The node is invoking the callback script due to change in quorum membership.

User response: None. Informational message only.

6027-2745 [E] File system is not mounted.

Explanation: A command was issued, which requires that the file system be mounted.

User response: Mount the file system and reissue the command.

6027-2746 [E] Too many disks unavailable for this server to continue serving a RecoveryGroup.

Explanation: RecoveryGroup panic: Too many disks unavailable to continue serving this RecoveryGroup. This server will resign, and failover to an alternate server will be attempted.

User response: Ensure the alternate server took over. Determine what caused this event and address the situation. Prior messages may help determine the cause of the event.

6027-2747 [E] Inconsistency detected between the local node number retrieved from 'mmsdrfs' (*nodeNumber*) and the node number retrieved from 'mmfs.cfg' (*nodeNumber*).

Explanation: The node number retrieved by obtaining the list of nodes in the **mmsdrfs** file did not match the node number contained in **mmfs.cfg**. There may have been a recent change in the IP addresses being used by network interfaces configured at the node.

User response: Stop and restart GPFS daemon.

6027-2748 Terminating because a conflicting program on the same inode space *inodeSpace* is running.

Explanation: A program detected that it must terminate because a conflicting program is running.

User response: Reissue the command after the conflicting program ends.

6027-2749 Specified locality group '*number*' does not match disk '*name*' locality group '*number*'. To change locality groups in an SNC environment, please use the mmdeldisk and mmadddisk commands.

Explanation: The locality group specified on the **mmchdisk** command does not match the current locality group of the disk.

User response: To change locality groups in an SNC environment, use the **mmdeldisk** and **mmadddisk** commands.

6027-2750 [I] Node *NodeName* is now the Group Leader.

Explanation: A new cluster Group Leader has been assigned.

User response: None. Informational message only.

6027-2751 [I] Starting new election: Last elected: *NodeNumber* Sequence: *SequenceNumber*

Explanation: A new disk election will be started. The disk challenge will be skipped since the last elected node was either none or the local node.

User response: None. Informational message only.

6027-2752 [I] This node got elected. Sequence: *SequenceNumber*

Explanation: Local node got elected in the disk election. This node will become the cluster manager.

User response: None. Informational message only.

6027-2753 [N] Responding to disk challenge: response: *ResponseValue*. Error code: *ErrorCode*.

Explanation: A disk challenge has been received, indicating that another node is attempting to become a Cluster Manager. Issuing a challenge response, to confirm the local node is still alive and will remain the Cluster Manager.

User response: None. Informational message only.

6027-2754 [X] Challenge thread did not respond to challenge in time: took *TimeIntervalSecs* seconds.

Explanation: Challenge thread took too long to respond to a disk challenge. Challenge thread will exit, which will result in the local node losing quorum.

User response: None. Informational message only.

6027-2755 [N] Another node committed disk election with sequence *CommittedSequenceNumber* (our sequence was *OurSequenceNumber*).

Explanation: Another node committed a disk election with a sequence number higher than the one used when this node used to commit an election in the past. This means that the other node has become, or is becoming, a Cluster Manager. To avoid having two Cluster Managers, this node will lose quorum.

User response: None. Informational message only.

6027-2756 Attention: In file system *FileSystemName*, *FileSetName* (*Default*) *QuotaLimitType*(*QuotaLimit*) for *QuotaTypeUserName/GroupName/FilesetName* is too small. Suggest setting it higher than *minQuotaLimit*.

Explanation: Users set too low quota limits. It will cause unexpected quota behavior. MinQuotaLimit is computed through:

1. for block: QUOTA_THRESHOLD * MIN_SHARE_BLOCKS * subblocksize
2. for inode: QUOTA_THRESHOLD * MIN_SHARE_INODES

User response: Users should reset quota limits so that they are more than MinQuotaLimit. It is just a warning. Quota limits will be set anyway.

6027-2757 [E] The peer snapshot is in progress. Queue cannot be flushed now.

Explanation: The Peer Snapshot is in progress. Queue cannot be flushed now.

User response: Reissue the command once the peer snapshot has ended.

6027-2758 [E] The AFM target does not support this operation. Run mmafmconfig on the AFM target cluster.

Explanation: The **.afmctl** file is probably not present on the AFM target cluster.

User response: Run **mmafmconfig** on the AFM target cluster to configure the AFM target cluster.

6027-2759 [N] Disk lease period expired in cluster *ClusterName***. Attempting to reacquire lease.**

Explanation: The disk lease period expired, which will prevent the local node from being able to perform disk I/O. This can be caused by a temporary communication outage.

User response: If message is repeated then the communication outage should be investigated.

6027-2760 [N] Disk lease reacquired in cluster *ClusterName***.**

Explanation: The disk lease has been reacquired, and disk I/O will be resumed.

User response: None. Informational message only.

6027-2761 Unable to run *command* **on '***fileSystem***' while the file system is mounted in restricted mode.**

Explanation: A command that can alter data in a file system was issued while the file system was mounted in restricted mode.

User response: Mount the file system in read-only or read-write mode or unmount the file system and then reissue the command.

6027-2762 Unable to run *command* **on '***fileSystem***' while the file system is suspended.**

Explanation: A command that can alter data in a file system was issued while the file system was suspended.

User response: Resume the file system and reissue the command.

6027-2763 Unable to start *command* **on '***fileSystem***' because conflicting program** *name* **is running. Waiting until it completes.**

Explanation: A program detected that it cannot start because a conflicting program is running. The program will automatically start once the conflicting program has ended as long as there are no other conflicting programs running at that time.

User response: None. Informational message only.

6027-2764 Terminating *command* **on** *fileSystem* **because a conflicting program** *name* **is running.**

Explanation: A program detected that it must terminate because a conflicting program is running.

User response: Reissue the command after the conflicting program ends.

6027-2765 *command* **on '***fileSystem***' is finished waiting. Processing continues ...** *name*

Explanation: A program detected that it can now continue the processing since a conflicting program has ended.

User response: None. Informational message only.

6027-2766 [I] User script has chosen to expel node *nodeName* **instead of node** *nodeName***.**

Explanation: User has specified a callback script that is invoked whenever a decision is about to be taken on what node should be expelled from the active cluster. As a result of the execution of the script, GPFS will reverse its decision on what node to expel.

User response: None.

6027-2767 [E] Error *errorNumber* **while accessing tiebreaker devices.**

Explanation: An error was encountered while reading from or writing to the tiebreaker devices. When such error happens while the cluster manager is checking for challenges, it will cause the cluster manager to lose cluster membership.

User response: Verify the health of the tiebreaker devices.

6027-2770 Disk *diskName* **belongs to a write-affinity enabled storage pool. Its failure group cannot be changed.**

Explanation: The failure group specified on the **mmchdisk** command does not match the current failure group of the disk.

User response: Use the **mmdeldisk** and **mmadddisk** commands to change failure groups in a write-affinity enabled storage pool.

6027-2771 *fileSystem***: Default per-fileset quotas are disabled for** *quotaType***.**

Explanation: A command was issued to modify default fileset-level quota, but default quotas are not enabled.

User response: Ensure the **--perfileset-quota** option is in effect for the file system, then use the

mmdefquotaon command to enable default fileset-level quotas. After default quotas are enabled, issue the failed command again.

6027-2772 **Cannot close disk** *name***.**

Explanation: Could not access the specified disk.

User response: Check the disk hardware and the path to the disk. Refer to "Unable to access disks" on page 353.

6027-2773 *fileSystem*:*filesetName***: default quota for** *quotaType* **is disabled.**

Explanation: A command was issued to modify default quota, but default quota is not enabled.

User response: Ensure the **-Q yes** option is in effect for the file system, then enable default quota with the **mmdefquotaon** command.

6027-2774 *fileSystem***: Per-fileset quotas are not enabled.**

Explanation: A command was issued to modify fileset-level quota, but per-fileset quota management is not enabled.

User response: Ensure that the **--perfileset-quota** option is in effect for the file system and reissue the command.

6027-2775 **Storage pool named** *poolName* **does not exist.**

Explanation: The **mmlspool** command was issued, but the specified storage pool does not exist.

User response: Correct the input and reissue the command.

6027-2776 **Attention: A disk being stopped reduces the degree of system metadata replication (***value***) or data replication (***value***) to lower than tolerable.**

Explanation: The **mmchdisk stop** command was issued, but the disk cannot be stopped because of the current file system metadata and data replication factors.

User response: Make more disks available, delete unavailable disks, or change the file system metadata replication factor. Also check the current value of the **unmountOnDiskFail** configuration parameter.

6027-2777 [E] **Node** *nodeName* **is being expelled because of an expired lease. Pings sent:** *pingsSent***. Replies received:** *pingRepliesReceived***.**

Explanation: The node listed did not renew its lease in a timely fashion and is being expelled from the cluster.

User response: Check the network connection between this node and the node listed in the message.

6027-2778 [I] **Node** *nodeName***: ping timed out. Pings sent:** *pingsSent***. Replies received:** *pingRepliesReceived***.**

Explanation: Ping timed out for the node listed, which should be the cluster manager. A new cluster manager will be chosen while the current cluster manager is expelled from the cluster.

User response: Check the network connection between this node and the node listed in the message.

6027-2779 [E] Challenge thread stopped.

Explanation: A tiebreaker challenge thread stopped because of an error. Cluster membership will be lost.

User response: Check for additional error messages. File systems will be unmounted, then the node will rejoin the cluster.

6027-2780 [E] Not enough quorum nodes reachable: *reachableNodes***.**

Explanation: The cluster manager cannot reach a sufficient number of quorum nodes, and therefore must resign to prevent cluster partitioning.

User response: Determine if there is a network outage or if too many nodes have failed.

6027-2781 [E] Lease expired for *numSecs* **seconds (shutdownOnLeaseExpiry).**

Explanation: Disk lease expired for too long, which results in the node losing cluster membership.

User response: None. The node will attempt to rejoin the cluster.

6027-2782 [E] This node is being expelled from the cluster.

Explanation: This node received a message instructing it to leave the cluster, which might indicate communication problems between this node and some other node in the cluster.

User response: None. The node will attempt to rejoin the cluster.

6027-2783 [E] New leader elected with a higher ballot number.

Explanation: A new group leader was elected with a higher ballot number, and this node is no longer the

leader. Therefore, this node must leave the cluster and rejoin.

User response: None. The node will attempt to rejoin the cluster.

6027-2784 [E] No longer a cluster manager or lost quorum while running a group protocol.

Explanation: Cluster manager no longer maintains quorum after attempting to run a group protocol, which might indicate a network outage or node failures.

User response: None. The node will attempt to rejoin the cluster.

6027-2785 [X] A severe error was encountered during cluster probe.

Explanation: A severe error was encountered while running the cluster probe to determine the state of the nodes in the cluster.

User response: Examine additional error messages. The node will attempt to rejoin the cluster.

6027-2786 [E] Unable to contact any quorum nodes during cluster probe.

Explanation: This node has been unable to contact any quorum nodes during cluster probe, which might indicate a network outage or too many quorum node failures.

User response: Determine whether there was a network outage or whether quorum nodes failed.

6027-2787 [E] Unable to contact enough other quorum nodes during cluster probe.

Explanation: This node, a quorum node, was unable to contact a sufficient number of quorum nodes during cluster probe, which might indicate a network outage or too many quorum node failures.

User response: Determine whether there was a network outage or whether quorum nodes failed.

6027-2788 [E] Attempt to run leader election failed with error *errorNumber***.**

Explanation: This node attempted to run a group leader election but failed to get elected. This failure might indicate that two or more quorum nodes attempted to run the election at the same time. As a result, this node will lose cluster membership and then attempt to rejoin the cluster.

User response: None. The node will attempt to rejoin the cluster.

6027-2789 [E] Tiebreaker script returned a non-zero value.

Explanation: The tiebreaker script, invoked during group leader election, returned a non-zero value, which results in the node losing cluster membership and then attempting to rejoin the cluster.

User response: None. The node will attempt to rejoin the cluster.

6027-2790 Attention: Disk parameters were changed. Use the mmrestripefs command with the -r option to relocate data and metadata.

Explanation: The **mmchdisk** command with the **change** option was issued.

User response: Issue the **mmrestripefs -r** command to relocate data and metadata.

6027-2791 Disk *diskName* **does not belong to file system** *deviceName***.**

Explanation: The input disk name does not belong to the specified file system.

User response: Correct the command line.

6027-2792 The current file system version does not support default per-fileset quotas.

Explanation: The current version of the file system does not support default fileset-level quotas.

User response: Use the **mmchfs -V** command to activate the new function.

6027-2793 [E] Contents of local *fileName* **file are invalid. Node may be unable to be elected group leader.**

Explanation: In an environment where tie-breaker disks are used, the contents of the ballot file have become invalid, possibly because the file has been overwritten by another application. This node will be unable to be elected group leader.

User response: Run **mmcommon resetTiebreaker**, which will ensure the GPFS daemon is down on all quorum nodes and then remove the given file on this node. After that, restart the cluster on this and on the other nodes.

6027-2794 [E] Invalid content of disk paxos sector for disk *diskName***.**

Explanation: In an environment where tie-breaker disks are used, the contents of either one of the tie-breaker disks or the ballot files became invalid, possibly because the file has been overwritten by another application.

User response: Examine **mmfs.log** file on all quorum nodes for indication of a corrupted ballot file. If 6027-2793 is found then follow instructions for that message. If problem cannot be resolved, shut down GPFS across the cluster, undefine, and then redefine the **tiebreakerdisks** configuration variable, and finally restart the cluster.

6027-2795 **An error occurred while executing command for** *fileSystem***.**

Explanation: A quota command encountered a problem on a file system. Processing continues with the next file system.

User response: None. Informational message only.

6027-2796 [W] **Callback event** *eventName* **is not supported on this node; processing continues ...**

Explanation: informational

User response:

6027-2797 [I] **Node** *nodeName***: lease request received late. Pings sent:** *pingsSent***. Maximum pings missed:** *maxPingsMissed***.**

Explanation: The cluster manager reports that the lease request from the given node was received late, possibly indicating a network outage.

User response: Check the network connection between this node and the node listed in the message.

6027-2798 [E] **The node** *nodeName* **does not have a valid Extended License to run the requested command.**

Explanation: The file system manager node does not have a valid extended license to run ILM, AFM, or CNFS commands.

User response: Make sure `gpfs.ext` package is installed correctly on file system manager node and try again.

6027-2799 **Option '***option***' is incompatible with option '***option***'.**

Explanation: The options specified on the command are incompatible.

User response: Do not specify these two options together.

6027-2800 **Available memory exceeded on request to allocate** *number* **bytes. Trace point** *sourceFile-tracePoint***.**

Explanation: The available memory was exceeded during an allocation request made from the cited source file and trace point.

User response: Try shutting down and then restarting GPFS. If the problem recurs, contact the IBM Support Center.

6027-2801 **Policy set syntax version** *versionString* **not supported.**

Explanation: The policy rules do not comply with the supported syntax.

User response: Rewrite the policy rules, following the documented, supported syntax and keywords.

6027-2802 **Object name '***poolName_or_filesetName***' is not valid.**

Explanation: The cited name is not a valid GPFS object, names an object that is not valid in this context, or names an object that no longer exists.

User response: Correct the input to identify a GPFS object that exists and is valid in this context.

6027-2803 **Policy set must start with VERSION.**

Explanation: The policy set does not begin with **VERSION** as required.

User response: Rewrite the policy rules, following the documented, supported syntax and keywords.

6027-2804 **Unexpected SQL result code -** *sqlResultCode***.**

Explanation: This could be an IBM programming error.

User response: Check that your SQL expressions are correct and supported by the current release of GPFS. If the error recurs, contact the IBM Support Center.

6027-2805 [I] **Loaded policy '***policyFileName or filesystemName***':** *summaryOfPolicyRules*

Explanation: The specified loaded policy has the specified policy rules.

User response: None. Informational message only.

6027-2806 [E] **Error while validating policy '***policyFileName or filesystemName***': rc=***errorCode***:** *errorDetailsString*

Explanation: An error occurred while validating the specified policy.

User response: Correct the policy rules, heeding the error details in this message and other messages issued immediately before or after this message. Use the **mmchpolicy** command to install a corrected policy rules file.

6027-2807 [W] Error in evaluation of placement policy for file *fileName*: *errorDetailsString*

Explanation: An error occurred while evaluating the installed placement policy for a particular new file. Although the policy rules appeared to be syntactically correct when the policy was installed, evidently there is a problem when certain values of file attributes occur at runtime.

User response: Determine which file names and attributes trigger this error. Correct the policy rules, heeding the error details in this message and other messages issued immediately before or after this message. Use the **mmchpolicy** command to install a corrected policy rules file.

6027-2808 In rule '*ruleName***' (***ruleNumber***), '***wouldBePoolName***' is not a valid pool name.**

Explanation: The cited name that appeared in the cited rule is not a valid pool name. This may be because the cited name was misspelled or removed from the file system.

User response: Correct or remove the rule.

6027-2809 Validated policy '*policyFileName or filesystemName***': *summaryOfPolicyRules***

Explanation: The specified validated policy has the specified policy rules.

User response: None. Informational message only.

6027-2810 [W] There are *numberOfPools* **storage pools but the policy file is missing or empty.**

Explanation: The cited number of storage pools are defined, but the policy file is missing or empty.

User response: You should probably install a policy with placement rules using the **mmchpolicy** command, so that at least some of your data will be stored in your nonsystem storage pools.

6027-2811 Policy has no storage pool placement rules!

Explanation: The policy has no storage pool placement rules.

User response: You should probably install a policy with placement rules using the **mmchpolicy** command, so that at least some of your data will be stored in your nonsystem storage pools.

6027-2812 Keyword '*keywordValue***' begins a second** *clauseName* **clause - only one is allowed.**

Explanation: The policy rule should only have one clause of the indicated type.

User response: Correct the rule and reissue the policy command.

6027-2813 This '*ruleName***' rule is missing a** *clauseType* **required clause.**

Explanation: The policy rule must have a clause of the indicated type.

User response: Correct the rule and reissue the policy command.

6027-2814 This '*ruleName***' rule is of unknown type or not supported.**

Explanation: The policy rule set seems to have a rule of an unknown type or a rule that is unsupported by the current release of GPFS.

User response: Correct the rule and reissue the policy command.

6027-2815 The value '*value***' is not supported in a '***clauseType***' clause.**

Explanation: The policy rule clause seems to specify an unsupported argument or value that is not supported by the current release of GPFS.

User response: Correct the rule and reissue the policy command.

6027-2816 Policy rules employ features that would require a file system upgrade.

Explanation: One or more policy rules have been written to use new features that cannot be installed on a back-level file system.

User response: Install the latest GPFS software on all nodes and upgrade the file system or change your rules. (Note that **LIMIT** was introduced in GPFS Release 3.2.)

6027-2817 Error on popen/pclose (*command_string***): rc=***return_code_from_popen_or_pclose***

Explanation: The execution of the *command_string* by **popen/pclose** resulted in an error.

User response: To correct the error, do one or more of the following:

Check that the standard **m4** macro processing command is installed on your system as **/usr/bin/m4**.

Or:

Set the **MM_M4_CMD** environment variable.

Or:

Correct the macro definitions in your policy rules file.

If the problem persists, contact the IBM Support Center.

6027-2818 **A problem occurred during m4 processing of policy rules. rc =** *return_code_from_popen_pclose_or_m4*

Explanation: An attempt to expand the policy rules with an **m4** subprocess yielded some warnings or errors or the **m4** macro wrote some output to standard error. Details or related messages may follow this message.

User response: To correct the error, do one or more of the following:

Check that the standard **m4** macro processing command is installed on your system as **/usr/bin/m4**.

Or:

Set the **MM_M4_CMD** environment variable.

Or:

Correct the macro definitions in your policy rules file.

If the problem persists, contact the IBM Support Center.

6027-2819 **Error opening temp file** *temp_file_name***:** *errorString*

Explanation: An error occurred while attempting to open the specified temporary work file.

User response: Check that the path name is defined and accessible. Check the file and then reissue the command.

6027-2820 **Error reading temp file** *temp_file_name***:** *errorString*

Explanation: An error occurred while attempting to read the specified temporary work file.

User response: Check that the path name is defined and accessible. Check the file and then reissue the command.

6027-2821 **Rule '***ruleName***' (***ruleNumber***) specifies a THRESHOLD for EXTERNAL POOL '***externalPoolName***'. This is not supported.**

Explanation: GPFS does not support the **THRESHOLD** clause within a migrate rule that names an external pool in the **FROM POOL** clause.

User response: Correct or remove the rule.

6027-2822 **This file system does not support fast extended attributes, which are needed for encryption.**

Explanation: Fast extended attributes need to be supported by the file system for encryption to be activated.

User response: Enable the fast extended attributes feature in this file system.

6027-2823 [E] **Encryption activated in the file system, but node not enabled for encryption.**

Explanation: The file system is enabled for encryption, but this node is not.

User response: Ensure the GPFS encryption packages are installed. Verify if encryption is supported on this node architecture.

6027-2824 **This file system version does not support encryption rules.**

Explanation: This file system version does not support encryption.

User response: Update the file system to a version which supports encryption.

6027-2825 **Duplicate encryption set name '***setName***'.**

Explanation: The given set name is duplicated in the policy file.

User response: Ensure each set name appears only once in the policy file.

6027-2826 **The encryption set '***setName***' requested by rule '***rule***' could not be found.**

Explanation: The given set name used in the rule cannot be found.

User response: Verify if the set name is correct. Add the given set if it is missing from the policy.

6027-2827 [E] **Error in evaluation of encryption policy for file** *fileName***:** *%s*

Explanation: An error occurred while evaluating the encryption rules in the given policy file.

User response: Examine the other error messages produced while evaluating the policy file.

6027-2828 [E] **Encryption not supported on Windows. Encrypted file systems are not allowed when Windows nodes are present in the cluster.**

Explanation: Self-explanatory.

User response: To activate encryption, ensure there are

no Windows nodes in the cluster.

6027-2950 [E] Trace value '*value*' after class '*class*' must be from 0 to 14.

Explanation: The specified trace value is not recognized.

User response: Specify a valid trace integer value.

6027-2951 [W] Value *value* for worker1Threads must be <= than the original setting *value*

Explanation: An attempt to dynamically set **worker1Threads** found the value out of range. The dynamic value must be 2 <= *value* <= the original setting when the GPFS daemon was started.

6027-2952 [E] Unknown assert class '*assertClass*'.

Explanation: The assert class is not recognized.

User response: Specify a valid assert class.

6027-2953 [E] Non-numeric assert value '*value*' after class '*class*'.

Explanation: The specified assert value is not recognized.

User response: Specify a valid assert integer value.

6027-2954 [E] Assert value '*value*' after class '*class*' must be from 0 to 127.

Explanation: The specified assert value is not recognized.

User response: Specify a valid assert integer value.

6027-2955 [W] Time-of-day may have jumped back. Late by *delaySeconds* seconds to wake certain threads.

Explanation: Time-of-day may have jumped back, which has resulted in some threads being awakened later than expected. It is also possible that some other factor has caused a delay in waking up the threads.

User response: Verify if there is any problem with network time synchronization, or if time-of-day is being incorrectly set.

6027-2956 [E] Invalid crypto engine type (encryptionCryptoEngineType): *cryptoEngineType*.

Explanation: The specified value for **encryptionCryptoEngineType** is incorrect.

User response: Specify a valid value for **encryptionCryptoEngineType**.

6027-2957 [E] Invalid cluster manager selection choice (clusterManagerSelection): *clusterManagerSelection*.

Explanation: The specified value for **clusterManagerSelection** is incorrect.

User response: Specify a valid value for **clusterManagerSelection**.

6027-2958 [E] Invalid NIST compliance type (nistCompliance): *nistComplianceValue*.

Explanation: The specified value for **nistCompliance** is incorrect.

User response: Specify a valid value for **nistCompliance**.

6027-2959 [E] The CPU architecture on this node does not support tracing in *traceMode* mode. Switching to *traceMode* mode.

Explanation: The CPU does not have constant time stamp counter capability required for overwrite trace mode. The trace has been enabled in blocking mode.

User response: Update configuration parameters to use trace facility in blocking mode or replace this node with modern CPU architecture.

6027-2960 [W] Unable to establish a session with Active Directory server for the domain '*domainServer*'. ID mapping through Microsoft Identity Management for Unix will be unavailable.

Explanation: GPFS tried to establish an LDAP session with the specified Active Directory server but was unable to do so.

User response: Ensure that the specified domain controller is available.

6027-2961 [I] Established a session with Active Directory server for the domain '*domainServer*'.

Explanation: GPFS was able to successfully establish an LDAP session with the specified Active Directory server.

User response: None.

6027-3101 Pdisk rotation rate invalid in option '*option*'.

Explanation: When parsing disk lists, the pdisk rotation rate is not valid.

User response: Specify a valid rotation rate (SSD, NVRAM, or 1025 through 65535).

6027-3102 **Pdisk FRU number too long in option '*option*', maximum length *length*.**

Explanation: When parsing disk lists, the pdisk FRU number is too long.

User response: Specify a valid FRU number that is shorter than or equal to the maximum length.

6027-3103 **Pdisk location too long in option '*option*', maximum length *length*.**

Explanation: When parsing disk lists, the pdisk location is too long.

User response: Specify a valid location that is shorter than or equal to the maximum length.

6027-3104 **Pdisk failure domains too long in option '*name1name2*', maximum length *name3*.**

Explanation: When parsing disk lists, the pdisk failure domains are too long.

User response: Specify valid failure domains, shorter than the maximum.

6027-3105 **Pdisk nPathActive invalid in option '*option*'.**

Explanation: When parsing disk lists, the **nPathActive** value is not valid.

User response: Specify a valid **nPathActive** value (0 to 255).

6027-3106 **Pdisk nPathTotal invalid in option '*option*'.**

Explanation: When parsing disk lists, the **nPathTotal** value is not valid.

User response: Specify a valid *nPathTotal* value (0 to 255).

6027-3107 **Pdisk nsdFormatVersion invalid in option '*name1name2*'.**

Explanation: The nsdFormatVersion that is entered while parsing the disk is invalid.

User response: Specify valid nsdFormatVersion, 1 or 2.

6027-3108 **Declustered array name *name1* appears more than once in the declustered array stanzas.**

Explanation: when parsing declustered array lists a duplicate name is found.

User response: Remove duplicate MSG_PARSE_DUPNAME which is not documented.

6027-3200 **AFM ERROR: command *pCacheCmd* fileset *filesetName* fileids [*parentId.childId.tParentId.targetId,ReqCmd*] original error *oerr* application error *aerr* remote error *remoteError***

Explanation: AFM operations on a particular file failed.

User response: For asynchronous operations that are requeued, run the **mmafmctl** command with the **resumeRequeued** option after fixing the problem at the home cluster.

6027-3201 **AFM ERROR DETAILS: type: *remoteCmdType* snapshot name *snapshotName* snapshot ID *snapshotId***

Explanation: Peer snapshot creation or deletion failed.

User response: Fix snapshot creation or deletion error.

6027-3204 **AFM: Failed to set xattr on inode *inodeNum* error *err*, ignoring.**

Explanation: Setting extended attributes on an inode failed.

User response: None.

6027-3205 **AFM: Failed to get xattrs for inode *inodeNum*, ignoring.**

Explanation: Getting extended attributes on an inode failed.

User response: None.

6027-3209 **Home NFS mount of *host*:*path* failed with error *err***

Explanation: NFS mounting of path from the home cluster failed.

User response: Make sure the exported path can be mounted over NFSv3.

6027-3210 **Cannot find AFM control file for fileset *filesetName* in the exported file system at home. ACLs and extended attributes will not be synchronized. Sparse files will have zeros written for holes.**

Explanation: Either home path does not belong to GPFS, or the AFM control file is not present in the exported path.

User response: If the exported path belongs to a GPFS file system, run the **mmafmconfig** command with the **enable** option on the export path at home.

6027-3211 **Change in home export detected. Caching will be disabled.**

Explanation: A change in home export was detected or the home path is stale.

User response: Ensure the exported path is accessible.

6027-3212 **AFM ERROR: Cannot enable AFM for fileset** *filesetName* **(error** *err***)**

Explanation: AFM was not enabled for the fileset because the root file handle was modified, or the remote path is stale.

User response: Ensure the remote export path is accessible for NFS mount.

6027-3213 **Cannot find snapshot link directory name for exported file system at home for fileset** *filesetName***. Snapshot directory at home will be cached.**

Explanation: Unable to determine the snapshot directory at the home cluster.

User response: None.

6027-3214 [E] **AFM: Unexpiration of fileset** *filesetName* **failed with error** *err***. Use mmafmctl to manually unexpire the fileset.**

Explanation: Unexpiration of fileset failed after a home reconnect.

User response: Run the **mmafmctl** command with the **unexpire** option on the fileset.

6027-3215 [W] **AFM: Peer snapshot delayed due to long running execution of operation to remote cluster for fileset** *filesetName***. Peer snapshot continuing to wait.**

Explanation: Peer snapshot command timed out waiting to flush messages.

User response: None.

6027-3216 **Fileset** *filesetName* **encountered an error synchronizing with the remote cluster. Cannot synchronize with the remote cluster until AFM recovery is executed.**

Explanation: Cache failed to synchronize with home because of an out of memory or conflict error. Recovery, resynchronization, or both will be performed by GPFS to synchronize cache with the home.

User response: None.

6027-3217 **AFM ERROR Unable to unmount NFS export for fileset** *filesetName*

Explanation: NFS unmount of the path failed.

User response: None.

6027-3220 **AFM: Home NFS mount of** *host***:***path* **failed with error** *err* **for file system** *fileSystem* **fileset id** *filesetName***. Caching will be disabled and the mount will be tried again after** *mountRetryTime* **seconds, on next request to gateway**

Explanation: NFS mount of the home cluster failed. The mount will be tried again after *mountRetryTime* seconds.

User response: Make sure the exported path can be mounted over NFSv3.

6027-3221 **AFM: Home NFS mount of** *host***:***path* **succeeded for file system** *fileSystem* **fileset** *filesetName***. Caching is enabled.**

Explanation: NFS mount of the path from the home cluster succeeded. Caching is enabled.

User response: None.

6027-3224 [I] **AFM: Failed to set extended attributes on file system** *fileSystem* **inode** *inodeNum* **error** *err***, ignoring.**

Explanation: Setting extended attributes on an inode failed.

User response: None.

6027-3225 [I] **AFM: Failed to get extended attributes for file system** *fileSystem* **inode** *inodeNum***, ignoring.**

Explanation: Getting extended attributes on an inode failed.

User response: None.

6027-3226 [I] **AFM: Cannot find control file for file system** *fileSystem* **fileset** *filesetName* **in the exported file system at home. ACLs and extended attributes will not be synchronized. Sparse files will have zeros written for holes.**

Explanation: Either the home path does not belong to GPFS, or the AFM control file is not present in the exported path.

User response: If the exported path belongs to a GPFS file system, run the **mmafmconfig** command with the **enable** option on the export path at home.

6027-3227 [E] AFM: Cannot enable AFM for file system *fileSystem* **fileset** *filesetName* **(error** *err***)**

Explanation: AFM was not enabled for the fileset because the root file handle was modified, or the remote path is stale.

User response: Ensure the remote export path is accessible for NFS mount.

6027-3228 [E] AFM: Unable to unmount NFS export for file system *fileSystem* **fileset** *filesetName*

Explanation: NFS unmount of the path failed.

User response: None.

6027-3229 [E] AFM: File system *fileSystem* **fileset** *filesetName* **encountered an error synchronizing with the remote cluster. Cannot synchronize with the remote cluster until AFM recovery is executed.**

Explanation: The cache failed to synchronize with home because of an out of memory or conflict error. Recovery, resynchronization, or both will be performed by GPFS to synchronize the cache with the home.

User response: None.

6027-3230 [I] AFM: Cannot find snapshot link directory name for exported file system at home for file system *fileSystem* **fileset** *filesetName***. Snapshot directory at home will be cached.**

Explanation: Unable to determine the snapshot directory at the home cluster.

User response: None.

6027-3232 *type* **AFM:** *pCacheCmd* **file system** *fileSystem* **fileset** *filesetName* **file IDs [***parentId***.***childId***.***tParentId***.***targetId***,***flag***] name** *sourceName origin* **error** *err*

Explanation: AFM operations on a particular file failed.

User response: For asynchronous operations that are requeued, run the **mmafmctl** command with the **resumeRequeued** option after fixing the problem at the home cluster.

6027-3233 [I] AFM: Previous error repeated *repeatNum* **times.**

Explanation: Multiple AFM operations have failed.

User response: None.

6027-3234 [E] AFM: Unable to start thread to unexpire filesets.

Explanation: Failed to start thread for unexpiration of fileset.

User response: None.

6027-3235 [I] AFM: Stopping recovery for the file system *fileSystem* **fileset** *filesetName*

Explanation: AFM recovery terminated because the current node is no longer MDS for the fileset.

User response: None.

6027-3236 [E] AFM: Recovery on file system *fileSystem* **fileset** *filesetName* **failed with error** *err***. Recovery will be retried on next access after recovery retry interval (***timeout* **seconds) or manually resolve known problems and recover the fileset.**

Explanation: Recovery failed to complete on the fileset. The next access will restart recovery.

Explanation: AFM recovery failed. Fileset will be temporarily put into dropped state and will be recovered on accessing fileset after timeout mentioned in the error message. User can recover the fileset manually by running **mmafmctl** command with **recover** option after rectifying any known errors leading to failure.

User response: None.

6027-3239 [E] AFM: Remote command *remoteCmdType* **on file system** *fileSystem* **snapshot** *snapshotName* **snapshot ID** *snapshotId* **failed.**

Explanation: A failure occurred when creating or deleting a peer snapshot.

User response: Examine the error details and retry the operation.

6027-3240 [E] AFM: *pCacheCmd* **file system** *fileSystem* **fileset** *filesetName* **file IDs [***parentId***.***childId***.***tParentId***.***targetId***,***flag***] error** *err*

Explanation: Operation failed to execute on home in independent-writer mode.

User response: None.

6027-3241 [I] AFM: GW queue transfer started for file system *fileSystem* **fileset** *filesetName***. Transferring to** *nodeAddress***.**

Explanation: An old GW initiated the queue transfer because a new GW node joined the cluster, and the

fileset now belongs to the new GW node.

User response: None.

6027-3242 [I] AFM: GW queue transfer started for file system *fileSystem* **fileset** *filesetName***. Receiving from** *nodeAddress***.**

Explanation: An old MDS initiated the queue transfer because this node joined the cluster as GW and the fileset now belongs to this node.

User response: None.

6027-3243 [I] AFM: GW queue transfer completed for file system *fileSystem* **fileset** *filesetName***. error** *error*

Explanation: A GW queue transfer completed.

User response: None.

6027-3244 [I] AFM: Home mount of *afmTarget* **succeeded for file system** *fileSystem* **fileset** *filesetName***. Caching is enabled.**

Explanation: A mount of the path from the home cluster succeeded. Caching is enabled.

User response: None.

6027-3245 [E] AFM: Home mount of *afmTarget* **failed with error** *error* **for file system** *fileSystem* **fileset ID** *filesetName***. Caching will be disabled and the mount will be tried again after** *mountRetryTime* **seconds, on the next request to the gateway.**

Explanation: A mount of the home cluster failed. The mount will be tried again after *mountRetryTime* seconds.

User response: Verify that the *afmTarget* can be mounted using the specified protocol.

6027-3246 [I] AFM: Prefetch recovery started for the file system *fileSystem* **fileset** *filesetName***.**

Explanation: Prefetch recovery started.

User response: None.

6027-3247 [I] AFM: Prefetch recovery completed for the file system *fileSystem* **fileset** *filesetName***. error** *error*

Explanation: Prefetch recovery completed.

User response: None.

6027-3248 [E] AFM: Cannot find the control file for fileset *filesetName* **in the exported file system at home. This file is required to operate in primary mode. The fileset will be disabled.**

Explanation: Either the home path does not belong to GPFS, or the AFM control file is not present in the exported path.

User response: If the exported path belongs to a GPFS file system, run the **mmafmconfig** command with the enable option on the export path at home.

6027-3249 [E] AFM: Target for fileset *filesetName* **is not a secondary-mode fileset or file system. This is required to operate in primary mode. The fileset will be disabled.**

Explanation: The AFM target is not a secondary fileset or file system.

User response: The AFM target fileset or file system should be converted to secondary mode.

6027-3250 [E] AFM: Refresh intervals cannot be set for fileset.

Explanation: Refresh intervals are not supported on primary and secondary-mode filesets.

User response: None.

6027-3252 [I] AFM: Home has been restored for cache *filesetName***. Synchronization with home will be resumed.**

Explanation: A change in home export was detected that caused the home to be restored. Synchronization with home will be resumed.

User response: None.

6027-3253 [E] AFM: Change in home is detected for cache *filesetName***. Synchronization with home is suspended until the problem is resolved.**

Explanation: A change in home export was detected or the home path is stale.

User response: Ensure the exported path is accessible.

6027-3254 [W] AFM: Home is taking longer than expected to respond for cache *filesetName***. Synchronization with home is temporarily suspended.**

Explanation: A pending message from gateway node to home is taking longer than expected to respond. This could be the result of a network issue or a problem at the home site.

User response: Ensure the exported path is accessible.

6027-3255 [E] AFM: Target for fileset *filesetName* **is a secondary-mode fileset or file system. Only a primary-mode, read-only or local-update mode fileset can operate on a secondary-mode fileset. The fileset will be disabled.**

Explanation: The AFM target is a secondary fileset or file system. Only a primary-mode, read-only, or local-update fileset can operate on a secondary-mode fileset.

User response: Use a secondary-mode fileset as the target for the primary-mode, read-only or local-update mode fileset.

6027-3256 [I] AFM: The RPO peer snapshot was missed for file system *fileSystem* **fileset** *filesetName***.**

Explanation: The periodic RPO peer snapshot was not taken in time for the primary fileset.

User response: None.

6027-3257 [E] AFM: Unable to start thread to verify primary filesets for RPO.

Explanation: Failed to start thread for verification of primary filesets for RPO.

User response: None.

6027-3258 [I] [I] AFM: AFM is not enabled for fileset *filesetName* **in filesystem** *fileSystem***. Some features will not be supported, see documentation for enabling AFM and unsupported features.**

Explanation: Either the home path does not belong to GPFS, or the AFM control file is not present in the exported path.

User response: If the exported path belongs to a GPFS file system, run the **mmafmconfig** command with the enable option on the export path at home.

6027-3259 [I] [I] AFM: Remote file system *fileSystem* **is panicked due to unresponsive messages on fileset** *filesetName***, re-mount the file system after it becomes responsive.**

Explanation: *SGPanic* is triggered if remote file system is unresponsive to bail out inflight messages that are stuck in the queue.

User response: Re-mount the remote file system when it is responsive.

6027-3300 Attribute afmShowHomeSnapshot cannot be changed for a single-writer fileset.

Explanation: Changing **afmShowHomeSnapshot** is not supported for single-writer filesets.

User response: None.

6027-3301 Unable to quiesce all nodes; some processes are busy or holding required resources.

Explanation: A timeout occurred on one or more nodes while trying to quiesce the file system during a snapshot command.

User response: Check the GPFS log on the file system manager node.

6027-3302 Attribute afmShowHomeSnapshot cannot be changed for a *afmMode* **fileset.**

Explanation: Changing **afmShowHomeSnapshot** is not supported for single-writer or independent-writer filesets.

User response: None.

6027-3303 Cannot restore snapshot; quota management is active for *fileSystem***.**

Explanation: File system quota management is still active. The file system must be unmounted when restoring global snapshots.

User response: Unmount the file system and reissue the restore command.

6027-3304 Attention: Disk space reclaim on *number* **of** *number* **regions in** *fileSystem* **returned errors.**

Explanation: Free disk space reclaims on some regions failed during tsreclaim run. Typically this is due to the lack of space reclaim support by the disk controller or operating system. It may also be due to utilities such as **mmdefragfs** or **mmfsck** running concurrently.

User response: Verify that the disk controllers and the operating systems in the cluster support thin-provisioning space reclaim. Or, rerun the **mmfsctl reclaimSpace** command after **mmdefragfs** or **mmfsck** completes.

6027-3305 AFM Fileset *filesetName* **cannot be changed as it is in beingDeleted state**

Explanation: The user specified a fileset to **tschfileset** that cannot be changed.

User response: None. You cannot change the attributes of the root fileset.

6027-3306 **Fileset cannot be changed because it is unlinked.**

Explanation: The fileset cannot be changed when it is unlinked.

User response: Link the fileset and then try the operation again.

6027-3307 **Fileset cannot be changed.**

Explanation: Fileset cannot be changed.

User response: None.

6027-3308 **This AFM option cannot be set for a secondary fileset.**

Explanation: This AFM option cannot be set for a secondary fileset. The fileset cannot be changed.

User response: None.

6027-3309 **The AFM attribute specified cannot be set for a primary fileset.**

Explanation: This AFM option cannot be set for a primary fileset. The fileset cannot be changed.

User response: None.

6027-3310 **A secondary fileset cannot be changed.**

Explanation: A secondary fileset cannot be changed.

User response: None.

6027-3311 **A primary fileset cannot be changed.**

Explanation: A primary fileset cannot be changed.

User response: None.

6027-3312 **No inode was found matching the criteria.**

Explanation: No inode was found matching the criteria.

User response: None.

6027-3313 **File system scan RESTARTED due to resume of all disks being emptied.**

Explanation: The parallel inode traversal (PIT) phase is restarted with a file system restripe.

User response: None.

6027-3314 **File system scan RESTARTED due to new disks to be emptied.**

Explanation: The file system restripe was restarted after a new disk was suspended.

User response: None.

6027-3315 **File system scan CANCELLED due to new disks to be emptied or resume of all disks being emptied.**

Explanation: The parallel inode traversal (PIT) phase is cancelled during the file system restripe.

User response: None.

6027-3316 **Unable to create file system because there is not enough space for the log files. Number of log files:** *numberOfLogFiles*. **Log file size:** *logFileSize*. **Change one or more of the following as suggested and try again:**

Explanation: There is not enough space available to create all the required log files. This can happen when the storage pool is not large enough.

User response: Refer to the details given and correct the file system parameters.

6027-3317 **Warning: file system is not 4K aligned due to small** *reasonString*. **Native 4K sector disks cannot be added to this file system unless the disk that is used is dataOnly and the data block size is at least 128K.**

Explanation: The file system is created with a small inode or block size. Native 4K sector disk cannot be added to the file system, unless the disk that is used is dataOnly and the data block size is at least 128K.

User response: None.

6027-3318 **Fileset** *filesetName* **cannot be deleted as it is in compliant mode and it contains user files.**

Explanation: An attempt was made to delete a non-empty fileset that is in compliant mode.

User response: None.

6027-3319 **The AFM attribute** *optionName* **cannot be set for a primary fileset.**

Explanation: This AFM option cannot be set for a primary fileset. Hence, the fileset cannot be changed.

User response: None.

6027-3320 *commandName*: **indefiniteRetentionProtection is enabled. File system cannot be deleted.**

Explanation: Indefinite retention is enabled for the file system so it cannot be deleted.

User response: None.

6027-3321 **Snapshot** *snapshotName* **is an internal pcache recovery snapshot and cannot be deleted by user.**

Explanation: The snapshot cannot be deleted by user as it is an internal pcache recovery snapshot.

User response: None.

6027-3400 **Attention: The file system is at risk. The specified replication factor does not tolerate unavailable metadata disks.**

Explanation: The default metadata replication was reduced to one while there were unavailable, or stopped, metadata disks. This condition prevents future file system manager takeover.

User response: Change the default metadata replication, or delete unavailable disks if possible.

6027-3401 **Failure group** *value* **for disk** *diskName* **is not valid.**

Explanation: An explicit failure group must be specified for each disk that belongs to a write affinity enabled storage pool.

User response: Specify a valid failure group.

6027-3402 [X] **An unexpected device mapper path** *dmDevice* **(***nsdId***) was detected. The new path does not have Persistent Reserve enabled. The local access to disk** *diskName* **will be marked as down.**

Explanation: A new device mapper path was detected, or a previously failed path was activated after the local device discovery was finished. This path lacks a Persistent Reserve and cannot be used. All device paths must be active at mount time.

User response: Check the paths to all disks in the file system. Repair any failed paths to disks then rediscover the local disk access.

6027-3404 [E] **The current file system version does not support write caching.**

Explanation: The current file system version does not allow the write caching option.

User response: Use **mmchfs -V** to convert the file system to version 14.04 (4.1.0.0) or higher and reissue the command.

6027-3405 [E] **Cannot change the rapid repair, \"***fileSystemName***\" is mounted on** *number* **node(s).**

Explanation: Rapid repair can only be changed on unmounted file systems.

User response: Unmount the file system before running this command.

6027-3406 **Error: Cannot add 4K native dataOnly disk** *diskName* **to non-4K aligned file system unless the file system version is at least 4.1.1.4.**

Explanation: An attempt was made through the mmadddisk command to add a 4K native disk to a non-4K aligned file system while the file system version is not at 4.1.1.4 or later.

User response: Upgrade the file system to 4.1.1.4 or later, and then retry the command.

6027-3407 [E] **Disk failure. Volume** *name***. rc =** *value***, and physical volume** *name***.**

Explanation: An I/O request to a disk or a request to fence a disk is failed in such a manner that GPFS can no longer use the disk.

User response: Check the disk hardware and the software subsystems in the path to the disk.

6027-3408 [X] **File System** *fileSystemName* **unmounted by the system with return code** *value***, reason code** *value***, at line** *value* **in** *name***.**

Explanation: Console log entry caused by a forced unmount due to a problem such as disk or communication failure.

User response: Correct the underlying problem and remount the file system.

6027-3450 **Error** *errorNumber* **when purging key (file system** *fileSystem***). Key name format possibly incorrect.**

Explanation: An error was encountered when purging a key from the key cache. The specified key name might have been incorrect, or an internal error was encountered.

User response: Ensure that the key name specified in the command is correct.

6027-3451 Error *errorNumber* **when emptying cache (file system** *fileSystem***).**

Explanation: An error was encountered when purging all the keys from the key cache.

User response: Contact the IBM Support Center.

6027-3452 [E] Unable to create encrypted file *fileName* **(inode** *inodeNumber***, fileset** *filesetNumber***, file system** *fileSystem***).**

Explanation: Unable to create a new encrypted file. The key required to encrypt the file might not be available.

User response: Examine the error message following this message for information on the specific failure.

6027-3453 [E] Unable to open encrypted file: inode *inodeNumber***, fileset** *filesetNumber***, file system** *fileSystem***.**

Explanation: Unable to open an existing encrypted file. The key used to encrypt the file might not be available.

User response: Examine the error message following this message for information on the specific failure.

6027-3457 [E] Unable to rewrap key with name *Keyname* **(inode** *inodeNumber***, fileset** *filesetNumber***, file system** *fileSystem***).**

Explanation: Unable to rewrap the key for a specified file because of an error with the key name.

User response: Examine the error message following this message for information on the specific failure.

6027-3458 [E] Invalid length for the Keyname string.

Explanation: The *Keyname* string has an incorrect length. The length of the specified string was either zero or it was larger than the maximum allowed length.

User response: Verify the *Keyname* string.

6027-3459 [E] Not enough memory.

Explanation: Unable to allocate memory for the *Keyname* string.

User response: Restart GPFS. Contact the IBM Support Center.

6027-3460 [E] Incorrect format for the Keyname string.

Explanation: An incorrect format was used when specifying the *Keyname* string.

User response: Verify the format of the *Keyname* string.

6027-3461 [E] Error code: *errorNumber***.**

Explanation: An error occurred when processing a key ID.

User response: Contact the IBM Support Center.

6027-3462 [E] Unable to rewrap key: original key name: *originalKeyname***, new key name:** *newKeyname* **(inode** *inodeNumber***, fileset** *filesetNumber***, file system** *fileSystem***).**

Explanation: Unable to rewrap the key for a specified file, possibly because the existing key or the new key cannot be retrieved from the key server.

User response: Examine the error message following this message for information on the specific failure.

6027-3463 [E] Rewrap error.

Explanation: An internal error occurred during key rewrap.

User response: Examine the error messages surrounding this message. Contact the IBM Support Center.

6027-3464 [E] New key is already in use.

Explanation: The new key specified in a key rewrap is already being used.

User response: Ensure that the new key specified in the key rewrap is not being used by the file.

6027-3465 [E] Cannot retrieve original key.

Explanation: Original key being used by the file cannot be retrieved from the key server.

User response: Verify that the key server is available, the credentials to access the key server are correct, and that the key is defined on the key server.

6027-3466 [E] Cannot retrieve new key.

Explanation: Unable to retrieve the new key specified in the rewrap from the key server.

User response: Verify that the key server is available, the credentials to access the key server are correct, and that the key is defined on the key server.

6027-3468 [E] Rewrap error code *errorNumber***.**

Explanation: Key rewrap failed.

User response: Record the error code and contact the IBM Support Center.

6027-3469 [E] Encryption is enabled but the crypto module could not be initialized. Error code: *number*. Ensure that the GPFS crypto package was installed.

Explanation: Encryption is enabled, but the cryptographic module required for encryption could not be loaded.

User response: Ensure that the packages required for encryption are installed on each node in the cluster.

6027-3470 [E] Cannot create file *fileName*: extended attribute is too large: *numBytesRequired* bytes (*numBytesAvailable* available) (fileset *filesetNumber*, file system *fileSystem*).

Explanation: Unable to create an encryption file because the extended attribute required for encryption is too large.

User response: Change the encryption policy so that the file key is wrapped fewer times, reduce the number of keys used to wrap a file key, or create a file system with a larger inode size.

6027-3471 [E] At least one key must be specified.

Explanation: No key name was specified.

User response: Specify at least one key name.

6027-3472 [E] Could not combine the keys.

Explanation: Unable to combine the keys used to wrap a file key.

User response: Examine the keys being used. Contact the IBM Support Center.

6027-3473 [E] Could not locate the RKM.conf file.

Explanation: Unable to locate the `RKM.conf` configuration file.

User response: Contact the IBM Support Center.

6027-3474 [E] Could not open *fileType* file ('*fileName*' was specified).

Explanation: Unable to open the specified configuration file. Encryption files will not be accessible.

User response: Ensure that the specified configuration file is present on all nodes.

6027-3475 [E] Could not read file '*fileName*'.

Explanation: Unable to read the specified file.

User response: Ensure that the specified file is accessible from the node.

6027-3476 [E] Could not seek through file '*fileName*'.

Explanation: Unable to seek through the specified file. Possible inconsistency in the local file system where the file is stored.

User response: Ensure that the specified file can be read from the local node.

6027-3477 [E] Could not wrap the FEK.

Explanation: Unable to wrap the file encryption key.

User response: Examine other error messages. Verify that the encryption policies being used are correct.

6027-3478 [E] Insufficient memory.

Explanation: Internal error: unable to allocate memory.

User response: Restart GPFS. Contact the IBM Support Center.

6027-3479 [E] Missing combine parameter string.

Explanation: The combine parameter string was not specified in the encryption policy.

User response: Verify the syntax of the encryption policy.

6027-3480 [E] Missing encryption parameter string.

Explanation: The encryption parameter string was not specified in the encryption policy.

User response: Verify the syntax of the encryption policy.

6027-3481 [E] Missing wrapping parameter string.

Explanation: The wrapping parameter string was not specified in the encryption policy.

User response: Verify the syntax of the encryption policy.

6027-3482 [E] '*combineParameter*' could not be parsed as a valid combine parameter string.

Explanation: Unable to parse the combine parameter string.

User response: Verify the syntax of the encryption policy.

6027-3483 [E] '*encryptionParameter*' could not be parsed as a valid encryption parameter string.

Explanation: Unable to parse the encryption parameter string.

User response: Verify the syntax of the encryption policy.

6027-3484 [E] '*wrappingParameter*' could not be parsed as a valid wrapping parameter string.

Explanation: Unable to parse the wrapping parameter string.

User response: Verify the syntax of the encryption policy.

6027-3485 [E] The Keyname string cannot be longer than *number* characters.

Explanation: The specified *Keyname* string has too many characters.

User response: Verify that the specified *Keyname* string is correct.

6027-3486 [E] The KMIP library could not be initialized.

Explanation: The KMIP library used to communicate with the key server could not be initialized.

User response: Restart GPFS. Contact the IBM Support Center.

6027-3487 [E] The RKM ID cannot be longer than *number* characters.

Explanation: The remote key manager ID cannot be longer than the specified length.

User response: Use a shorter remote key manager ID.

6027-3488 [E] The length of the key ID cannot be zero.

Explanation: The length of the specified key ID string cannot be zero.

User response: Specify a key ID string with a valid length.

6027-3489 [E] The length of the RKM ID cannot be zero.

Explanation: The length of the specified RKM ID string cannot be zero.

User response: Specify an RKM ID string with a valid length.

6027-3490 [E] The maximum size of the RKM.conf file currently supported is *number* bytes.

Explanation: The `RKM.conf` file is larger than the size that is currently supported.

User response: User a smaller `RKM.conf` configuration file.

6027-3491 [E] The string '*Keyname*' could not be parsed as a valid key name.

Explanation: The specified string could not be parsed as a valid key name.

User response: Specify a valid *Keyname* string.

6027-3493 [E] *numKeys* keys were specified but a maximum of *numKeysMax* is supported.

Explanation: The maximum number of specified key IDs was exceeded.

User response: Change the encryption policy to use fewer keys.

6027-3494 [E] Unrecognized cipher mode.

Explanation: Unable to recognize the specified cipher mode.

User response: Specify one of the valid cipher modes.

6027-3495 [E] Unrecognized cipher.

Explanation: Unable to recognize the specified cipher.

User response: Specify one of the valid ciphers.

6027-3496 [E] Unrecognized combine mode.

Explanation: Unable to recognize the specified combine mode.

User response: Specify one of the valid combine modes.

6027-3497 [E] Unrecognized encryption mode.

Explanation: Unable to recognize the specified encryption mode.

User response: Specify one of the valid encryption modes.

6027-3498 [E] Invalid key length.

Explanation: An invalid key length was specified.

User response: Specify a valid key length for the chosen cipher mode.

6027-3499 [E] Unrecognized wrapping mode.

Explanation: Unable to recognize the specified wrapping mode.

User response: Specify one of the valid wrapping modes.

6027-3500 [E] Duplicate Keyname string '*keyIdentifier*'.

Explanation: A given *Keyname* string has been specified twice.

User response: Change the encryption policy to eliminate the duplicate.

6027-3501 [E] Unrecognized combine mode ('*combineMode*').

Explanation: The specified combine mode was not recognized.

User response: Specify a valid combine mode.

6027-3502 [E] Unrecognized cipher mode ('*cipherMode*').

Explanation: The specified cipher mode was not recognized.

User response: Specify a valid cipher mode.

6027-3503 [E] Unrecognized cipher ('*cipher*').

Explanation: The specified cipher was not recognized.

User response: Specify a valid cipher.

6027-3504 [E] Unrecognized encryption mode ('*mode*').

Explanation: The specified encryption mode was not recognized.

User response: Specify a valid encryption mode.

6027-3505 [E] Invalid key length ('*keyLength*').

Explanation: The specified key length was incorrect.

User response: Specify a valid key length.

6027-3506 [E] Mode '*mode1*' is not compatible with mode '*mode2*', aborting.

Explanation: The two specified encryption parameters are not compatible.

User response: Change the encryption policy and specify compatible encryption parameters.

6027-3509 [E] Key '*keyID*:*RKMID*' could not be fetched (RKM reported error *errorNumber*).

Explanation: The key with the specified name cannot be fetched from the key server.

User response: Examine the error messages to obtain information about the failure. Verify connectivity to the key server and that the specified key is present at the server.

6027-3510 [E] Could not bind symbol *symbolName* (*errorDescription*).

Explanation: Unable to find the location of a symbol in the library.

User response: Contact the IBM Support Center.

6027-3512 [E] The specified type '*type*' for backend '*backend*' is invalid.

Explanation: An incorrect type was specified for a key server backend.

User response: Specify a correct backend type in `RKM.conf`.

6027-3513 [E] Duplicate backend '*backend*'.

Explanation: A duplicate backend name was specified in `RKM.conf`.

User response: Specify unique RKM backends in `RKM.conf`.

6027-3517 [E] Could not open library (*libName*).

Explanation: Unable to open the specified library.

User response: Verify that all required packages are installed for encryption. Contact the IBM Support Center.

6027-3518 [E] The length of the RKM ID string is invalid (must be between 0 and *length* characters).

Explanation: The length of the RKM backend ID is invalid.

User response: Specify an RKM backend ID with a valid length.

6027-3519 [E] '*numAttempts*' is not a valid number of connection attempts.

Explanation: The value specified for the number of connection attempts is incorrect.

User response: Specify a valid number of connection attempts.

6027-3520 [E] '*sleepInterval*' is not a valid sleep interval.

Explanation: The value specified for the sleep interval is incorrect.

User response: Specify a valid sleep interval value (in microseconds).

6027-3521 [E] '*timeout*' is not a valid connection timeout.

Explanation: The value specified for the connection timeout is incorrect.

User response: Specify a valid connection timeout (in seconds).

6027-3522 [E] '*url*' is not a valid URL.

Explanation: The specified string is not a valid URL for the key server.

User response: Specify a valid URL for the key server.

6027-3524 [E] '*tenantName*' is not a valid tenantName.

Explanation: An incorrect value was specified for the tenant name.

User response: Specify a valid tenant name.

6027-3527 [E] Backend '*backend*' could not be initialized (error *errorNumber*).

Explanation: Key server backend could not be initialized.

User response: Examine the error messages. Verify connectivity to the server. Contact the IBM Support Center.

6027-3528 [E] Unrecognized wrapping mode ('*wrapMode*').

Explanation: The specified key wrapping mode was not recognized.

User response: Specify a valid key wrapping mode.

6027-3529 [E] An error was encountered while processing file '*fileName*':

Explanation: An error was encountered while processing the specified configuration file.

User response: Examine the error messages that follow and correct the corresponding conditions.

6027-3530 [E] Unable to open encrypted file: key retrieval not initialized (inode *inodeNumber*, fileset *filesetNumber*, file system *fileSystem*).

Explanation: File is encrypted but the infrastructure required to retrieve encryption keys was not initialized, likely because processing of `RKM.conf` failed.

User response: Examine error messages at the time the file system was mounted.

6027-3533 [E] Invalid encryption key derivation function.

Explanation: An incorrect key derivation function was specified.

User response: Specify a valid key derivation function.

6027-3534 [E] Unrecognized encryption key derivation function ('*keyDerivation*').

Explanation: The specified key derivation function was not recognized.

User response: Specify a valid key derivation function.

6027-3535 [E] Incorrect client certificate label '*clientCertLabel*' for backend '*backend*'.

Explanation: The specified client keypair certificate label is incorrect for the backend.

User response: Ensure that the correct client certificate label is used in `RKM.conf`.

6027-3537 [E] Setting default encryption parameters requires empty combine and wrapping parameter strings.

Explanation: A non-empty combine or wrapping parameter string was used in an encryption policy rule that also uses the default parameter string.

User response: Ensure that neither the combine nor the wrapping parameter is set when the default parameter string is used in the encryption rule.

6027-3540 [E] The specified RKM backend type (*rkmType*) is invalid.

Explanation: The specified RKM type in `RKM.conf` is incorrect.

User response: Ensure that only supported RKM types are specified in `RKM.conf`.

6027-3541 [E] Encryption is not supported on Windows.

Explanation: Encryption cannot be activated if there are Windows nodes in the cluster.

User response: Ensure that encryption is not activated if there are Windows nodes in the cluster.

6027-3543 [E] The integrity of the file encrypting key could not be verified after unwrapping; the operation was cancelled.

Explanation: When opening an existing encrypted file, the integrity of the file encrypting key could not be verified. Either the cryptographic extended attributes were damaged, or the master key(s) used to unwrap the FEK have changed.

User response: Check for other symptoms of data corruption, and verify that the configuration of the key server has not changed.

6027-3545 [E] Encryption is enabled but there is no valid license. Ensure that the GPFS crypto package was installed properly.

Explanation: The required license is missing for the GPFS encryption package.

User response: Ensure that the GPFS encryption package was installed properly.

6027-3546 [E] Key '*keyID*:*rkmID*' could not be fetched. The specified RKM ID does not exist; check the RKM.conf settings.

Explanation: The specified RKM ID part of the key name does not exist, and therefore the key cannot be retrieved. The corresponding RKM might have been removed from `RKM.conf`.

User response: Check the set of RKMs specified in `RKM.conf`.

6027-3547 [E] Key '*keyID*:*rkmID*' could not be fetched. The connection was reset by the peer while performing the TLS handshake.

Explanation: The specified key could not be retrieved from the server, because the connection with the server was reset while performing the TLS handshake.

User response: Check connectivity to the server. Check credentials to access the server. Contact the IBM Support Center.

6027-3548 [E] Key '*keyID*:*rkmID*' could not be fetched. The IP address of the RKM could not be resolved.

Explanation: The specified key could not be retrieved from the server because the IP address of the server could not be resolved.

User response: Ensure that the hostname of the key server is correct. Verify whether there are problems with name resolutions.

6027-3549 [E] Key '*keyID*:*rkmID*' could not be fetched. The TCP connection with the RKM could not be established.

Explanation: Unable to establish a TCP connection with the key server.

User response: Check the connectivity to the key server.

6027-3550 Error when retrieving encryption attribute: *errorDescription*.

Explanation: Unable to retrieve or decode the encryption attribute for a given file.

User response: File could be damaged and may need to be removed if it cannot be read.

6027-3551 Error flushing work file *fileName*: *errorString*

Explanation: An error occurred while attempting to flush the named work file or socket.

User response: None.

6027-3552 Failed to fork a new process to *operationString* **file system.**

Explanation: Failed to fork a new process to suspend/resume file system.

User response: None.

6027-3553 Failed to sync fileset *filesetName*.

Explanation: Failed to sync fileset.

User response: None.

6027-3554 The restore command encountered an out-of-memory error.

Explanation: The fileset snapshot restore command encountered an out-of-memory error.

User response: None.

6027-3555 *name* **must be combined with FileInherit, DirInherit or both.**

Explanation: **NoPropagateInherit** must be accompanied by other inherit flags. Valid values are **FileInherit** and **DirInherit**.

User response: Specify a valid NFSv4 option and reissue the command.

6027-3556 *cmdName* **error: insufficient memory.**

Explanation: The command exhausted virtual memory.

User response: Consider some of the command parameters that might affect memory usage. Contact the IBM Support Center.

6027-3557 *cmdName* **error: could not create a temporary file.**

Explanation: A temporary file could not be created in the current directory.

User response: Ensure that the file system is not full and that files can be created. Contact the IBM Support Center.

6027-3558 *cmdName* **error: could not initialize the key management subsystem (error** *returnCode***).**

Explanation: An internal component of the cryptographic library could not be properly initialized.

User response: Ensure that the `gpfs.gskit` package was installed properly. Contact the IBM Support Center.

6027-3559 *cmdName* **error: could not create the key database (error** *returnCode***).**

Explanation: The key database file could not be created.

User response: Ensure that the file system is not full and that files can be created. Contact the IBM Support Center.

6027-3560 *cmdName* **error: could not create the new self-signed certificate (error** *returnCode***).**

Explanation: A new certificate could not be successfully created.

User response: Ensure that the supplied canonical name is valid. Contact the IBM Support Center.

6027-3561 *cmdName* **error: could not extract the key item (error** *returnCode***).**

Explanation: The public key item could not be extracted successfully.

User response: Contact the IBM Support Center.

6027-3562 *cmdName* **error: base64 conversion failed (error** *returnCode***).**

Explanation: The conversion from or to the **BASE64** encoding could not be performed successfully.

User response: Contact the IBM Support Center.

6027-3563 *cmdName* **error: could not extract the private key (error** *returnCode***).**

Explanation: The private key could not be extracted successfully.

User response: Contact the IBM Support Center.

6027-3564 *cmdName* **error: could not initialize the ICC subsystem (error** *returnCode returnCode***).**

Explanation: An internal component of the cryptographic library could not be properly initialized.

User response: Ensure that the `gpfs.gskit` package was installed properly. Contact the IBM Support Center.

6027-3565 *cmdName* **error: I/O error.**

Explanation: A terminal failure occurred while performing I/O.

User response: Contact the IBM Support Center.

6027-3566 *cmdName* **error: could not open file '***fileName***'.**

Explanation: The specified file could not be opened.

User response: Ensure that the specified path and file name are correct and that you have sufficient permissions to access the file.

6027-3567 *cmdName* **error: could not convert the private key.**

Explanation: The private key material could not be converted successfully.

User response: Contact the IBM Support Center.

6027-3568 *cmdName* **error: could not extract the private key information structure.**

Explanation: The private key could not be extracted successfully.

User response: Contact the IBM Support Center.

6027-3569 *cmdName* **error: could not convert the private key information to DER format.**

Explanation: The private key material could not be converted successfully.

User response: Contact the IBM Support Center.

6027-3570 *cmdName* **error: could not encrypt the private key information structure (error** *returnCode***).**

Explanation: The private key material could not be encrypted successfully.

User response: Contact the IBM Support Center.

6027-3571 *cmdName* **error: could not insert the key in the keystore, check your system's clock (error** *returnCode***).**

Explanation: Insertion of the new keypair into the keystore failed because the local date and time are not properly set on your system.

User response: Synchronize the local date and time on your system and try this command again.

6027-3572 *cmdName* **error: could not insert the key in the keystore (error** *returnCode***).**

Explanation: Insertion of the new keypair into the keystore failed.

User response: Contact the IBM Support Center.

6027-3573 *cmdName* **error: could not insert the certificate in the keystore (error** *returnCode***).**

Explanation: Insertion of the new certificate into the keystore failed.

User response: Contact the IBM Support Center.

6027-3574 *cmdName* **error: could not initialize the digest algorithm.**

Explanation: Initialization of a cryptographic algorithm failed.

User response: Contact the IBM Support Center.

6027-3575 *cmdName* **error: error while computing the digest.**

Explanation: Computation of the certificate digest failed.

User response: Contact the IBM Support Center.

6027-3576 *cmdName* **error: could not initialize the SSL environment (error** *returnCode***).**

Explanation: An internal component of the cryptographic library could not be properly initialized.

User response: Ensure that the `gpfs.gskit` package was installed properly. Contact the IBM Support Center.

6027-3577 **Failed to sync fileset** *filesetName***.** *errString***.**

Explanation: Failed to sync fileset.

User response: Check the error message and try again. If the problem persists, contact the IBM Support Center.

6027-3578 [E] *pathName* **is not a valid argument for this command. You must specify a path name within a single GPFS snapshot.**

Explanation: This message is similar to message number 6027-872, but the *pathName* does not specify a path that can be scanned. The value specified for *pathName* might be a **.snapdir** or similar object.

User response: Correct the command invocation and reissue the command.

6027-3579 *cmdName* **error: the cryptographic library could not be initialized in FIPS mode.**

Explanation: The cluster is configured to operate in FIPS mode but the cryptographic library could not be initialized in that mode.

User response: Verify that the `gpfs.gskit` package has been installed properly and that GPFS supports FIPS mode on your platform. Contact the IBM Support Center.

6027-3580 **Failed to sync file system:** *fileSystem* **Error:** *errString***.**

Explanation: Failed to sync file system.

User response: Check the error message and try again. If the problem persists, contact the IBM Support Center.

6027-3581 **Failed to create the operation list file.**

Explanation: Failed to create the operation list file.

User response: Verify that the file path is correct and check the additional error messages.

6027-3582 [E] Compression is not supported for clone or clone-parent files.

Explanation: File compression is not supported as the file being compressed is a clone or a clone parent file.

User response: None.

6027-3583 [E] Compression is not supported for snapshot files.

Explanation: The file being compressed is within a snapshot and snapshot file compression is not supported.

User response: None.

6027-3584 [E] Current file system version does not support compression.

Explanation: The current file system version is not recent enough for file compression support.

User response: Upgrade the file system to the latest version and retry the command.

6027-3585 [E] Compression is not supported for AFM cached files.

Explanation: The file being compressed is cached in an AFM cache fileset and compression is not supported for such files.

User response: None.

6027-3586 [E] Compression/uncompression failed.

Explanation: Compression or uncompression failed.

User response: Refer to the error message below this line for the cause of the compression failure.

6027-3587 [E] Aborting compression as the file is opened in hyper allocation mode.

Explanation: Compression operation is not performed because the file is opened in hyper allocation mode.

User response: Compress this file after the file is closed.

6027-3588 [E] Aborting compression as the file is currently memory mapped, opened in direct I/O mode, or stored in a horizontal storage pool.

Explanation: Compression operation is not performed because it is inefficient or unsafe to compress the file at this time.

User response: Compress this file after the file is no longer memory mapped, opened in direct I/O mode, or stored in a horizontal storage pool.

6027-3589 *cmdName* **error: Cannot set the password twice.**

Explanation: An attempt was made to set the password by using different available options.

User response: Set the password either through the CLI or by specifying a file that contains it.

6027-3590 *cmdName* **error: Could not access file** *fileName* **(error** *errorCode***).**

Explanation: The specified file could not be accessed.

User response: Check whether the file name is correct and verify whether you have required access privileges to access the file.

6027-3591 *cmdName* **error: The password specified in file** *fileName* **exceeds the maximum length of** *length* **characters.**

Explanation: The password stored in the specified file is too long.

User response: Pick a shorter password and retry the operation.

6027-3592 *cmdName* **error: Could not read the password from file** *fileName***.**

Explanation: The password could not be read from the specified file.

User response: Ensure that the file can be read.

6027-3593 [E] Compression is supported only for regular files.

Explanation: The file is not compressed because compression is supported only for regular files.

User response: None.

6027-3594 [E] [E] Failed to synchronize the being restored fileset:*filesetName***. [I] Please stop the activities in the fileset and rerun the command.**

Explanation: Failed to synchronize the being restored fileset due to some conflicted activities in the fileset.

User response: Stop the activities in the fileset and try the command again. If the problem persists, contact the IBM Support Center.

6027-3595 [E] [E] Failed to synchronize the being restored file system:*fileSystem***. [I] Please stop the activities in the file system and rerun the command.**

Explanation: Failed to synchronize the being restored file system due to some conflicted activities in the file system.

User response: Stop the activities in the file system and try the command again. If the problem persists, contact the IBM Support Center.

6027-3596 *cmdName* **error: could not read/write file from/to directory '***pathName***'. This path does not exist.**

Explanation: A file could not be read from/written to the specified directory.

User response: Ensure that the path exists.

6027-3597 *cmdName* **error: Could not open directory '***pathName***' (error** *errorCode***).**

Explanation: The specified directory could not be opened.

User response: Ensure that the path exists and it is readable.

6027-3598 *cmdName* **error: Could not insert the key in the keystore. Another key with the specified label already exists (error** *errorCode***).**

Explanation: The key could not be inserted into the keystore with the specified label, since another key or certificate already exists with that label.

User response: Use another label for the key.

6027-3599 *cmdName* **error: A certificate with the label '***certLabel***' already exists (error** *errorCode***).**

Explanation: The certificate could not be inserted into the keystore with the specified label, since another key or certificate already exists with that label.

User response: Use another label for the certificate.

6027-3600 *cmdName* **error: The certificate '***certFilename***' already exists in the keystore under another label (error** *errorCode***).**

Explanation: The certificate could not be inserted into the keystore, since it is already stored in the keystore.

User response: None, as the certificate already exists in the keystore under another label. The command does not need to be rerun.

6027-3601 [E] Current file system version does not support compression library selection.

Explanation: The current file system version does not support file compression library selection.

User response: Upgrade the file system to the latest version and retry the command.

6027-3602 [E] Current file system version does not support the specified compression library. Supported libraries include \"z\" and \"lz4\".

Explanation: The current file system version does not support the compression library that is specified by the user.

User response: Select "z" or "lz4" as the compression library.

6027-3602 [E] Snapshot data migration to external pool, *externalPoolName***, is not supported.**

Explanation: Snapshot data can only be migrated among internal pools.

User response: None.

6027-3604 [E] Fail to load compression library.

Explanation: Compression library failed to load.

User response: Make sure that the *gpfs.compression* package is installed.

6027-3700 [E] Key '*keyID***' was not found on RKM ID '***rkmID***'.**

Explanation: The specified key could not be retrieved from the key server.

User response: Verify that the key is present at the server. Verify that the name of the keys used in the encryption policy is correct.

6027-3701 [E] Key '*keyID***:***rkmID***' could not be fetched. The authentication with the RKM was not successful.**

Explanation: Unable to authenticate with the key server.

User response: Verify that the credentials used to authenticate with the key server are correct.

6027-3702 [E] Key '*keyID***:***rkmID***' could not be fetched. Permission denied.**

Explanation: Unable to authenticate with the key server.

User response: Verify that the credentials used to authenticate with the key server are correct.

6027-3703 [E] I/O error while accessing the keystore file '*keystoreFileName***'.**

Explanation: An error occurred while accessing the keystore file.

User response: Verify that the name of the keystore file in `RKM.conf` is correct. Verify that the keystore file can be read on each node.

6027-3704 [E] The keystore file '*keystoreFileName***' has an invalid format.**

Explanation: The specified keystore file has an invalid format.

User response: Verify that the format of the keystore file is correct.

6027-3705 [E] Incorrect FEK length after unwrapping; the operation was cancelled.

Explanation: When opening an existing encrypted file, the size of the FEK that was unwrapped did not correspond to the one recorded in the file's extended attributes. Either the cryptographic extended attributes

were damaged, or the master key(s) used to unwrap the FEK have changed.

User response: Check for other symptoms of data corruption, and verify that the configuration of the key server has not changed.

6027-3706 [E] The crypto library with FIPS support is not available for this architecture. Disable FIPS mode and reattempt the operation.

Explanation: GPFS is operating in FIPS mode, but the initialization of the cryptographic library failed because FIPS mode is not yet supported on this architecture.

User response: Disable FIPS mode and attempt the operation again.

6027-3707 [E] The crypto library could not be initialized in FIPS mode. Ensure that the crypto library package was correctly installed.

Explanation: GPFS is operating in FIPS mode, but the initialization of the cryptographic library failed.

User response: Ensure that the packages required for encryption are properly installed on each node in the cluster.

6027-3708 [E] Incorrect passphrase for backend '*backend*'.

Explanation: The specified passphrase is incorrect for the backend.

User response: Ensure that the correct passphrase is used for the backend in `RKM.conf`.

6027-3709 [E] Error encountered when parsing line *lineNumber*: expected a new RKM backend stanza.

Explanation: An error was encountered when parsing a line in `RKM.conf`. Parsing of the previous backend is complete, and the stanza for the next backend is expected.

User response: Correct the syntax in `RKM.conf`.

6027-3710 [E] Error encountered when parsing line *lineNumber*: invalid key '*keyIdentifier*'.

Explanation: An error was encountered when parsing a line in `RKM.conf`.

User response: Specify a well-formed stanza in `RKM.conf`.

6027-3711 [E] Error encountered when parsing line *lineNumber*: invalid key-value pair.

Explanation: An error was encountered when parsing a line in `RKM.conf`: an invalid key-value pair was found.

User response: Correct the specification of the RKM backend in `RKM.conf`.

6027-3712 [E] Error encountered when parsing line *lineNumber*: incomplete RKM backend stanza '*backend*'.

Explanation: An error was encountered when parsing a line in `RKM.conf`. The specification of the backend stanza was incomplete.

User response: Correct the specification of the RKM backend in `RKM.conf`.

6027-3713 [E] An error was encountered when parsing line *lineNumber*: duplicate key '*key*'.

Explanation: A duplicate keyword was found in `RKM.conf`.

User response: Eliminate duplicate entries in the backend specification.

6027-3714 [E] Incorrect permissions for the /var/mmfs/etc/RKM.conf configuration file on node *nodeName*: the file must be owned by the root user and be in the root group, must be a regular file and be readable and writable by the owner only.

Explanation: The permissions for the `/var/mmfs/etc/RKM.conf` configuration file are incorrect. The file must be owned by the root user, must be in the root group, must be a regular file, and must be readable and writeable by the owner only.

User response: Fix the permissions on the file and retry the operation.

6027-3715 [E] Error encountered when parsing line *lineNumber*: RKM ID '*RKMID*' is too long, it cannot exceed *length* characters.

Explanation: The RKMID chosen at the specified line of `/var/mmfs/etc/RKM.conf` contains too many characters.

User response: Choose a shorter string for the *RKMID*.

6027-3716 [E] Key '*keyID*:*rkmID*' could not be fetched. The TLS handshake could not be completed successfully.

Explanation: The specified key could not be retrieved

from the server because the TLS handshake did not complete successfully.

User response: Ensure that the configurations of GPFS and the remote key management (RKM) server are compatible when it comes to the version of the TLS protocol used upon key retrieval (GPFS uses the **nistCompliance** configuration variable to control that). In particular, if **nistCompliance=SP800-131A** is set in GPFS, ensure that the TLS v1.2 protocol is enabled in the RKM server. If this does not resolve the issue, contact the IBM Support Center.

6027-3717 [E] Key '*keyID*:*rkmID*' **could not be fetched. The RKM is in quarantine after experiencing a fatal error.**

Explanation: GPFS has quarantined the remote key management (RKM) server and will refrain from initiating further connections to it for a limited amount of time.

User response: Examine the error messages that precede this message to determine the cause of the quarantine.

6027-3718 [E] Key '*keyID*:*rkmID*' **could not be fetched. Invalid request.**

Explanation: The key could not be fetched because the remote key management (RKM) server reported that the request was invalid.

User response: Ensure that the RKM server trusts the client certificate that was used for this request. If this does not resolve the issue, contact the IBM Support Center.

6027-3719 [W] Wrapping parameter string '*oldWrappingParameter*' **is not safe and will be replaced with '***newWrappingParameter*'**.**

Explanation: The wrapping parameter specified by the policy should no longer be used since it may cause data corruption or weaken the security of the system. For this reason, the wrapping parameter specified in the message will be used instead.

User response: Change the policy file and replace the specified wrapping parameter with a more secure one. Consult the *IBM Spectrum Scale: Administration Guide* for a list of supported wrapping parameters.

6027-3720 [E] *binaryName* **error: Invalid command type '***command*'**.**

Explanation: The command supplied to the specified binary is invalid.

User response: Specify a valid command. Refer to the documentation for a list of supported commands.

6027-3721 [E] *binaryName* **error: Invalid arguments.**

Explanation: The arguments supplied to the specified binary are invalid.

User response: Supply valid arguments. Refer to the documentation for a list of valid arguments.

6027-3722 [E] An error was encountered while processing file '*fileName*'**:** *errorString*

Explanation: An error was encountered while processing the specified configuration file.

User response: Examine the error message and correct the corresponding conditions.

6027-3723 [E] Incorrect permissions for the configuration file *fileName* **on node** *nodeName***.**

Explanation: The permissions for the specified configuration file are incorrect. The file must be owned by the root user, must be in the root group, must be a regular file, and must be readable and writeable by the owner only.

User response: Fix the permissions on the file and retry the operation.

6027-3726 [E] Key '*keyID*:*rkmID*' **could not be fetched. Bad certificate.**

Explanation: The key could not be fetched from the remote key management (RKM) server because of a problem with the validation of the certificate.

User response: Verify the steps used to generate the server and client certificates. Check whether the NIST settings are correct on the server. If this does not resolve the issue, contact the IBM Support Center.

6027-3727 [E] Key '*keyID*:*rkmID*' **could not be fetched. Invalid tenantName.**

Explanation: The key could not be fetched from the remote key management (RKM) server because the tenantName specified in the RKM.conf file stanza was invalid.

User response: Verify that the tenantName specified in the RKM.conf file stanza is valid, and corresponds to an existing Device Group in the RKM server.

6027-3728 [E] The keyStore permissions are incorrect for *fileName***. Access should be only granted to root, and no execute permission is allowed for the file.**

Explanation: The specified file allows access from a non-root user, or has execute permission, which is not allowed.

User response: Ensure the specified file is not granted access to non root. Explanation: The specified file allows access from a non-root user, or has execute permission, which is not allowed.

6027-3729 [E] Key '*keyID*:*rkmID*' could not be fetched. The SSL connection cannot be initialized.

Explanation: The specified key could not be retrieved from the server, because the SSL connection with the server cannot be initialized. Key server daemon may be unresponsive.

User response: Check connectivity to the server. Check credentials to access the server. Perform problem determination on key server daemon. Contact the IBM Support Center.

6027-3730 [E] Certificate with label '*clientCertLabel*' for backend '*backend*' has expired.

Explanation: The certificate identified by the specified label for the backend has expired.

User response: Create new client credentials.

6027-3900 Invalid flag '*flagName*' in the criteria file.

Explanation: An invalid flag was found in the criteria file.

User response: None.

6027-3901 Failed to receive inode list: *listName*.

Explanation: A failure occurred while receiving an inode list.

User response: None.

6027-3902 Check file '*fileName*' on *fileSystem* for inodes that were found matching the criteria.

Explanation: The named file contains the inodes generated by parallel inode traversal (PIT) with interesting flags; for example, **dataUpdateMiss** or **BROKEN**.

User response: None.

6027-3903 [W] *quotaType* quota is disabled or quota file is invalid.

Explanation: The corresponding quota type is disabled or invalid, and cannot be copied.

User response: Verify that the corresponding quota type is enabled.

6027-3904 [W] *quotaType* quota file is not a metadata file. File was not copied.

Explanation: The quota file is not a metadata file, and it cannot be copied in this way.

User response: Copy quota files directly.

6027-3905 [E] Specified directory does not exist or is invalid.

Explanation: The specified directory does not exist or is invalid.

User response: Check the spelling or validity of the directory.

6027-3906 [W] *backupQuotaFile* already exists.

Explanation: The destination file for a metadata quota file backup already exists.

User response: Move or delete the specified file and retry.

6027-3907 [E] No other quorum node found during cluster probe.

Explanation: The node could not renew its disk lease and there was no other quorum node available to contact.

User response: Determine whether there was a network outage, and also ensure the cluster is configured with enough quorum nodes. The node will attempt to rejoin the cluster.

6027-3908 Check file '*fileName*' on *fileSystem* for inodes with broken disk addresses or failures.

Explanation: The named file contains the inodes generated by parallel inode traversal (PIT) with interesting flags; for example, **dataUpdateMiss** or **BROKEN**.

User response: None.

6027-3909 The file (*backupQuotaFile*) is a quota file in *fileSystem* already.

Explanation: The file is a quota file already. An incorrect file name might have been specified.

User response: None.

6027-3910 [I] Delay *number* seconds for safe recovery.

Explanation: When disk lease is in use, wait for the existing lease to expire before performing log and token manager recovery.

User response: None.

6027-3911 Error reading message from the file system daemon: *errorString* **: The system ran out of memory buffers or memory to expand the memory buffer pool.**

Explanation: The system ran out of memory buffers or memory to expand the memory buffer pool. This prevented the client from receiving a message from the file system daemon.

User response: Try again later.

6027-3912 [E] File *fileName* **cannot run with error** *errorCode***:** *errorString***.**

Explanation: The named shell script cannot run.

User response: Verify that the file exists and that the access permissions are correct.

6027-3913 Attention: disk *diskName* **is a 4K native dataOnly disk and it is used in a non-4K aligned file system. Its usage is not allowed to change from dataOnly.**

Explanation: An attempt was made through the mmchdisk command to change the usage of a 4K native disk in a non-4K aligned file system from dataOnly to something else.

User response: None.

6027-3914 [E] Current file system version does not support compression.

Explanation: File system version is not recent enough for file compression support.

User response: Upgrade the file system to the latest version, then retry the command.

6027-3915 Invalid file system name provided: '*FileSystemName***'.**

Explanation: The specified file system name contains invalid characters.

User response: Specify an existing file system name or one which only contains valid characters.

6027-3916 [E] *fileSystemName* **is a clone of** *fileSystemName***, which is mounted already.**

Explanation: A cloned file system is already mounted internally or externally with the same stripe group ID. The mount will be rejected.

User response: Unmount the cloned file system and remount.

6027-3917 [E] The file *fileName* **does not exist in the root directory of** *fileSystemName***.**

Explanation: The backup file for quota does not exist in the root directory.

User response: Check the file name and root directory and rerun the command after correcting the error.

6027-3918 [N] Disk lease period expired *number* **seconds ago in cluster** *clusterName***. Attempting to reacquire the lease.**

Explanation: The disk lease period expired, which will prevent the local node from being able to perform disk I/O. May be caused by a temporary communication outage.

User response: If message is repeated then investigate the communication outage.

6027-3919 [E] No attribute found.

Explanation: The attribute does not exist.

User response: None.

6027-3920 [E] Cannot find an available quorum node that would be able to successfully run `Expel` command.

Explanation: Expel command needs to be run on quorum node but cannot find any available quorum node that would be able to successfully run the **`Expel`** command. All quorum nodes are either down or being expelled.

User response: None.

6027-3921 Partition `*partitionName***' is created and policy broadcasted to all nodes.**

Explanation: A partition is created and policy broadcasted to all nodes.

User response: None.

6027-3922 Partition `*partitionName***' is deleted and policy broadcasted to all nodes.**

Explanation: A partition is deleted and policy broadcasted to all nodes.

User response: None.

6027-3923 Partition `*partitionName***' does not exist for the file system** *fileSystemName***.**

Explanation: Given policy partition does not exist for the file system.

User response: Verify the partition name and try again.

6027-3924 **Null partition:** *partitionName*

Explanation: Null partition.

User response: None.

6027-3925 **No partitions defined.**

Explanation: No partitions defined.

User response: None.

6027-3926 **Empty policy file.**

Explanation: Empty policy file.

User response: None.

6027-3927 **Failed to read policy partition '***partitionName***' for file system '***fileSystemName***'.**

Explanation: Could not read the given policy partition for the file system.

User response: Reissue the command. If the problem persists, re-install the subject policy partition.

6027-3928 **Failed to list policy partitions for file system '***fileSystemName***'.**

Explanation: Could not list the policy partitions for the file system.

User response: Reissue the command. If the problem persists, contact IBM Support.

6027-3929 **Policy file for file system** *fileSystemName* **contains** *numOfPartitions* **partitions.**

Explanation: Number of partitions in the policy file of the file system.

User response: None.

6027-3930 **No policy partitions defined for file system** *fileSystemName***.**

Explanation: Policy partitions are not defined for the file system.

User response: None.

6027-3931 **Policy partitions are not enabled for file system '***fileSystemName***'.**

Explanation: The cited file system must be upgraded to use policy partitions.

User response: Upgrade the file system by using the **mmchfs -V** command.

6027-3932 [E] GPFS daemon has an incompatible version.

Explanation: The GPFS daemon has an incompatible version.

User response: None.

6027-4000 [I] *descriptorType* **descriptor on this NSD can be updated by running the following command from the node physically connected to NSD** *nsdName***:**

Explanation: This message is displayed when a descriptor validation thread finds a valid NSD, or disk, or stripe group descriptor but with a different ID. This can happen if a device is reused for another NSD.

User response: None. After this message, another message is displayed with a command to fix the problem.

6027-4001 [I] 'mmfsadm writeDesc <device> *descriptorType descriptorId***:***descriptorId nsdFormatVersion pdiskStatus***', where device is the device name of that NSD.**

Explanation: This message displays the command that must run to fix the NSD or disk descriptor on that device. The *deviceName* must be supplied by system administrator or obtained from **mmlsnsd -m** command. The *descriptorId* is a hexadecimal value.

User response: Run the command that is displayed on that NSD server node and replace *deviceName* with the device name of that NSD.

6027-4002 [I] Before running this command, check both NSDs. You might have to delete one of the NSDs.

Explanation: Informational message.

User response: The system administrator should decide which NSD to keep before running the command to fix it. If you want to keep the NSD found on disk, then you do not run the command. Instead, delete the other NSD found in cache (the NSD ID shown in the command).

6027-4003 [E] The on-disk *descriptorType* **descriptor of** *nsdName descriptorIdName descriptorId***:***descriptorId* **is not valid because of bad** *corruptionType***:**

Explanation: The descriptor validation thread found an on-disk descriptor that is corrupted. GPFS will automatically fix it.

User response: None.

6027-4004 [D] **On-disk NSD descriptor: nsdId** *nsdId* **nsdMagic** *nsdMagic* **nsdFormatVersion** *nsdFormatVersion* **on disk nsdChecksum** *nsdChecksum* **calculated checksum** *calculatedChecksum* **nsdDescSize** *nsdDescSize* **firstPaxosSector** *firstPaxosSector* **nPaxosSectors** *nPaxosSectors* **nsdIsPdisk** *nsdIsPdisk*

Explanation: Description of an on-disk NSD descriptor.

User response: None.

6027-4005 [D] **Local copy of NSD descriptor: nsdId** *nsdId* **nsdMagic** *nsdMagic* **formatVersion** *formatVersion* **nsdDescSize** *nsdDescSize* **firstPaxosSector** *firstPaxosSector* **nPaxosSectors** *nPaxosSectors*

Explanation: Description of the cached NSD descriptor.

User response: None.

6027-4006 [I] **Writing NSD descriptor of** *nsdName* **with local copy: nsdId** *nsdId* **nsdFormatVersion** *formatVersion* **firstPaxosSector** *firstPaxosSector* **nPaxosSectors** *nPaxosSectors* **nsdDescSize** *nsdDescSize* **nsdIsPdisk** *nsdIsPdisk* **nsdChecksum** *nsdChecksum*

Explanation: Description of the NSD descriptor that was written.

User response: None.

6027-4007 *errorType* **descriptor on** *descriptorType* **nsdId** *nsdId***:***nsdId* **error** *error*

Explanation: This message is displayed after reading and writing NSD, disk and stripe group descriptors.

User response: None.

6027-4008 [E] **On-disk** *descriptorType* **descriptor of** *nsdName* **is valid but has a different UID: uid** *descriptorId***:***descriptorId* **on-disk uid** *descriptorId***:***descriptorId* **nsdId** *nsdId***:***nsdId*

Explanation: While verifying an on-disk descriptor, a valid descriptor was found but with a different ID. This can happen if a device is reused for another NSD with the **mmcrnsd -v no** command.

User response: After this message there are more messages displayed that describe the actions to follow.

6027-4009 [E] **On-disk NSD descriptor of** *nsdName* **is valid but has a different ID. ID in cache is** *cachedId* **and ID on-disk is** *ondiskId*

Explanation: While verifying an on-disk NSD descriptor, a valid descriptor was found but with a different ID. This can happen if a device is reused for another NSD with the **mmcrnsd -v no** command.

User response: After this message, there are more messages displayed that describe the actions to follow.

6027-4010 [I] **This corruption can happen if the device is reused by another NSD with the -v option and a file system is created with that reused NSD.**

Explanation: Description of a corruption that can happen when an NSD is reused.

User response: Verify that the NSD was not reused to create another NSD with the **-v** option and that the NSD was not used for another file system.

6027-4011 [D] **On-disk disk descriptor: uid** *descriptorID***:***descriptorID* **magic** *descMagic* **formatVersion** *formatVersion* **descSize** *descSize* **checksum on disk** *diskChecksum* **calculated checksum** *calculatedChecksum* **firstSGDescSector** *firstSGDescSector* **nSGDescSectors** *nSGDescSectors* **lastUpdateTime** *lastUpdateTime*

Explanation: Description of the on-disk disk descriptor.

User response: None.

6027-4012 [D] **Local copy of disk descriptor: uid** *descriptorID***:***descriptorID* **firstSGDescSector** *firstSGDescSector* **nSGDescSectors** *nSGDescSectors*

Explanation: Description of the cached disk descriptor.

User response: None.

6027-4013 [I] **Writing disk descriptor of** *nsdName* **with local copy: uid** *descriptorID***:***descriptorID***, magic** *magic***, formatVersion** *formatVersion* **firstSGDescSector** *firstSGDescSector* **nSGDescSectors** *nSGDescSectors* **descSize** *descSize*

Explanation: Writing disk descriptor to disk with local information.

User response: None.

6027-4014 [D] Local copy of StripeGroup descriptor: uid *descriptorID*:*descriptorID* **curFmtVersion** *curFmtVersion* **configVersion** *configVersion*

Explanation: Description of the cached stripe group descriptor.

User response: None.

6027-4015 [D] On-disk StripeGroup descriptor: uid *sgUid*:*sgUid* **magic** *magic* **curFmtVersion** *curFmtVersion* **descSize** *descSize* **on-disk checksum** *diskChecksum* **calculated checksum** *calculatedChecksum* **configVersion** *configVersion* **lastUpdateTime** *lastUpdateTime*

Explanation: Description of the on-disk stripe group descriptor.

User response: None.

6027-4016 [E] Data buffer checksum mismatch during write. File system *fileSystem* **tag** *tag1 tag2* **nBytes** *nBytes diskAddresses*

Explanation: GPFS detected a mismatch in the checksum of the data buffer content which means content of data buffer was changing while a direct I/O write operation was in progress.

User response: None.

6027-4017 [E] Current file system version does not support the initial disk status *BeingAddedByGNR*.

Explanation: File system version must be upgraded to specify *BeingAddedByGNR* as the initial disk status.

User response: Upgrade the file system version.

6027-4018 [E] Disk *diskName* **is not an existing vdisk, but initial status** *BeingAddedByGNR* **is specified**

Explanation: When you specify the initial disk status *BeingAddedByGNR*, all disks that are being added must be existing NSDs of type vdisk

User response: Ensure that NSDs are of type vdisk and try again.

6027-4019 [D] On-disk StripeGroup descriptor: uid 0x*sgUid*:*sgUid* **magic 0x***magic* **curFmtVersion** *curFmtVersion* **on-disk descSize** *descSize* **cached descSize** *descSize* **on-disk checksum 0x***diskChecksum* **calculated checksum 0x***calculatedChecksum* **configVersion** *configVersion* **lastUpdateTime** *lastUpdateTime*

Explanation: Description of the on-disk stripe group descriptor.

User response: None.

Accessibility features for IBM Spectrum Scale

Accessibility features help users who have a disability, such as restricted mobility or limited vision, to use information technology products successfully.

Accessibility features

The following list includes the major accessibility features in IBM Spectrum Scale:

- Keyboard-only operation
- Interfaces that are commonly used by screen readers
- Keys that are discernible by touch but do not activate just by touching them
- Industry-standard devices for ports and connectors
- The attachment of alternative input and output devices

IBM Knowledge Center, and its related publications, are accessibility-enabled. The accessibility features are described in IBM Knowledge Center (www.ibm.com/support/knowledgecenter).

Keyboard navigation

This product uses standard Microsoft Windows navigation keys.

IBM and accessibility

See the IBM Human Ability and Accessibility Center (www.ibm.com/able) for more information about the commitment that IBM has to accessibility.

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing IBM Corporation North Castle Drive, MD-NC119 Armonk, NY 10504-1785 US

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing Legal and Intellectual Property Law IBM Japan Ltd. 19-21, Nihonbashi-Hakozakicho, Chuo-ku Tokyo 103-8510, Japan

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Director of Licensing IBM Corporation North Castle Drive, MD-NC119 Armonk, NY 10504-1785 US

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

The performance data discussed herein is presented as derived under specific operating conditions. Actual results may vary.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Each copy or any portion of these sample programs or any derivative work must include a copyright notice as follows:

© (your company name) (year).
Portions of this code are derived from IBM Corp.
Sample Programs. © Copyright IBM Corp. _enter the year or years_.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at Copyright and trademark information at www.ibm.com/legal/copytrade.shtml.

Intel is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of the Open Group in the United States and other countries.

Terms and conditions for product documentation

Permissions for the use of these publications are granted subject to the following terms and conditions.

Applicability

These terms and conditions are in addition to any terms of use for the IBM website.

Personal use

You may reproduce these publications for your personal, noncommercial use provided that all proprietary notices are preserved. You may not distribute, display or make derivative work of these publications, or any portion thereof, without the express consent of IBM.

Commercial use

You may reproduce, distribute and display these publications solely within your enterprise provided that all proprietary notices are preserved. You may not make derivative works of these publications, or reproduce, distribute or display these publications or any portion thereof outside your enterprise, without the express consent of IBM.

Rights

Except as expressly granted in this permission, no other permissions, licenses or rights are granted, either express or implied, to the publications or any information, data, software or other intellectual property contained therein.

IBM reserves the right to withdraw the permissions granted herein whenever, in its discretion, the use of the publications is detrimental to its interest or, as determined by IBM, the above instructions are not being properly followed.

You may not download, export or re-export this information except in full compliance with all applicable laws and regulations, including all United States export laws and regulations.

IBM MAKES NO GUARANTEE ABOUT THE CONTENT OF THESE PUBLICATIONS. THE PUBLICATIONS ARE PROVIDED "AS-IS" AND WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, AND FITNESS FOR A PARTICULAR PURPOSE.

IBM Online Privacy Statement

IBM Software products, including software as a service solutions, (“Software Offerings”) may use cookies or other technologies to collect product usage information, to help improve the end user experience, to tailor interactions with the end user or for other purposes. In many cases no personally identifiable information is collected by the Software Offerings. Some of our Software Offerings can help enable you to

collect personally identifiable information. If this Software Offering uses cookies to collect personally identifiable information, specific information about this offering's use of cookies is set forth below.

This Software Offering does not use cookies or other technologies to collect personally identifiable information.

If the configurations deployed for this Software Offering provide you as customer the ability to collect personally identifiable information from end users via cookies and other technologies, you should seek your own legal advice about any laws applicable to such data collection, including any requirements for notice and consent.

For more information about the use of various technologies, including cookies, for these purposes, See IBM's Privacy Policy at http://www.ibm.com/privacy and IBM's Online Privacy Statement at http://www.ibm.com/privacy/details the section entitled "Cookies, Web Beacons and Other Technologies" and the "IBM Software Products and Software-as-a-Service Privacy Statement" at http://www.ibm.com/software/info/product-privacy.

Glossary

This glossary provides terms and definitions for IBM Spectrum Scale.

The following cross-references are used in this glossary:

- *See* refers you from a nonpreferred term to the preferred term or from an abbreviation to the spelled-out form.
- *See also* refers you to a related or contrasting term.

For other terms and definitions, see the IBM Terminology website (www.ibm.com/software/globalization/terminology) (opens in new window).

B

block utilization
The measurement of the percentage of used subblocks per allocated blocks.

C

cluster
A loosely-coupled collection of independent systems (nodes) organized into a network for the purpose of sharing resources and communicating with each other. See also *GPFS cluster*.

cluster configuration data
The configuration data that is stored on the cluster configuration servers.

Cluster Export Services (CES) nodes
A subset of nodes configured within a cluster to provide a solution for exporting GPFS file systems by using the Network File System (NFS), Server Message Block (SMB), and Object protocols.

cluster manager
The node that monitors node status using disk leases, detects failures, drives recovery, and selects file system managers. The cluster manager must be a quorum node. The selection of the cluster manager node favors the quorum-manager node with the lowest node number among the nodes that are operating at that particular time.

Note: The cluster manager role is not moved to another node when a node with a lower node number becomes active.

control data structures
Data structures needed to manage file data and metadata cached in memory. Control data structures include hash tables and link pointers for finding cached data; lock states and tokens to implement distributed locking; and various flags and sequence numbers to keep track of updates to the cached data.

D

Data Management Application Program Interface (DMAPI)
The interface defined by the Open Group's XDSM standard as described in the publication *System Management: Data Storage Management (XDSM) API Common Application Environment (CAE) Specification C429*, The Open Group ISBN 1-85912-190-X.

deadman switch timer
A kernel timer that works on a node that has lost its disk lease and has outstanding I/O requests. This timer ensures that the node cannot complete the outstanding I/O requests (which would risk causing file system corruption), by causing a panic in the kernel.

dependent fileset
A fileset that shares the inode space of an existing independent fileset.

disk descriptor
A definition of the type of data that the disk contains and the failure group to which this disk belongs. See also *failure group*.

disk leasing
A method for controlling access to storage devices from multiple host systems. Any host that wants to access a storage device configured to use disk leasing registers for a lease; in the event of a perceived failure, a host system can deny access,

preventing I/O operations with the storage device until the preempted system has reregistered.

disposition
The session to which a data management event is delivered. An individual disposition is set for each type of event from each file system.

domain
A logical grouping of resources in a network for the purpose of common management and administration.

E

ECKD™
See *extended count key data (ECKD)*.

ECKD device
See *extended count key data device (ECKD device)*.

encryption key
A mathematical value that allows components to verify that they are in communication with the expected server. Encryption keys are based on a public or private key pair that is created during the installation process. See also *file encryption key*, *master encryption key*.

extended count key data (ECKD)
An extension of the count-key-data (CKD) architecture. It includes additional commands that can be used to improve performance.

extended count key data device (ECKD device)
A disk storage device that has a data transfer rate faster than some processors can utilize and that is connected to the processor through use of a speed matching buffer. A specialized channel program is needed to communicate with such a device. See also *fixed-block architecture disk device*.

F

failback
Cluster recovery from failover following repair. See also *failover*.

failover
(1) The assumption of file system duties by another node when a node fails. (2) The process of transferring all control of the ESS to a single cluster in the ESS when the other clusters in the ESS fails. See also *cluster*. (3) The routing of all transactions to a second controller when the first controller fails. See also *cluster*.

failure group
A collection of disks that share common access paths or adapter connection, and could all become unavailable through a single hardware failure.

FEK See *file encryption key*.

fileset A hierarchical grouping of files managed as a unit for balancing workload across a cluster. See also *dependent fileset*, *independent fileset*.

fileset snapshot
A snapshot of an independent fileset plus all dependent filesets.

file clone
A writable snapshot of an individual file.

file encryption key (FEK)
A key used to encrypt sectors of an individual file. See also *encryption key*.

file-management policy
A set of rules defined in a policy file that GPFS uses to manage file migration and file deletion. See also *policy*.

file-placement policy
A set of rules defined in a policy file that GPFS uses to manage the initial placement of a newly created file. See also *policy*.

file system descriptor
A data structure containing key information about a file system. This information includes the disks assigned to the file system (*stripe group*), the current state of the file system, and pointers to key files such as quota files and log files.

file system descriptor quorum
The number of disks needed in order to write the file system descriptor correctly.

file system manager
The provider of services for all the nodes using a single file system. A file system manager processes changes to the state or description of the file system, controls the regions of disks that are allocated to each node, and controls token management and quota management.

fixed-block architecture disk device (FBA disk device)
A disk device that stores data in blocks of fixed size. These blocks are addressed by block number relative to the beginning of the file. See also *extended count key data device*.

fragment
The space allocated for an amount of data too small to require a full block. A fragment consists of one or more subblocks.

G

global snapshot
A snapshot of an entire GPFS file system.

GPFS cluster
A cluster of nodes defined as being available for use by GPFS file systems.

GPFS portability layer
The interface module that each installation must build for its specific hardware platform and Linux distribution.

GPFS recovery log
A file that contains a record of metadata activity, and exists for each node of a cluster. In the event of a node failure, the recovery log for the failed node is replayed, restoring the file system to a consistent state and allowing other nodes to continue working.

I

ill-placed file
A file assigned to one storage pool, but having some or all of its data in a different storage pool.

ill-replicated file
A file with contents that are not correctly replicated according to the desired setting for that file. This situation occurs in the interval between a change in the file's replication settings or suspending one of its disks, and the restripe of the file.

independent fileset
A fileset that has its own inode space.

indirect block
A block containing pointers to other blocks.

inode The internal structure that describes the individual files in the file system. There is one inode for each file.

inode space
A collection of inode number ranges reserved for an independent fileset, which enables more efficient per-fileset functions.

ISKLM
IBM Security Key Lifecycle Manager. For GPFS encryption, the ISKLM is used as an RKM server to store MEKs.

J

journaled file system (JFS)
A technology designed for high-throughput server environments, which are important for running intranet and other high-performance e-business file servers.

junction
A special directory entry that connects a name in a directory of one fileset to the root directory of another fileset.

K

kernel The part of an operating system that contains programs for such tasks as input/output, management and control of hardware, and the scheduling of user tasks.

M

master encryption key (MEK)
A key used to encrypt other keys. See also *encryption key*.

MEK See *master encryption key*.

metadata
Data structures that contain information that is needed to access file data. Metadata includes inodes, indirect blocks, and directories. Metadata is not accessible to user applications.

metanode
The one node per open file that is responsible for maintaining file metadata integrity. In most cases, the node that has had the file open for the longest period of continuous time is the metanode.

mirroring
The process of writing the same data to multiple disks at the same time. The

mirroring of data protects it against data loss within the database or within the recovery log.

Microsoft Management Console (MMC)
A Windows tool that can be used to do basic configuration tasks on an SMB server. These tasks include administrative tasks such as listing or closing the connected users and open files, and creating and manipulating SMB shares.

multi-tailed
A disk connected to multiple nodes.

N

namespace
Space reserved by a file system to contain the names of its objects.

Network File System (NFS)
A protocol, developed by Sun Microsystems, Incorporated, that allows any host in a network to gain access to another host or netgroup and their file directories.

Network Shared Disk (NSD)
A component for cluster-wide disk naming and access.

NSD volume ID
A unique 16 digit hex number that is used to identify and access all NSDs.

node An individual operating-system image within a cluster. Depending on the way in which the computer system is partitioned, it may contain one or more nodes.

node descriptor
A definition that indicates how GPFS uses a node. Possible functions include: manager node, client node, quorum node, and nonquorum node.

node number
A number that is generated and maintained by GPFS as the cluster is created, and as nodes are added to or deleted from the cluster.

node quorum
The minimum number of nodes that must be running in order for the daemon to start.

node quorum with tiebreaker disks
A form of quorum that allows GPFS to run with as little as one quorum node available, as long as there is access to a majority of the quorum disks.

non-quorum node
A node in a cluster that is not counted for the purposes of quorum determination.

P

policy A list of file-placement, service-class, and encryption rules that define characteristics and placement of files. Several policies can be defined within the configuration, but only one policy set is active at one time.

policy rule
A programming statement within a policy that defines a specific action to be performed.

pool A group of resources with similar characteristics and attributes.

portability
The ability of a programming language to compile successfully on different operating systems without requiring changes to the source code.

primary GPFS cluster configuration server
In a GPFS cluster, the node chosen to maintain the GPFS cluster configuration data.

private IP address
A IP address used to communicate on a private network.

public IP address
A IP address used to communicate on a public network.

Q

quorum node
A node in the cluster that is counted to determine whether a quorum exists.

quota The amount of disk space and number of inodes assigned as upper limits for a specified user, group of users, or fileset.

quota management
The allocation of disk blocks to the other nodes writing to the file system, and comparison of the allocated space to quota limits at regular intervals.

R

Redundant Array of Independent Disks (RAID)
A collection of two or more disk physical drives that present to the host an image of one or more logical disk drives. In the event of a single physical device failure, the data can be read or regenerated from the other disk drives in the array due to data redundancy.

recovery
The process of restoring access to file system data when a failure has occurred. Recovery can involve reconstructing data or providing alternative routing through a different server.

remote key management server (RKM server)
A server that is used to store master encryption keys.

replication
The process of maintaining a defined set of data in more than one location. Replication involves copying designated changes for one location (a source) to another (a target), and synchronizing the data in both locations.

RKM server
See *remote key management server*.

rule A list of conditions and actions that are triggered when certain conditions are met. Conditions include attributes about an object (file name, type or extension, dates, owner, and groups), the requesting client, and the container name associated with the object.

S

SAN-attached
Disks that are physically attached to all nodes in the cluster using Serial Storage Architecture (SSA) connections or using Fibre Channel switches.

Scale Out Backup and Restore (SOBAR)
A specialized mechanism for data protection against disaster only for GPFS file systems that are managed by IBM Spectrum Protect Hierarchical Storage Management (HSM).

secondary GPFS cluster configuration server
In a GPFS cluster, the node chosen to maintain the GPFS cluster configuration data in the event that the primary GPFS cluster configuration server fails or becomes unavailable.

Secure Hash Algorithm digest (SHA digest)
A character string used to identify a GPFS security key.

session failure
The loss of all resources of a data management session due to the failure of the daemon on the session node.

session node
The node on which a data management session was created.

Small Computer System Interface (SCSI)
An ANSI-standard electronic interface that allows personal computers to communicate with peripheral hardware, such as disk drives, tape drives, CD-ROM drives, printers, and scanners faster and more flexibly than previous interfaces.

snapshot
An exact copy of changed data in the active files and directories of a file system or fileset at a single point in time. See also *fileset snapshot*, *global snapshot*.

source node
The node on which a data management event is generated.

stand-alone client
The node in a one-node cluster.

storage area network (SAN)
A dedicated storage network tailored to a specific environment, combining servers, storage products, networking products, software, and services.

storage pool
A grouping of storage space consisting of volumes, logical unit numbers (LUNs), or addresses that share a common set of administrative characteristics.

stripe group
The set of disks comprising the storage assigned to a file system.

striping
A storage process in which information is split into blocks (a fixed amount of data) and the blocks are written to (or read from) a series of disks in parallel.

subblock
The smallest unit of data accessible in an I/O operation, equal to one thirty-second of a data block.

system storage pool
A storage pool containing file system control structures, reserved files, directories, symbolic links, special devices, as well as the metadata associated with regular files, including indirect blocks and extended attributes The **system storage pool** can also contain user data.

T

token management
A system for controlling file access in which each application performing a read or write operation is granted some form of access to a specific block of file data. Token management provides data consistency and controls conflicts. Token management has two components: the token management server, and the token management function.

token management function
A component of token management that requests tokens from the token management server. The token management function is located on each cluster node.

token management server
A component of token management that controls tokens relating to the operation of the file system. The token management server is located at the file system manager node.

transparent cloud tiering (TCT)
A separately installable add-on feature of IBM Spectrum Scale that provides a native cloud storage tier. It allows data center administrators to free up on-premise storage capacity, by moving out cooler data to the cloud storage, thereby reducing capital and operational expenditures. .

twin-tailed
A disk connected to two nodes.

U

user storage pool
A storage pool containing the blocks of data that make up user files.

V

VFS See *virtual file system*.

virtual file system (VFS)
A remote file system that has been mounted so that it is accessible to the local user.

virtual node (vnode)
The structure that contains information about a file system object in a virtual file system (VFS).

Index

Special characters

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Product Number: 5725-Q01
5641-GPF
5725-S28
5765-ESS

Printed in USA

SC27-9264-03