IBM Systems
White paper

Cross industry

IBM System Storage SAN Volume Controller

IP-based quorum support in version 7.6.0

This white paper provides a brief introduction to the IBM System Storage SAN Volume Controller (SVC) cluster with the Internet Protocol (IP) based quorum support feature in the 7.6.0 release. It includes an additional command-line interface (CLI) introduced for this feature along with the details of deployment with screen captures of IP-based quorum deployment.

Contents

Getting started

This section gives a brief idea about the basics of the IBM® System Storage SVC cluster with the IP-based quorum support feature along with the implementation details. This feature enables the use of a low-cost IP network-attached host as a quorum disk for simplified implementation and operation.

Introduction to quorum disk

A quorum disk is a managed disk (MDisk) that contains a reserved area for use exclusively by the system. The system uses quorum disks to break a tie when exactly half the nodes in the system remain after a failure and half of the nodes die at the same time. This situation is referred to as split brain. There are three candidate quorum disks that exist in the system. However, only one quorum disk is active at any time.

In other words, a quorum disk is a disk that contains a reserved area that is used exclusively for cluster membership management. The quorum disk is accessed when it is necessary to determine which half of the clustered system continues to read and write data. Quorum disks can either be MDisks or drives in the storage controller cluster.

There are two main use cases of quorum disk.

- As a tiebreaker if there is a storage area network (SAN) fault (when half of the nodes in the cluster die at same time)
- To hold a copy of important system configuration data

The quorum disk saves data slightly above 256 MB which is reserved for this purpose on each quorum disk candidate. The storage controller cluster system automatically selects the active quorum disk from the pool of assigned quorum disk candidates. If a tiebreaker condition occurs, one-half portion of the system nodes, which can reserve the quorum disk after the split-brain condition occurred, locks the quorum disk and continues to operate. At this stage, the other half stops its operation. This design prevents both the sites from claiming ownership of the cluster and start their own configuration nodes.

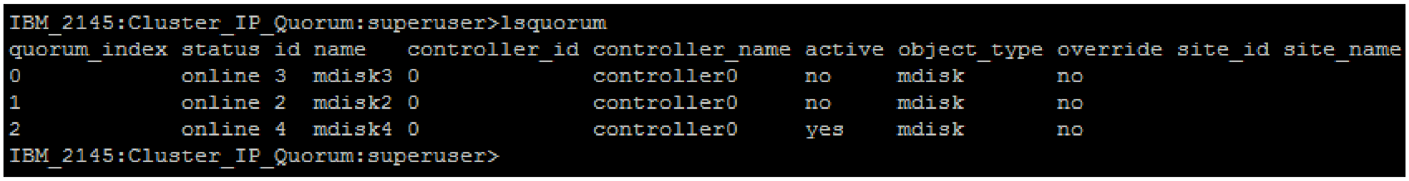
When MDisks are added to the SVC cluster, the system checks the MDisks to see whether they can be used as quorum disks. If the MDisks fulfill these requirements, then SVC assigns the MDisks that are added to the system as quorum candidates. One of these MDisks is selected as the active quorum disk. An internal mechanism of SVC places the quorum candidates on separate disk subsystems if possible. After the quorum disk is selected, no attempt is made to ensure that the other quorum candidates are presented through separate disk subsystems or not. This mechanism of quorum disk placement verification and adjustment to separate storage systems (if possible) reduce the dependency from a single storage system and can increase the quorum disk availability significantly.

Quorum disk commands

This section describes the basic commands of IBM SVC related to quorum disk with sample screen captures. For more details about the commands, refer:

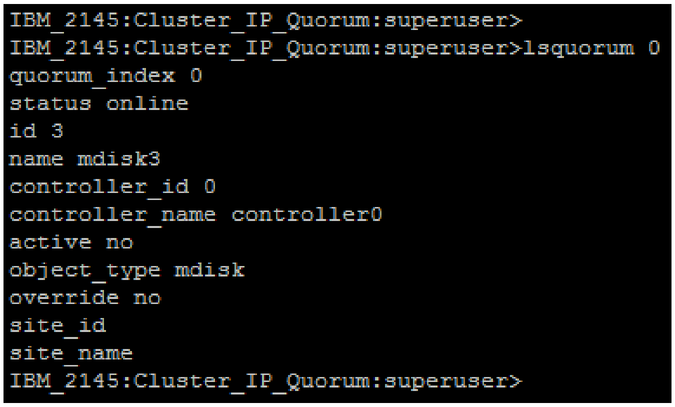
ibm.com/support/knowledgecenter/STPVGU_7.6.0/com.ibm.storage.svc.console.760.doc/svc_clicommandscontainer_229g0r.html?lang=en

lsquorum: This command displays a concise list or a detailed view of the MDisks or drives that the system is currently using to store quorum data. This information can be used to ensure that the quorum candidates are on separate storage subsystems.



```
IBM_2145:Cluster_IP_Quorum:superuser>lsquorum
quorum_index status id name   controller_id controller_name active object_type override site_id site_name
0            online 3  mdisk3 0             controller0     no     mdisk       no
1            online 2  mdisk2 0             controller0     no     mdisk       no
2            online 4  mdisk4 0             controller0     yes    mdisk       no
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 1: A concise invocation example of the lsquorum command

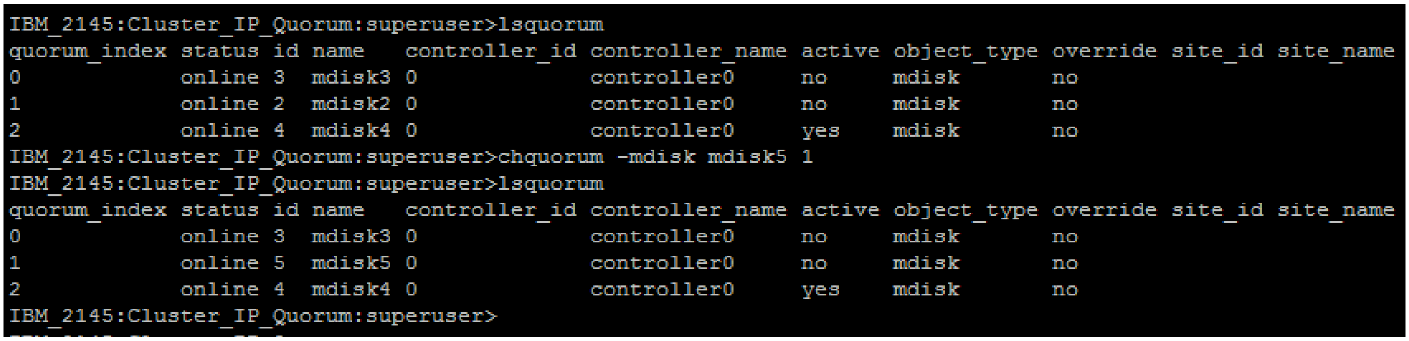


```
IBM_2145:Cluster_IP_Quorum:superuser>
IBM_2145:Cluster_IP_Quorum:superuser>lsquorum 0
quorum_index 0
status online
id 3
name mdisk3
controller_id 0
controller_name controller0
active no
object_type mdisk
override no
site_id
site_name
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 2: A concise invocation example of lsquorum detailed view

chquorum: This command is used to change the quorum disk association. To identify the drive or MDisk that owns the current active quorum disk, users need to run the lsquorum command.

The chquorum command is not synchronous, but usually takes only a few seconds to complete. In some situations, it can take several minutes. The quorum candidate holds a copy of important system metadata. When using an MDisk as a quorum disk, space is allocated from the storage pool. The number of extents required for a quorum depends on the extent size for the managed disk group containing the MDisk.



```
IBM_2145:Cluster_IP_Quorum:superuser>lsquorum
quorum_index status id name   controller_id controller_name active object_type override site_id site_name
0            online 3  mdisk3 0             controller0     no     mdisk       no
1            online 2  mdisk2 0             controller0     no     mdisk       no
2            online 4  mdisk4 0             controller0     yes    mdisk       no
IBM_2145:Cluster_IP_Quorum:superuser>chquorum -mdisk mdisk5 1
IBM_2145:Cluster_IP_Quorum:superuser>lsquorum
quorum_index status id name   controller_id controller_name active object_type override site_id site_name
0            online 3  mdisk3 0             controller0     no     mdisk       no
1            online 5  mdisk5 0             controller0     no     mdisk       no
2            online 4  mdisk4 0             controller0     yes    mdisk       no
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 3: A concise invocation example of chquorum command

What is new in IBM SVC 7.6.0

With respect to quorum disk, IP quorum is a main feature released in IBM SVC 7.6.0 release. IP based quorum support can enable the use of a low cost IP network-attached host as a quorum disk for simplified implementation and operation.

IBM SVC currently uses storage from MDisks or drives for quorum resolution. As stated earlier, normally three MDisks are either automatically or manually selected to be the quorum disk candidates. Storage system exclusively reserves some area in each of these disks to store quorum data. In the case of a quorum loss, that is split-brain scenario, where half of the cluster cannot see the other half, the quorum disks are used to break the tie. The first half in a split-brain scenario, to reach the quorum disks, assumes ownership of the cluster and locks the disks. All the nodes in a cluster must have access to the quorum disks.

In this implementation design, no extra hardware or networking is required beyond what is normally provisioned within a cluster [that is, Fibre Chanel (FC) or serial-attached SCSI (SAS) attached storage]. But in case of a stretched cluster environment, the need for accessibility to the quorum device in the case of a site failure necessitates the presence of a third independent domain for quorum resolution. In versions before 7.6.0, the requirement of a third site is necessary over FC connectivity to the third site with some associated storage controller at the third site. Maintaining this third site and storage controller over FC makes the system costly in case of site recovery implementation of IBM SVC.

To overcome this limitation of maintaining a third site with FC connectivity along with a site 3 controller, the alternative is to implement Ethernet-attached quorum servers at the third site that can be run on hosts. Ethernet connectivity is generally easier and more economical to provide than FC connectivity, and hosts are typically less expensive than fully fledged network-attached storage controllers. This implementation of using host application over Ethernet connection can reduce the implementation and maintenance cost.

Real life use cases

This section includes customer scenarios, namely CLI addition in version 7.6.0 along with IP based quorum deployment.

Command line changes in version 7.6.0

In SVC 7.6.0, the mkquorumapp command is introduced to generate the quorum application. Refer to the following syntax of the mkquorumapp command to generate a Java™ application to use for the IP-based quorum.

Optionally, you can use the -ip_6 parameter to specify that the quorum application uses the IPv6 service addresses to connect to nodes.

Invocation examples

The following example creates the file, /dumps/ip_quorum.jar, to use for IP quorum on an IPv4 network.
IBM_2145:Cluster_IP_Quorum:superuser> mkquorumapp

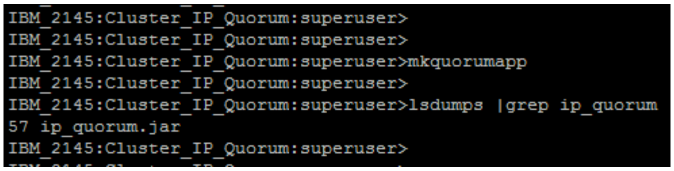
The following example creates the file, /dumps/ip_quorum.jar, to use for IP quorum on an IPv6 network.
IBM_2145:Cluster_IP_Quorum:superuser> mkquorumapp -ip_6

Deployment of IP-based quorum disks

The IP quorum application is a Java application that runs on a host at the third site. The IP network is used for communication between the IP quorum application and node canisters in the system. An IP quorum application can also act as the quorum device for systems that are configured with a standard topology that do not have any external storage configured. This configuration gives a system tie-break capability, automatically resuming I/O processing if half of the system's node canisters are inaccessible. The maximum number of IP quorum applications that can be deployed is five. Applications can be deployed on multiple hosts to provide redundancy.

Perform the following steps to configure IP quorum.

1. Enter the mkquorumapp command to provide a signed JAR file to the dumps directory on the config node. This command provides an error message if the service IP on node is not set.

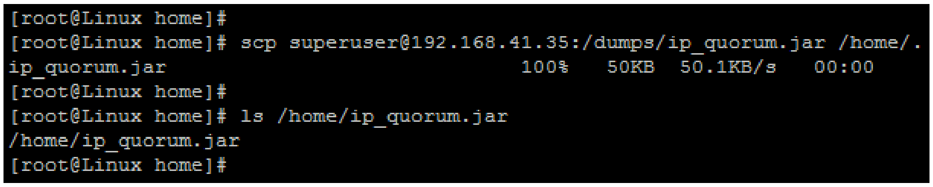


```
IBM_2145:Cluster_IP_Quorum:superuser>
IBM_2145:Cluster_IP_Quorum:superuser>
IBM_2145:Cluster_IP_Quorum:superuser>mkquorumapp
IBM_2145:Cluster_IP_Quorum:superuser>
IBM_2145:Cluster_IP_Quorum:superuser>lsdumps |grep ip_quorum
57 ip_quorum.jar
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 4: A sample output of the mkquorumapp command

2. Download the JAR file generated by the mkquorumapp command to the supported Linux® server where IBM Java 7.1, IBM Java 8 is already being installed.

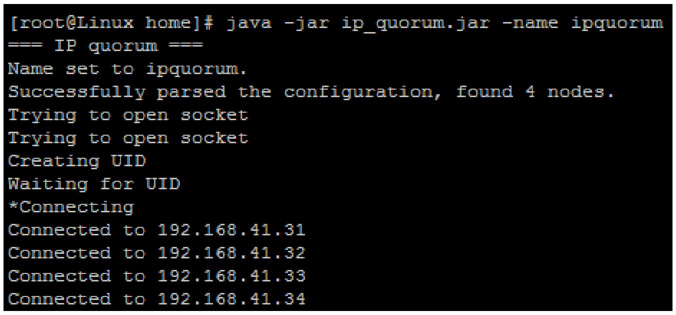
   You can download this JAR file using the management GUI or using Secure Copy Protocol (SCP) when the cluster is running. The JAR file contains the application and all additional files needed for the quorum application to work with the SVC cluster. The JAR file must be regenerated when the user changes a service IP, when the cluster is assigned a new Secure Sockets Layer (SSL) certificate, or when a node is permanently added or removed from the cluster. Only one JAR file needs to be generated, and can then be deployed on multiple hosts.



```
[root@Linux home]#
[root@Linux home]# scp superuser@192.168.41.35:/dumps/ip_quorum.jar /home/.
ip_quorum.jar                                  100%    50KB  50.1KB/s   00:00
[root@Linux home]#
[root@Linux home]# ls /home/ip_quorum.jar
/home/ip_quorum.jar
[root@Linux home]#
```

Figure 5: Sample output of SCP of ip_quorum jar file to Linux server

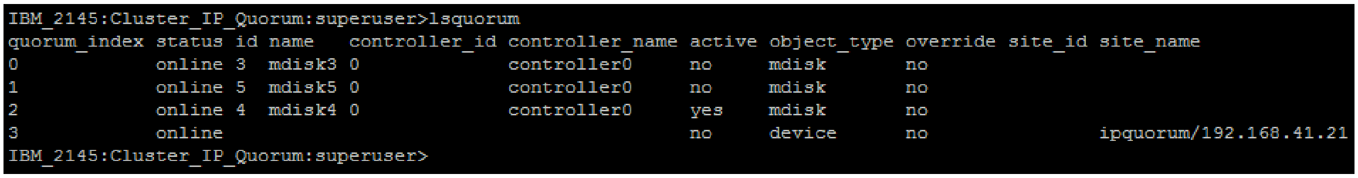
3. Initialize the quorum application on the host by running the quorum ip_quorum.jar file. The app creates a log file in the current directory, which can be used to assist in debugging issues. During usage, you can check the output from the app for any potential errors.



```
[root@Linux home]# java -jar ip_quorum.jar -name ipquorum
=== IP quorum ===
Name set to ipquorum.
Successfully parsed the configuration, found 4 nodes.
Trying to open socket
Trying to open socket
Creating UID
Waiting for UID
*Connecting
Connected to 192.168.41.31
Connected to 192.168.41.32
Connected to 192.168.41.33
Connected to 192.168.41.34
```

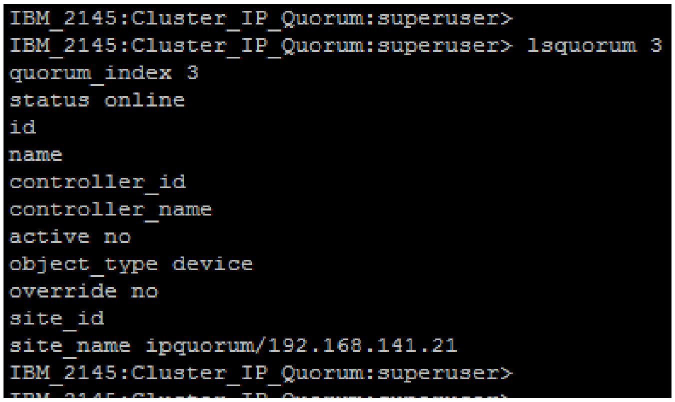
Figure 6: Starting ip_quorum.jar on Linux server

4. Run the lsquorum command (refer Figure 7) on the cluster to view the connected app. The quorum app will automatically be promoted to active the quorum device.



```
IBM_2145:Cluster_IP_Quorum:superuser>lsquorum
quorum_index status id name   controller_id controller_name active object_type override site_id site_name
0            online 3  mdisk3 0             controller0     no     mdisk       no
1            online 5  mdisk5 0             controller0     no     mdisk       no
2            online 4  mdisk4 0             controller0     yes    mdisk       no
3            online                                         no     device      no                ipquorum/192.168.41.21
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 7: lsquorum command output with IP quorum enabled



```
IBM_2145:Cluster_IP_Quorum:superuser>
IBM_2145:Cluster_IP_Quorum:superuser> lsquorum 3
quorum_index 3
status online
id
name
controller_id
controller_name
active no
object_type device
override no
site_id
site_name ipquorum/192.168.141.21
IBM_2145:Cluster_IP_Quorum:superuser>
```

Figure 8: Detailed view of the lsquorum command for IP quorum

The IP quorum setup is now complete and a new quorum disk is visible to the cluster and can be used by the SVC cluster for tie break and other purposes.

Summary

This white paper provided a brief introduction to the IBM System Storage SVC cluster with the IP-based quorum support in the 7.6.0 release. Improvements in the quorum disk area enable to use the IP-based quorum hosted over a Linux server. This feature can be effectively used in a stretched cluster or IBM HyperSwap® topology configurations, or both, where a third site is mandatory to break the tie in case of a split-brain scenario. This IP-based quorum saves the Fibre Channel infrastructure for the third site and can be used over an IP network. This feature is easy to deploy and users can migrate their current FC-based third-site quorum to an IP-based quorum by adding the IP quorum and then removing the older FC quorum at the third site. This feature supports only one JAR file to deploy from multiple Linux servers to have more than one IP quorum available.

Resources

The following websites provide useful references to supplement the information contained in this paper:

IBM Systems on PartnerWorld®
ibm.com/partnerworld/systems

IBM Power Development Platform
ibm.com/partnerworld/pdp

IBM Power Systems Knowledge Center
http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp

IBM Redbooks®
ibm.com/redbooks

About the authors

Kushal S. Patel is a software engineer in the IBM Systems and Storwize team. You can reach Kushal at kushpate@in.ibm.com.

Sarvesh S. Patel is a senior staff software engineer in IBM Systems SVC and Storwize team. You can reach Sarvesh at sarvepat@in.ibm.com.

Shrikant V. Karve is a storage architect in the IBM Systems SVC and Storwize team. You can reach Shrikant at shrkarve@in.ibm.com.

IBM Corporation
IBM Systems and Technology Group
3039 Cornwallis Road
RTP, NC 27709

Produced in the United States of America
April 2016
All Rights Reserved

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.

♲ Please Recycle