



IBM
Spectrum
Archive

IBM Spectrum Archive™ Solutions

An introduction to archive solutions with
IBM Spectrum Archive Enterprise Edition leveraging the
Linear Tape File System (LTFS)

Version 6.0 (August 2020)

Table of content:

1.	Introduction	3
2.	Overview	4
2.1	File archiving in scaling environments.....	4
2.1.1	Archiving requirements.....	4
2.1.2	File archiving process	5
2.1.3	Archiving and ILM	6
2.2	File archive solution.....	6
2.2.1	IBM Spectrum Scale	6
2.2.1.1	IBM Spectrum Scale RAID.....	7
2.2.1.2	Data Lifecycle Management.....	8
2.2.2	IBM Spectrum Archive Enterprise Edition	9
2.2.2.1	Characteristics.....	9
2.2.2.2	Migration Policies	10
3.	IBM Spectrum Archive solutions.....	12
3.1	IBM Elastic Storage Server (ESS) configuration.....	13
3.1.1	File systems	13
3.2	Application server configuration	14
3.2.1	Network File System users	14
3.2.2	Archive application.....	15
3.3	Spectrum Archive configuration	16
3.3.1	IBM Spectrum Archive EE server setup	17
3.3.2	Migration policies for NFS users	18
3.3.3	Migration policies for Archive application.....	20
3.3.4	Configuring migration callback.....	21
4.	Summary.....	22
4.1.1	Total cost of ownership.....	22
4.1.2	Scalability	22
4.1.3	Data availability and protection.....	22
4.1.4	WORM protection	23
4.1.5	Future proof	23
5.	Appendix	24
5.1	References.....	24
5.2	Disclaimer and Trademarks	25

Appreciations

Thanks to my valued colleague Achim Christ for the thorough review and the insights into IBM Spectrum Archive EE. Thanks also to Nick Cvetkovic – a former IT consultant and native English speaker – for his proof-reading.

1. Introduction

In the backdrop of exponentially growing data accelerated by new IT trends - such as cloud, analytics, mobile and social – file archiving is becoming more and more important. The ever growing volumes of data along with stringent regulatory and industry compliance requirements are the main drivers for archiving. Just using backups to fulfill the requirements for archiving is not sufficient.

The majority of unstructured data across industries is stored in files managed by file systems. Files are typically created and processed for a while until they idle. Instead of deleting files when they are no longer accessed, files are typically kept for long periods of time, causing file storage systems to constantly grow. And so does the total cost of ownership for storing and managing massive amounts of files.

File archiving helps to optimize the cost by moving files that are no longer or rarely accessed to cost efficient storage media such as tape. Tape is cost efficient, unlike disk, because it does not consume power when data is not accessed. Over long file lifecycles the savings in power costs for large amounts of data stored on tape accumulate significantly, as independent studies have shown [1].

Naturally, the cost savings associated with archiving can only be justified if the archive process is automated and the access to files is transparent to the user. This enables also facilitates the seamless integration of archiving processes into existing file storage environments.

The requirement to scale out file storage environments is becoming more and more important. Scale out means to scale the file storage independently in multiple dimension such as capacity, performance and file access. In addition, a scale-out file storage must address requirements for business continuity, global collaboration and operational efficiency.

In this paper we present an archive solution for files in a highly scalable environment that is based on IBM Spectrum Archive™ Enterprise Edition. This innovative solution allows storing files on the most appropriate storage medium over their lifecycle, helping to optimize total cost of ownership. Files that are no longer accessed are automatically migrated to the Linear Tape File System (LTFS) while providing transparent access to the user in a global file system name space. The LTFS tape is self-describing based on standardized LTFS format [4] and can be seamlessly exchanged with other systems.

2. Overview

In this chapter general [file archiving process and requirements](#) are highlighted before main concepts and components of the file archive solution are discussed. The file archive solution is based on the integration of the [IBM Spectrum Scale™](#) clustered file system with [IBM Spectrum Archive Enterprise Edition](#) (IBM Spectrum Archive EE). IBM Spectrum Scale provides Information Lifecycle Management functions allowing to identify files that need to be archived based on file properties. IBM Spectrum Archive EE transparently migrates identified files to LTFS formatted tapes.

2.1 File archiving in scaling environments

More and more files are stored in file systems and over time many files are never or rarely accessed. Deleting files sounds easy but requires additional decision processes and approvals in a business environment. Thus, most files are kept in the file storage system causing constant data growth.

Storing all data on disk storage devices becomes costly especially for large volumes of data (100s of Terabytes), because disk requires power and cooling to operate [1]. To cope with the data growth, more and more disk drives must be added leading to further cost increases. This demonstrates the need to archive files that are no longer or rarely accessed to less expensive storage media such as tape.

In the following sections the [general requirements for archiving](#) are presented in order to define the scope of the file archive solution. Subsequently the typical [process for file archiving](#) is outlined, followed by a discussion about the role of [Information Lifecycle Management](#) in file archiving.

2.1.1 Archiving requirements

Archiving in a scale-out file storage environment has the following requirements:

- Optimization of total cost of ownership
- Flexibility and scalability in accordance with changing business requirements
- Data availability and protection
- WORM protection of data
- Future proof

The first and perhaps most important requirement for archiving is *optimizing total cost of ownership*. Cost for storing and managing data accumulates over long time periods and includes a variety of cost factors such as operational efforts, power and cooling and purchasing cost. Operational costs are driven by the day-to-day efforts associated with the maintenance and monitoring of the archive system and for data migrations. Over long periods of time operational cost can become tremendous, especially if operational processes are not automated. Power and cooling cost is mainly driven by the storage medium. While disk drives consume power and require cooling constantly, tapes are much more cost efficient. The purchasing cost is typically not the dominant factor and includes the cost of purchasing hardware, software and services. Information Lifecycle Management (ILM) is one technique to drive optimization of total cost of ownership (see section [Archiving and ILM](#))

Archiving in a scale-out file storage environment must *scale* with the file data. As more files are stored more might have to be archived. Archiving processes must

be designed to adapt to changing business requirements quickly. For example the archive process must allow for flexible changes of selection criteria for files to be archived (see [File archiving process](#)).

Because of the long lifecycles associated with archiving the *likelihood of failures and disaster increases*. Archiving of files in a scale-out storage environment must assure access to files even in case of undesired events.

WORM - write once read many - means that data that has been written once cannot be modified or deleted. Legal obligations demand that certain data needs to be WORM protected. Very often this data has to be kept for longer periods of time such as 5 or 7 years or even longer. This data has to be archived and retained WORM protected in order to meet these requirements.

One of the key challenges associated with archiving is the *technological progress* and the associated obsolescence of infrastructure components. Archiving must assure that files can be easily read through standardized interfaces, because these provide long life spans. It must also assure that files and metadata can be easily migrated to new infrastructure. Automation is essential to accelerate the migration.

In section [Summary](#) we demonstrate how the file archive solution presented herein meet the requirements discussed above.

2.1.2 File archiving process

The archive process for files in a scale-out storage environment is comprised of two steps:

1. Selection of files to be archived
2. Transparent migration of files from disk to tape while maintaining access from a user perspective

The selection of files is automated and flexible. Files can be selected based on access behavior, file names, user names, file size and additional file attributes. A rule based policy system used to describe the selection criteria is advantageous and reduces operational cost.

The migration of files is transparent to the user. Thus the user does not notice that the file has been migrated to tape and he has transparent access to the file. This means that the user sees the file in the file system after it has been migrated. When the user accesses the file, it is automatically copied back (recalled) from tape to disk. Because the access behavior is slower with tape, the user may have to wait longer before he can open the file. This is a disadvantage of tape which is offset by the cost savings.

Another advantage of file archiving on tape is that files can be seamlessly accessed on tape, even if the scale-out file storage is not available. Just like the Linear Tape File System (LTFS) which allows access data on tape just like in a file system. This also fosters the interchangeability of data using high capacity tapes.

In the section [IBM Spectrum Archive solutions](#) we demonstrate how this archive process is implemented with this archive solution for files.

2.1.3 Archiving and ILM

File archiving in scale-out storage environments is closely related to the concepts of Information Lifecycle Management (ILM). The goal of ILM is to optimize total cost of ownership by storing data on the most appropriate storage medium over its lifetime. This starts with the creation of data and continues over the lifecycle until deletion. Cost optimization is particularly important for data that is kept for long periods of time because cost is cumulative.

There are two key techniques associated with ILM: placement and migration. File placement assures that a file is initially created on the most appropriate storage medium. Migration assures that files are migrated to the most appropriate storage medium over their life cycle. When files are no longer accessed, they can be migrated to a lower cost storage medium like tape. This frees up space on the more expensive disk storage and reduces storage cost because tapes do not consume power.

ILM processes for placement and migration of files must be fully automated in order to drive further cost savings, especially for operational cost. This also includes transparent access to files from a user perspective, eliminating the need for manual interventions.

In section [Data Lifecycle Management](#) the ILM function implemented in IBM Spectrum Scale are explained.

2.2 File archive solution

The file archiving solution presented in this document combines the clustered, scale-out file system [IBM Spectrum Scale](#) [2] with [IBM Spectrum Archive Enterprise Edition](#) [14]. This enabled you to automatically move infrequently accessed data from disk to tape to lower costs while retaining ease of use and without the need for proprietary tape applications.

Both components – IBM Spectrum Scale and IBM Spectrum Archive EE - are members of the IBM Spectrum Storage™ family providing software define storage.

2.2.1 IBM Spectrum Scale

Big data. Cloud storage. It doesn't matter what you call it, today's exponential growth of data, transactions and digitally aware devices are demanding larger and larger amounts of unstructured data storage.

IBM has taken on this challenge with software defined storage file system: IBM Spectrum Scale™ [2]. As a high-performance enterprise platform for optimizing data and file management, IBM Spectrum Scale is used extensively across industries worldwide.

IBM Spectrum Scale – also known as GPFS – is a file system software providing global namespace. It simplifies data management with integrated tools designed to help organizations manage Petabytes of data and billions of files—as well as to control the cost of managing these ever-growing data volumes. Spectrum Scale includes innovative features, such as:

- Graphical User Interface

- Transparent storage tiering on flash, disk and tape
- Spectrum Scale RAID
- Clustered NFS server exporting providing file services via NFS, SMB and Swift / S3
- File system snapshot and quota support
- Global file sharing with multi-cluster and Active File Management (AFM)
- Multisite high availability and disaster protection through synchronous and asynchronous replication
- Integrated Backup and restore
- Selective file encryption and secure erase
- Selective File compression
- Immutability [18]

To provide clients with an easy to use storage system, the Elastic Storage Server (ESS) has been developed that enables companies to store large quantities of file data in a global name space. ESS incorporates the leading IBM Spectrum Scale file system software, comes in different “T-Shirt sizes” and can be flexibly combined to build large file storage clusters [8].

An ESS system comprises two servers running the IBM Spectrum Scale RAID software (see section [IBM Spectrum Scale RAID](#)) and many SAS attached disk drives where the RAID protected data is stored. The two Spectrum Scale servers provide the Spectrum Scale file systems that can be accessed via Ethernet or InfiniBand networks. An ESS system is typically used as file system storage from application servers running all kind of applications including backup server, database server, analytics software, genomic software, file and object storage and many more. These applications can benefit from the data lifecycle management function of IBM Spectrum Scale.

In the next section two key IBM Spectrum Scale features that are relevant for this archive solution are explained: [IBM Spectrum Scale RAID](#) and [Data Lifecycle Management](#). Since Spectrum Scale and Spectrum Scale are two different names for the same storage software, we will use the name Spectrum Scale in the subsequent sections.

2.2.1.1 IBM Spectrum Scale RAID

IBM Spectrum Scale RAID [17] eliminates the need for expensive storage RAID controllers because it implements RAID protection within the file system software. Consequently, IBM Spectrum Scale RAID stores the RAID protected data on standard storage devices, such as SSD or disk drives.

IBM Spectrum Scale RAID employs a de-clustered approach to RAID, reducing the impact of drive failures by spreading data over all the available storage devices. This improves application I/O and storage recovery performance. IBM Spectrum Scale RAID delivers high reliability through an 8+2P or 8+3P Reed-Solomon–based RAID algorithm that divides each block of a file into eight data chunks and 2 or 3 parity chunks. Even higher reliability can be achieved with three- or four-way replication that creates two or three copies for each data block written. These data protection algorithm scale easily, starting with as few as 11 storage devices and growing to over 300 per system.

Recovery times from disk failures are fast with IBM Spectrum Scale RAID, measured in minutes rather than hours. Because there is no single-disk bottleneck during the rebuild, the workload is evenly distributed over many disks minimizing the performance impact of a rebuild. In addition, with faster recovery

times of disk failures the likelihood of another disk failure within the same RAID-array is lower, minimizing the risk of data access loss.

IBM Spectrum Scale RAID also provides end-to-end checksum protection to ensure data integrity. Checksum information is stored on disk and verified all the way up to the Spectrum Scale file systems layer. In case the checksum of a data block does not match a rebuild of the data block is started automatically, using parity or block replicas.

Spreading the data over many devices provides more predictable storage performance because all devices are utilized equally. A storage administrator does not have to worry about RAID and LUN layout and distribution, IBM Spectrum Scale RAID takes care for all of this.

2.2.1.2 Data Lifecycle Management

IBM Spectrum Scale helps achieving data lifecycle management efficiencies through policy-driven automation and tiered storage management [16]. The use of storage pools and user-defined policies allow administrators to better match the cost of storage to the value of data.

An IBM Spectrum Scale file system is composed of storage pools that are used to manage groups of storage devices with the same characteristics. Storage pool allow creating tiers of storage by grouping storage devices based on performance, locality or reliability characteristics. For example, one pool contains volumes protected by 4-way replication providing high performance and reliability while another pool contains volumes protected by the 8+2P RAID code providing more capacity.

IBM Spectrum Scale allows setting up placement and migration policies to manage the lifecycle of files automatically [16]. A policy comprises one or more rules. A placement rule defines that files matching certain criteria are placed in a specific storage pool. For example the following policy example includes two rules placing file names ending with ".mp3" in "pool1" while all other files are placed in pool "system":

```
RULE 'mp3' SET POOL 'pool1' WHERE ((LOWER(NAME) LIKE '%.mp3')
RULE 'default' SET POOL 'system'
```

A migration policies defines the migration of files from one storage pool to another one. The file migration can be triggered automatically based on usage of file system pools. For example the following policy rule will kick in when the usage of system pool exceeds 80% and it will migrate files accessed the longest time ago from the system pool to pool1, until the pool usage drops down to 70 %:

```
RULE 'systemMigration' MIGRATE FROM POOL 'system' THRESHOLD(80,70)
WEIGHT(CURRENT_TIMESTAMP - ACCESS_TIME) TO POOL 'pool1'
```

Placement is fully automated once the placement policies are applied. Migration policies allow migrating files to tape storage (see section [IBM Spectrum Archive Enterprise Edition](#)). Migration based on threshold is also fully automated if the threshold based migration policies is configured for the file system. Other migration processes can be fully automated using Spectrum Scale automation

framework [19]. User access to files is transparent, independent on the storage medium the files are stored in.

2.2.2 IBM Spectrum Archive Enterprise Edition

IBM Spectrum Archive Enterprise Edition gives organizations an easy way to use cost-effective IBM tape drives and libraries within a tiered storage infrastructure. By using tape libraries instead of disks for data that is stored for long-term retention, organizations can improve efficiency and reduce costs. IBM Spectrum Archive EE seamlessly integrates with the scalability, manageability and performance of IBM Spectrum Scale [10] [14].

IBM Spectrum Archive EE is based on the LTFS standard maintained by the Linear Tape Open consortium (LTO) and Storage Network and Industry Association (SNIA [4]). Spectrum Archive EE enhances the capabilities of the common LTFS editions – single drive edition and library edition [11] – by providing advanced tape management functions for scaling environment.

Spectrum Archive EE can be used as an external pool for IBM Spectrum Scale file systems allowing files to be seamlessly migrated from the disk tier to LTFS tape while providing transparent access to the user (see [Data Lifecycle Management](#)). This means that even if a file has been migrated to LTFS tape, it remains visible to the user in the Spectrum Scale file system name space. When the user accesses the file Spectrum Archive EE will recall it from the tape and copy it back to the Spectrum Scale file system before the file can actually be opened by the user. This operation may take a few seconds to a couple of minutes because the tape has to be mounted and positioned.

IBM Spectrum Archive EE has certain advantages for archiving in a scale-out storage environment as demonstrated in the sections.

2.2.2.1 Characteristics

IBM Spectrum Archive EE *scales* with the IBM Spectrum Scale cluster. A subset of IBM Spectrum Scale cluster servers can be configured as IBM Spectrum Archive EE enabled servers. Each IBM Spectrum Archive EE enabled server has tape drives attached where all tape drives must be installed in the same tape library. IBM Spectrum Archive EE operations such as migrations, recalls, reclamation and reconciliation are distributed across all IBM Spectrum Archive EE enabled servers by the workload manager, allowing for optimal resource utilization.

All IBM Spectrum Archive EE enabled servers *share* the same tape cartridges in the tape library. Adding IBM Spectrum Archive EE enabled server is seamless and allows the archive solution to scale dynamically.

Migrations and recalls are *transparent* while IBM Spectrum Archive EE also supports pre-migration of files. With pre-migration files are not moved from disk to tape but copied. This allows having two copies of files, one on disk and one on tape. Subsequent migrations do not have to copy the files to tape again, but can just delete the file data in the IBM Spectrum Scale file system and create a stub file. This allows for very fast migrations in situations where the file system is overfilled. Pre-migration with IBM Spectrum Archive EE however is not a backup solution because if a file in the IBM Spectrum Scale file system is deleted, it cannot be easily restored from LTFS.

During migration and pre-migration *multiple copies* of files can be created on distinct tapes. This mitigates the risk of data loss resulting from a damaged tape. If a tape is damaged the files can be recovered from other copy tapes.

When many files need to be recalled within a short period of time Spectrum Archive EE offers the *tape optimized* bulk recall function. With the bulk recall function the administrator creates a list of files to be recalled and passes this to the `eeadm recall` command. Spectrum Archive EE will sort the files by the tape-ID and the position on tape before it sequentially recalls all files residing on one tape.

One of the key advantages of LTFS is that it provides a *standardized* file system on tape, making the use of tapes much simpler. A LTFS tape is self-describing because it includes the file system index (directory structure) and the files stored on the same tape cartridge. IBM Spectrum Archive EE creates the same IBM Spectrum Scale file and directory structure on the LTFS tapes. This fosters the interchangeability of data on tape, instead of using slow and unreliable WAN connections. Consequently, IBM Spectrum Archive EE supports the *export* and *import* of tapes.

IBM Spectrum Archive EE also provides functions for *automated housekeeping*. With reclamation space on under-utilized tapes is freed up by copying the active files to another tape. The reconciliation function synchronizes the IBM Spectrum Scale file system structure with the file system structure on LTFS accounting for file renames and deletions in the IBM Spectrum Scale file system. Both functions can be automated by schedules.

Finally, IBM Spectrum Archive EE integrates with IBM Spectrum Scale immutable fileset [18]. These allow to set file to immutable with a retention period during which the files cannot be modified or deleted. After the expiration of the retention time an immutable file can be deleted, but not modified. IBM Spectrum Archive EE can migrate immutable files to WORM tapes ensuring that immutable files cannot be modified or deleted even if they reside on tape.

2.2.2.2 Migration Policies

The migration from disk to tape within a IBM Spectrum Scale file system is fully automated, utilizing the IBM Spectrum Scale policy engine (see section [Data Lifecycle Management](#)). It is possible to migrate files from a disk pool to LTFS tape when the disk pool has reached a certain capacity. The following policy example will migrate files from the internal pool „pool1“ to an LTFS tape in Spectrum Archive pool “EEPOOL” when the usage of pool1 exceeds 90%:

```
/*define the external pool being Spectrum Archive EE */
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/eeadm'
OPTS 'EEPOOL'

/* Migrate from pool1 to ltfs if 90% are reached, only migrate files
having a size greater 0 KB */
RULE 'MigToTape' MIGRATE FROM POOL 'pool1' THRESHOLD(90,70)
  WEIGHT(CURRENT_TIMESTAMP - ACCESS_TIME) TO POOL 'ltfs' WHERE
  (KB_ALLOCATED > 0)
```

The OPTS argument of the first rule defines the name of the IBM Spectrum Archive EE tape pool (EEPOOL) which is used as destination for the migration.

With Spectrum Archive EE version 1.2 the syntax of the tape pool has changed to `poolname @ libraryname`. With Spectrum Archive EE version 1.2 up to two tape

libraries can be configured in one Spectrum Archive EE cluster, necessitating the addition of the library name to the pool name. Find below an example for defining two destination pools in two different libraries in one external pool rule:

```
/*define the external pool with two destination pools (R 1.2 format*/  
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/ltfsee'  
OPTS '-p pool1@lib1 pool2@lib2'
```

Another way of running migrations can be based on periodic schedules, where files matching certain criteria are migrated. The following policy example shows the migration of files from "pool1" to pool "ltfs" which have not been accessed for 30 days:

```
/* Define LTFS as external pool */  
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/ltfsee'  
OPTS '-p pool1@lib1'  
  
/* Migrate files from pool1 to ltfs if they have not been accessed  
for more than 30 days*/  
RULE 'MigAge' MIGRATE FROM POOL 'pool1' TO POOL 'ltfs'  
WHERE (KB_ALLOCATED > 0) AND  
(DAYS(CURRENT_TIMESTAMP) - DAYS(ACCESS_TIME) > 30)
```

The range of migration policies is comprehensive and allows for great flexibility for migrating files from disk to tape based on file attributes. More examples for file attributes used for migration in the context of archiving are:

- File size
- Files and directory names
- Files types based on file names
- File ownership

3. IBM Spectrum Archive solutions

The archive solution presented in this chapter is based on IBM Elastic Storage Server systems (ESS) running the Spectrum Scale software and representing a global file system name space to users and applications. Figure 1 shows this archive solution and separates it into different domains that are explained below:

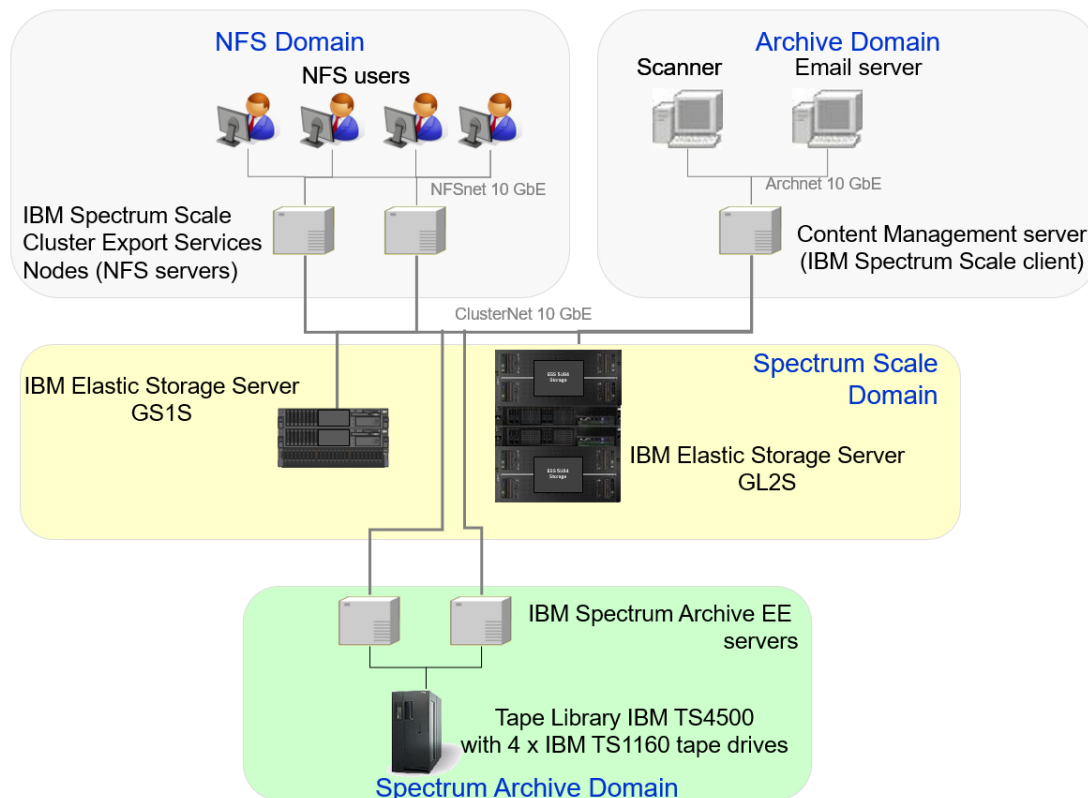


Figure 1: Spectrum archive solution

The central part of the solution is the *IBM Spectrum Scale domain* based on two ESS building blocks representing one Spectrum Scale cluster. One ESS building block is the model GS1S including 24 SSD drives. This building block is used for file system metadata. The second ESS building block is model GL2S including 166 NL-SAS drives. This building block is used for the actual file system data. The ESS building blocks provide two file systems to the users and applications and one file system solely used by Spectrum Archive EE for metadata. For more information about the ESS building block configuration refer to section [Elastic Storage Server \(ESS\) configuration](#).

The IBM Spectrum Scale cluster includes five more IBM Spectrum Scale servers that are configured as Spectrum Scale clients: two IBM Spectrum Scale Cluster Export Services servers in the NFS domain, one application server in the archive domain and two IBM Spectrum Archive EE servers in the IBM Spectrum Archive domain. As member of the Spectrum Scale cluster all servers have access to the Spectrum Scale file systems provisioned by the ESS building blocks. Conceptually, it is possible to consolidate the five additional servers by assigning multiple roles to one server. For example, with two servers each server would act as NFS server, archive application server and IBM Spectrum Archive EE server. In order to make the description of this archive solution understandable each server has one role. This also allows creating different administrative domains.

In the *NFS domain* users access the user files system through NFS. For this reason, two additional servers are installed and configured with Cluster Export Service (CES). CES is a function of IBM Spectrum Scale and provides clustered NFS services. The CES servers are configured as IBM Spectrum Scale clients and as such part of the IBM Spectrum Scale cluster. The CES servers export the Spectrum Scale files system for users via NFS. For more information about the NFS server configuration refer to section [Network File System users](#).

In the *Archive domain* the Content Management Application (CM) is installed on an additional server which is also configured as an IBM Spectrum Scale client and accesses the file system used by the CM application. The CM application obtains data from various sources, indexes it and stores it in the file system managed by the ESS building blocks. For more information about the archive application configuration refer to section [Archive application](#).

The *Spectrum Archive domain* comprises two IBM Spectrum Archive EE servers that are configured as IBM Spectrum Scale clients within the IBM Spectrum Scale cluster. The two Spectrum Archive EE servers perform the archiving process by selecting and migrating files from both file systems (*users* and *archive*) to tape. For this reason, two IBM TS1160 tape drives in one IBM TS4500 tape library are attached to each Spectrum Archive EE server. IBM Spectrum Archive EE version 1.2 and above support up to two libraries to be connected to one cluster. In this example we will focus on one tape library.

The Spectrum Archive EE servers use an extra file system for LTFS metadata allowing exchanging the tape inventory and status information.

3.1 IBM Elastic Storage Server (ESS) configuration

The IBM ESS building blocks provide three file systems used by the NFS and Archive domain and space managed by the IBM Spectrum Archive EE domain. Each IBM ESS building block includes two servers running the IBM Spectrum Scale RAID and standard storage devices, such as SSD and NL-SAS disk drives. No RAID controllers are required because each IBM ESS building block provides RAID protection on the storage devices (see section [Spectrum Scale RAID](#)).

3.1.1 File systems

The two IBM ESS building blocks provide three file systems. The file system "*users*" is exported via NFS by the CES servers and used by the NFS clients in the NFS domain. The file system "*archive*" is used by the content management application as POSIX file system in the Archive domain. And the file system "*ltfsmeta*" is used by IBM Spectrum Archive EE servers in the Spectrum Archive domain for metadata. Figure 2 shows the *users* file system configuration in the Elastic Storage cluster:

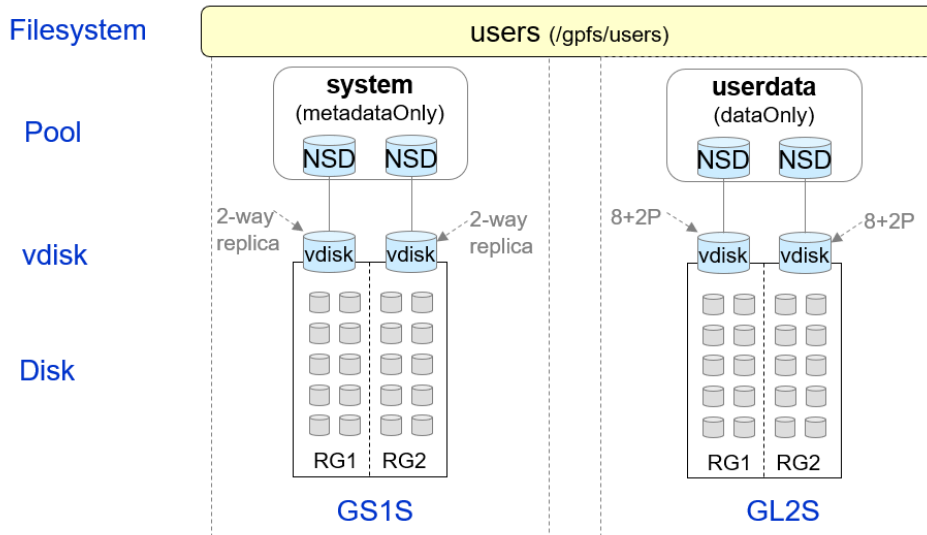


Figure 2: File system layout in IBM Elastic Storage building blocks

The *users* file system has two pools: *system* and *userdata*. The *system* pool is configured with two NSD using two vdisks configured in distinct recovery groups of the GS1S system. The NSD of the *system* pool are configured for metadata only usage. The file system "*users*" is mounted on */gpfs/users*.

The *userdata* pool is configured with two NSD using two vdisks configured in distinct recovery groups of the GL2S system. The NSD of the *userdata* pool are configured for data only usage.

The file system *archive* is configured in a similar fashion, the pools are named *system* and *archivedata*. The file system *archive* is mounted on */gpfs/archive*.

The *lftsmeta* file system does not have metadata on the GS1S system. Instead it uses one pool (*system*) configured on the GL2S system with 8+2P RAID protection. All NSD are configured for data and metadata usage. This file system *lftsmeta* is also used by CES servers for the CES shared root directory.

3.2 Application server configuration

There are two kinds of applications using the global file systems provided by IBM ESS: CES servers exporting NFS exports and content management application for archiving.

3.2.1 Network File System users

More than 100 users can access the IBM Spectrum Scale file system *users* through NFS. Each user has their own NFS export represented by a subdirectory in the *users* file system. In order to provide resiliency two clustered NFS servers are implemented.

IBM Spectrum Scale CES allows for configuring a subset of the servers in the cluster providing high availability for exporting IBM Spectrum Scale file systems using NFS. With CES NFS a failure of one CES server is automatically recovered

by IBM Spectrum Scale, moving the NFS server IP address from the failing CES server to another CES server.

The two CES servers are based on standard x86 servers running Red Hat Linux and the IBM Spectrum Scale software. Both CES servers are members of the Spectrum Scale cluster managed by the ESS building blocks. As shown in Figure 1, the two CES servers connect to two separate networks. One 10 GBit Ethernet network (NFSnet) connects the CES servers to the users providing access to the NFS exports. The other 10 GBit Ethernet network (ClusterNet) connects the CES servers with the IBM Spectrum Scale cluster and provides them access to the IBM Spectrum Scale file system *users* (see section [File systems](#)). It is possible to consolidate the two networks, however from a performance perspective it has been decided to use separate networks.

IBM Spectrum Scale CES is configured to export one directory of the *users* file system (/gpfs/users) for each user. Each exported directory is an IBM Spectrum Scale file set. A file set is a partition of the IBM Spectrum Scale file system – represented by a directory - allowing for management functions such as quota, snapshot, placement, migration, backup, active file management etc. The CES configuration requires configuring a shared root directory that is accessible for all CES servers. For efficiency reasons, the shared root directory required by CES is configured on the file system ltfsmeta as (/gpfs/ltfsmeta/CESrootdir). Each CES server has 10 IP addresses configured as CES interfaces, so the users have 20 IP addresses to access their NFS export. Load balancing is implemented with DNS round-robin.

The archive policy for the NFS users is based on last access dates of the files. In order to optimize storage cost, files that have not been accessed for 1 year will be migrated to tape. In addition, if the file system utilization is above 90 % files will be migrated to LTFs tape until the file system utilization drops below 70%. The actual implementation of the archive policies is shown in section [Migration policies for NFS users](#).

3.2.2 Archive application

The archive application is a content management system (CM) that archives files in a file system provided by IBM Spectrum Scale (file system *archive*). The CM system is connected to different client systems creating the files to be archived. One client system is a scan station that scans paper documents. Scanned documents are sent to the CM system that indexes (using optical character recognition, OCR) and stores these documents in the file system. Another client is an email server that copies all incoming and outgoing emails to the CM system. The CM system indexes the emails and stores them in files in the file system.

The CM software runs on a standard x86 server under Red Hat Linux and IBM Spectrum Scale. As shown in Figure 1, the CM server connects to two separate networks. One 10 GBit Ethernet network (Archnet) connects to the clients generating the files to be archived (scanner and email server). The other 10 GBit Ethernet network (ClusterNet) connects the CM servers with the IBM Spectrum Scale cluster and provides them access to the Spectrum Scale file system *archive*. Using two networks is not mandatory but has been decided for performance reasons. Note, it is possible to provision two servers for the CM application in order to provide high availability. To simplify the archive solution only one server is used.

The CM system is comprised of a database - for storing and managing the index information and metadata – and a file storage - for storing the archived files and documents. Both entities are stored on the Elastic Storage file system named *archive* (see section [File systems](#)).

Note, it has been considered – but not yet implemented - storing the CM database on faster SSD storage with lower latency facilitating better database performance. There are two ways to accomplish this. The first one is to create a new file system on the IBM ESS GS1S building block that comprises SSD drives. The other way is to allocate more storage capacity on the GS1S building block for the pool 'system' and use placement policies to store data on SSD (see section [File systems](#)). In this case the newly allocated storage capacity must have the NSD usage type dataAndMetadata. Here is an example of a placement policy that stores all files in directory `/gpfs/archive/dbdirectory` on SSD provided by the GS1S building block:

```
RULE 'cmdb' SET POOL 'system'  
  WHERE ((LOWER(NAME) LIKE '/gpfs/archive/dbdirectory')  
  
RULE 'default' SET POOL 'archivedata'
```

Alternatively, the database directory (`/gpfs/archive/dbdirectory`) can be configured as file set. A file set is a logical partition in a file system allowing for placement of files stored in the directory represented by the file set. When the database is stored in a directory `/gpfs/archive/dbdirectory` represented by file set named "dbfileset", the placement policy for the database may look like this:

```
RULE 'cmdb' SET POOL 'system' FOR FILESET ('dbfileset')  
  
RULE 'default' SET POOL 'archivedata'
```

The archive policy for the CM system is based on file system utilization. If the file system utilization is above 80% files will be migrated to tape until the file system utilization drops below 70%. In addition two copies of all migrated data will be created on two different sets of tapes. The actual implementation of the archive policies is shown in section [Migration policies for Archive application](#).

3.3 Spectrum Archive configuration

The IBM Spectrum Archive EE software integrating with Spectrum Scale - as described in section [Spectrum Archive Enterprise Edition](#) - is installed on two additional servers. Both servers represent an IBM Spectrum Archive EE system and are members of the IBM Spectrum Scale cluster. As Spectrum Scale clients both servers have access the file systems to be managed (*users* and *archive*) and the Spectrum Archive EE metadata file system (*Itfsmeta*).

The following section describe the Spectrum Archive EE configuration and policies used to archive the file systems according to the requirements of the applications (see section [Application servers](#)).

3.3.1 IBM Spectrum Archive EE server setup

An IBM Spectrum Archive EE environment must be sized according to the requirements. IBM provides a sizing tool to help identifying the appropriate infrastructure [6].

In this example, the Spectrum Archive EE servers are standard x86 servers comprising 2 cores, 64 GB RAM and 2 internal 300 GB drives for the operating system and software. Both servers are connected to the ESS building blocks via dual port 10 GB Ethernet connections (network ltfnet). Each server also has one dual port Fibre Channel card which connects to the tape SAN. The tape SAN is comprised of two SAN fabrics for redundancy where each Spectrum Archive EE server is connected to the tape drives on both SAN fabric.

The operating system is Red Hat Linux and is installed on the two internal drives configured in RAID-1 mode. In addition, the IBM Spectrum Scale software is installed and configured on both servers. The ESS building blocks had already been configured with the file systems which are available to the Spectrum Archive EE servers after adding them to the cluster (see section [File systems](#)).

All four IBM TS1160 tape drives are zoned to both IBM Spectrum Archive EE servers via the redundant SAN fabric. During the IBM Spectrum Archive EE configuration each server gets two distinct tape drives assigned. All tape drives are in the same logical partition of the tape library. In order to provide control paths failover, each tape drive is configured as control path device.

To install and configure IBM Spectrum Archive EE on both servers the following prerequisites must be satisfied:

- IBM Spectrum Scale must be installed and configured on Spectrum Archive EE servers
- The IBM Spectrum Archive EE servers must be part of the Spectrum Scale cluster
- The file systems to be managed (users, archive) must be mounted on the Spectrum Archive EE servers
 - File system users is mounted under `/gpfs/users`
 - File system archive is mounted under `/gpfs/archive`
- The file system for LTFS metadata (ltfsmeta) must be mounted on the Spectrum Archive EE servers
 - File system ltfsmeta is mounted under `/gpfs/ltfsmeta`
- The tape drive and library resources must be available to each Spectrum Archive EE server
 - Each server sees all four tape drives and four medium changer devices on two paths
 - Use the `itdt` tool provided by IBM [5] to ensure that the tape drives are available on both servers.
- Spectrum Archive EE software must be available

The Spectrum Archive EE installation and configuration consists of the following steps [10] [14] that must be carried out on each IBM Spectrum Archive server:

- Install the Linux prerequisites according to the requirements on both servers
- Install the Spectrum Archive EE software by extracting the package and running the command on both servers: `ltfsee_install -i`
- Configure the file system to be managed (users and archive) by Spectrum Archive EE with DMAPI capabilities using the command: `mmchfs <filesystemname> -z yes`
 - This requires un-mounting the file systems!

- Configure Spectrum Archive EE on first server: `ltfsee_config -m CLUSTER`
 - Select the file systems *users* and *archive* as space managed file systems
 - Select the file system *ltfsmeta* as LTFS metadata file system
- Configure Spectrum Archive EE on secondary servers: `ltfsee_config -m ADD_NODE`
 - Select the two tape drives connected to the server. Each server must have configured a distinct set of two tape drives.
- Start Spectrum Archive EE: `eadm cluster start`

After Spectrum Archive EE has been successfully installed and configured the IBM Spectrum Archive EE resources can be checked using the following commands:

Resource to be checked	Spectrum Archive EE command
System status	# eadm node list
Spectrum Archive EE servers	# eadm node list
Tape drives	# eadm drive list
Tape cartridges	# eadm tape list
Tape pools	# eadm pool list

To migrate files to tape, tape pools must be created, and tapes must be assigned. An IBM Spectrum Archive EE pool is comprised of a distinct set of tapes. A total of 3 pools need to be created: one for the file system *users* and two for the file system *archives* (because two redundant copies are required). To create a pool and add formatted tapes perform the following commands:

1. Create a pool: `eadm pool create poolname`
2. Format and add tapes: `eadm tape assign tape-id -p poolname -f`

For the file system *users* the tape pool *users_pool* has been created and tapes have been assigned to this pool .

For the file system *archive* the LTFS tape pools *archive_pool1* and *archive_pool2* have been created and tapes have been assigned to this pool.

Once the LTFS tape pools are created the policies can be implemented as shown in the next section.

3.3.2 Migration policies for NFS users

The archive requirements for the file system *users* - used by the NFS users - are the following (see also section [Network File System users](#)):

- Periodically migrate files that have not been accessed for 1 year
- Automatically migrate files when the file system utilization is above 90 % until the utilization drops below 70%
- Migrate to one tape pool (named *users_pool*)

The *users* file system (`/gpfs/users`) has two internal disk pools: *system* and *userdata*. The *system* pool is used exclusively for metadata (see section [File systems](#)). Thus, the migration source pool is the pool "userdata".

The periodic migration policy includes the following rules to migrate files from the file system `/gpfs/users` and pool "userdata" to the tape pool *users_pool* when they have been last accessed more than 365 days ago and have a size greater

than 0. The last condition assures that files which are already migrated are not selected.

```
/* Common excludes with Spectrum Archive */
RULE 'exclude' EXCLUDE WHERE
  ( PATH_NAME LIKE '%/.SpaceMan/%' OR
    PATH_NAME LIKE '%/.ltfsee/%' OR
    PATH_NAME LIKE '%/.mmSharedTmpDir/%' OR
    PATH_NAME LIKE '%/.snapshots/%' )

/* Define LTFS as external pool and specify the tape pool */
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/eeadm'
OPTS '-p users_pool@lib1'

/* Migrate files from userdata to ltfs if they have not been accessed
for more than 365 days*/
RULE 'usersDaily' MIGRATE FROM POOL 'userdata' TO POOL 'ltfs' WHERE
  (DAYS(CURRENT_TIMESTAMP) - DAYS(ACCESS_TIME) > 365) AND
  (KB_ALLOCATED > 0)
```

This policy uses an EXCLUDE rule to exclude certain files and directories [16]. This policy can be stored in a file.

The periodic migration policy is started every day using the following command, which might be run by scheduler such as cron or timerd:

```
# mmapplypolicy /gpfs/users -P users_policy.txt -m 2 -N
ltfsnode1,ltfsnode2 -B 1000 --single-instance
```

The parameters passed to `mmapplypolicy` have the following meaning [16]:

-P users_policy.txt	file containing the policy
-m 2	limit the number of migration thread to 2
-N ltfsnode1,ltfsnode2	IBM Spectrum Archive EE nodes run this policy
-B 1000	migrate 1000 files per migrate job
--single-instance	only one instance of mmapplypolicy should be run

Note, for running storage services like Spectrum Archive migrations on a periodic basis there is a framework for automation of Spectrum Scale storage services [19]. This framework essentially resolves some of the challenges with running storage service within a Spectrum Scale cluster.

The automatic migration policy includes the following rules to migrate files from the file system `/gpfs/users` and pool "userdata" to tape pool `users_pool` when the IBM Spectrum Scale pool "userdata" has reached 90%. Files which have been accessed longest ago are selected first for migration.

```
/* Common excludes with Spectrum Archive */
RULE 'exclude' EXCLUDE WHERE
  ( PATH_NAME LIKE '%/.SpaceMan/%' OR
    PATH_NAME LIKE '%/.ltfsee/%' OR
    PATH_NAME LIKE '%/.mmSharedTmpDir/%' OR
    PATH_NAME LIKE '%/.snapshots/%' )

/* Define LTFS as external pool and specify the tape pool */
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/eeadm'
OPTS '-p users_pool@lib1'

/* Migrate files from pool userdata to ltfs at 90% threshold*/
```

```
RULE 'usersAuto' MIGRATE FROM POOL 'userdata' THRESHOLD(90,70)
WEIGHT(CURRENT_TIMESTAMP - ACCESS_TIME) TO POOL 'ltfs'
WHERE (KB_ALLOCATED > 0)
```

To apply the automatic migration policy to the file system *users* the following command must be run, specifying a *policyfile* including the above policy and other policies, such as placement policies, if required (see section [Data Lifecycle Management](#)).

```
mmchpolicy users -P policyfile
```

To trigger a policy run when the threshold is reached a callback must be configured, see section [Configuring migration callback](#).

3.3.3 Migration policies for Archive application

The archive requirements for the file system *archive* - used by the content management application - are the following (see also section [Archive application](#)):

- Automatically migrate files when the file system utilization is above 80 % until the utilization drops below 70%
- Migrate to two LTFS tape pool (named *archive_pool1* and *archive_pool2*)

The *archive* file system (*/gpfs/archive*) has two pools: *system* and *archivedata* whereby the pool “*system*” is used exclusively for metadata (see [File systems](#)). Thus, the migration source pool is the pool “*archivedata*”.

The automatic migration policy includes the following rules to migrate files from the file system */gpfs/archive* and Spectrum Scale pool *archivedata* to LTFS pools *archive_pool1* and *archive_pool2* when the Spectrum Scale pool *archivedata* has reached 80%. Files which have been accessed longest ago are selected first for migration.

```
/* Common excludes with Spectrum Archive */
RULE 'exclude' EXCLUDE WHERE
  ( PATH_NAME LIKE '%/.SpaceMan/%' OR
    PATH_NAME LIKE '%/.ltfsee/%' OR
    PATH_NAME LIKE '%/.mmSharedTmpDir/%' OR
    PATH_NAME LIKE '%/.snapshots/%' )

/* Define LTFS as external pool and specify the tape pool */
RULE EXTERNAL POOL 'ltfs' EXEC '/opt/ibm/ltfsee/bin/eeadm'
OPTS '-p archive_pool1@lib1 archive_pool2@lib1'

/* Migrate files from pool archivedata to ltfs at 80% threshold*/
RULE 'archiveAuto' MIGRATE FROM POOL 'archivedata' THRESHOLD(80,70)
WEIGHT(CURRENT_TIMESTAMP - ACCESS_TIME) TO POOL 'ltfs'
WHERE (KB_ALLOCATED > 0)
```

To apply the automatic migration policy to the file system *archive* the following command must be performed, specifying a *policyfile* including the above policy and other policies, such as placement policies, if required (see section [Data Lifecycle Management](#)).

```
mmchpolicy archive -P policyfile
```

In order to trigger a policy run when the threshold is reached a callback must be configured, see section [Configuring migration callback](#).

3.3.4 Configuring migration callback

To start an automatic migration policy when a certain file system pool utilization threshold is reached, a callback must be configured. A callback in IBM Spectrum Scale is triggered upon defined events and invokes a script. The callback also allows sending additional parameters to this script. A Spectrum Scale callback has a cluster-wide scope and is invoked for all file systems when one of the events is triggered.

To configure a callback for migration to tape, the events "lowDiskSpace" and "noDiskSpace" must be specified. There is a standard script which starts the policies configured for the file systems (/usr/lpp/mmfs/bin/mmstartpolicy). The following creates a migration callback:

```
mmaddcallback MIGRATION --command /usr/lpp/mmfs/bin/mmstartpolicy
--event lowDiskSpace,noDiskSpace
--parms "%eventName %fsName -m 3 -N ltfsnode1,ltfsnode2 -B 1000 -n 1
--single-instance"
```

This callback is named MIGRATION and is triggered by the events "lowDiskSpace" and "noDiskSpace". When one of these events occurs then Spectrum Scale automatically invokes the callback script /usr/lpp/mmfs/bin/mmstartpolicy with the parameters "eventName, fsname" and some additional parameters required for mmapplypolicy (for an explanation of these parameters see section [Migration policies for NFS users](#)). The script /usr/lpp/mmfs/bin/mmstartpolicy invokes the mmapplypolicy command with the file system name and the parameters given with the --parms option of the callback. The command mmapplypolicy will run the policy applied with mmchpolicy.

4. Summary

The innovative IBM Spectrum Archive solution outlined in this document combines traditional scale-out file storage with archiving capabilities, allowing files to be stored on the most appropriate storage medium. From a user perspective access to files is transparent, regardless in which storage tier (disk or tape) the files are stored. The archiving process is fully automated and provides a great variety of archiving rules.

The combination of IBM Spectrum Scale and IBM Spectrum Archive Enterprise Edition meets the requirement for archiving - as outlined in section [Archiving requirements and techniques](#) - by providing techniques to:

- Optimize total cost of ownership
- Provide flexibility and scalability
- Facilitate data availability and protection
- Prevent deletion and modification of compliance data
- Provide future proof archive system

These techniques are detailed in the following sub-sections.

4.1.1 Total cost of ownership

The file archive solution presented leverages Information Lifecycle Management concepts to optimize total cost of ownership. It utilizes the data lifecycle management function in IBM Spectrum Scale (see section [Data Lifecycle Management](#)) which not only places the files on the most appropriate storage medium, but also allows migrating files during the lifecycle, matching the files information value with the cost of storage. When using tape as the storage tier for files that are no longer accessed, the highest cost saving can be achieved [1].

Further cost optimization is achieved by the automation of placement and migration processes built into Elastic Storage and Spectrum Archive EE, combined with transparent file access for users.

4.1.2 Scalability

This archive solution can scale-out in multiple dimensions. One dimension is performance and capacity for primary data which can be addressed by adding ESS building blocks (see section [File Archive solution](#)). Another dimension is performance and capacity for archived data which can be addressed by adding IBM Spectrum Archive EE servers and tape resources (see section [IBM Spectrum Archive Enterprise Edition](#)). And yet a third dimension is that application domains and users can be flexibly added to the Elastic Storage cluster.

The control for placement and migration is based on a comprehensive set of rules and policies. Files can be selected for migration based on a variety of attributes providing superior flexibility. Adapting new types of files to be archived is seamless and just requires new rules to be added.

4.1.3 Data availability and protection

The IBM Spectrum Scale file system architecture is based on clustering, providing high availability. With this archive solution files are continuously accessible regardless where they are stored.

The innovative IBM Spectrum Scale RAID technique implemented in the IBM Elastic Storage Server provides superior fault-tolerance for data (see section [IBM Spectrum Scale RAID](#)). In addition, files can be replicated to another IBM Spectrum Scale cluster in a secondary site leveraging the IBM Spectrum Scale Active File Management function. The cluster in the secondary site can be quickly activated upon a disaster.

In addition, IBM Spectrum Archive EE has capabilities to write two or three copies of files on distinct set of tapes. This does not only help to recover from failures reading data from tape, it also provides high availability because data is automatically read from copy tapes when the primary tape is damaged (see section [Characteristics](#)).

4.1.4 WORM protection

IBM Spectrum Archive integrates with the IBM Spectrum Scale immutability function [18] that provides WORM protection for files. Immutable file cannot be modified or deleted during the retention period. After the expiration of the retention period immutable files can be deleted but not modified.

IBM Spectrum Archive EE can migrate immutable files from the Spectrum Scale file system to WORM tapes where file remain immutable and cannot be modified or deleted. With this solution Spectrum Archive can be used for archiving data subject for regulatory compliance.

4.1.5 Future proof

In order to cope with technology obsolescence – one of the key challenges for archiving – migration of software, systems and data is inevitable [3]. The file archive solution presented herein is comprised of standard software and hardware that can be transparently replaced with newer components. One of the facilitators for this is Spectrum Scale providing concurrent upgrade capabilities.

In addition, this solution provides tiered storage capabilities allowing seamless migration of files from one storage medium to another (see section [Data Lifecycle Management](#)). Spectrum Archive EE supports different tape technologies coexisting within one system and facilitates the migration from one tape technology to another.

Standardized interfaces within an archive system are important because they prevent vendor lock-in and have long lifecycles with sufficient backward compatibility. The IBM Spectrum Scale file system as an interface to applications is based on the POSIX standard which fosters interoperability with many applications and provides long lifetimes. IBM Spectrum Scale Cluster Export Services provide high available access to data via standardized protocols NFS, SMB, OpenStack Swift and S3. IBM Spectrum Archive EE utilizes the market leading Linear Tape File System format which has been standardized by the Storage Networking and Industry Association (SNIA) [4]). The standardized data format makes an LTFs tape self-describing, allowing it to be read even years after it had been created without requiring an external index or interpreter.

5. Appendix

5.1 References

[1] Whitepaper - Analyzing the Economic Value of LTO Tape for Long-term Data Retention

https://www.lto.org/wp-content/uploads/2014/06/ESG-WP-LTO-EVV-Feb_2016.pdf

[2] Redbook: IBM Spectrum Scale (formerly GPFS)

<http://w3.itso.ibm.com/abstracts/sg248254.html?Open>

[3] SNIA Study on 100 year archive

<http://www.snia-europe.org/en/technology-topics/data-protection/dpc-publications.cfm/dpcosurvey>

[4] SNIA LTFS Format Specification

http://www.snia.org/tech_activities/standards/curr_standards/ltfs

[5] IBM Tape Diagnostic Tool (itdt)

<https://www.ibm.com/support/pages/ibm-tape-diagnostic-tool-itdt-v94>

[6] IBM Spectrum Archive EE sizing tool

<https://www.ibm.com/support/pages/node/6148113>

[8] IBM Elastic Storage Server (ESS):

<https://www.ibm.com/products/ibm-elastic-storage-server>

[10] IBM Spectrum Archive EE Knowledge Center:

https://www.ibm.com/support/knowledgecenter/en/ST9MBR_1.3.0/ltfs_ee_ichome.html

[11] IBM Spectrum Archive Editions:

<https://www.ibm.com/products/data-archive>

[14] Redbook: IBM Spectrum Archive EE Installation and Configuration Guide

<http://www.redbooks.ibm.com/abstracts/sg248333.html?Open>

[16] IBM Spectrum Scale ILM Policy Guide for IBM Spectrum Archive EE:

Internet: <https://www.ibm.com/support/pages/node/6260749>

[17] IBM Spectrum Scale RAID

https://www.ibm.com/support/knowledgecenter/en/SSYSP8_5.3.6/com.ibm.spectrum.scale.raid.v5r05.adm.doc/bl1adv_introduction.htm

[18] Red paper: Introduction to IBM Spectrum Scale immutability function:

<http://www.redbooks.ibm.com/abstracts/redp5507.html>

[19] Automation of storage services in an IBM Spectrum Scale cluster

<https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102676>

5.2 Disclaimer and Trademarks

This document reflects the understanding of the author in regard to questions asked about archiving solutions with IBM hardware and software. This document is presented "As-Is" and IBM does not assume responsibility for the statements expressed herein. It reflects the opinions of the author. These opinions are based on several years of joint work with the IBM Systems group. If you have questions about the contents of this document, please direct them to the Author (nils_haustein@de.ibm.com).

The Techdocs information, tools and documentation ("Materials") are being provided to IBM Business Partners to assist them with customer installations. Such Materials are provided by IBM on an "as-is" basis. IBM makes no representations or warranties regarding these Materials and does not provide any guarantee or assurance that the use of such Materials will result in a successful customer installation. These Materials may only be used by authorized IBM Business Partners for installation of IBM products and otherwise in compliance with the IBM Business Partner Agreement."

The following terms are trademarks or registered trademarks of the IBM Corporation in the United States or other countries or both: IBM, IBM Spectrum Archive, IBM Spectrum Scale, IBM Spectrum Protect.

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a world-wide basis.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Other company, product, and service names may be trademarks or service marks of others.