# IBM Storage Virtualize for SAN Volume Controller and FlashSystem Family

*Concept Guide - Version 861*

**Note**

Before using this information and the product it supports, read the information in Chapter 1, “Notices,” on page 1.

This edition applies to version 8, release 6, modification x, and to all subsequent modifications until otherwise indicated in new editions.

# Contents

# Chapter 1. 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*
*Armonk, NY 10504-1785*
*U.S.A.*

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 the names and addresses used by an actual business enterprise 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.

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.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel and Intel Xeon are trademarks or registered trademarks 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® and the Linux logo is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo 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.

Other product and service names might be trademarks of IBM or other companies.

# Chapter 2. Scenarios

Scenarios describe how you can achieve specific business goals by using IBM Storage FlashSystem. Each scenario introduces a product feature and walks you through the tasks that are required to configure that feature. The scenarios include useful links to other content to help you to gain a better understanding of the area in which you are interested.

## Policy-based High Availability

Policy-based High Availability (HA) provides a solution for two Storage Systems in different locations where the storage will automatically remain accessible to hosts if there is an event that impacts the infrastructure and makes one of the systems unavailable. Hosts will seamlessly fail over to the other system. The solution ensures that both data and configuration are kept consistent on both systems.

### Planning High Availability

Review the topics in this section to understand what is covered in this scenario, the reasons why a business might want to follow the scenario, and to see an overview of the solution proposed by the scenario.

#### Environment

Each location contains a system that supports Policy-based High Availability (HA) with Fibre Channel network connectivity between them.

The locations are sufficiently far apart such that anticipated disruptions affect only one location. The locations are sufficiently close together to not cause the response time for I/O to exceed the maximum that an application can tolerate because of latency in replication link communications.

Both systems have access to an external IP quorum app to arbitrate if a site or system loss occurs. Multiple quorum applications can be deployed for redundancy.

Hosts in a high-availability solution must use an ALUA-compliant multipath policy. For more information, see Host attachment.

#### Solution overview

The High Availability (HA) solution uses Storage Partitions, a configuration object that is the single point of management for HA objects. These partitions contain volumes, volume groups, hosts, and host-to-volume mappings.

Within a partition:

- All volumes are in volume groups.
- Mappings can be created only between volumes and hosts in the same partition.

An HA replication policy can be associated with a partition. The HA replication policy defines the two systems that are connected by a partnership that are providing the HA solution.

When the partitions are associated with an HA replication policy, all the objects that are contained in the partition are automatically configured across both systems that are associated with the replication policy. Hosts that are created in a partition will be able to discover paths to the mapped volumes through both systems.

Before creation of storage partitions, the systems are configured with “Partnerships” on page 10, Storage pools that are linked between systems (see “Creating Linked Pools” on page 13), “Replication policies” on page 16, and “IP quorum application configuration” on page 14. You can configure multiple partitions for HA.

Each partition that is associated with an HA replication policy has two properties - the preferred management system and the active management system.

The active management system is the system from which all configuration tasks must be performed. If an outage or other failure happens on the current active management system, the active management system will automatically fail over to the other system.

The preferred management system is the system that you would like to be the active management system under ideal conditions. If the active management system and the preferred management system are not the same system, the system will automatically failover the active management system back to the preferred management system when it is able. The preferred management system can be changed by the user.

Everything in a configured Storage Partition is highly available. When required, both management and I/O fail over between systems to maintain access.

You can configure more volumes, volume groups, hosts, and host-to-volume mappings at any time, either by adding to an existing partition or by creating a new one.

The IP Quorum application determines which system becomes the active management system and avoids a scenario where both halves of the High Availability solution continue to process the same partition.

# Implementing High Availability

This guide explains how to configure a Policy-based High Availability (HA) solution. You can setup Policy-based High Availability (HA) solution using the management GUI or the CLI.

## Configuring and monitoring High Availability by using the GUI

High Availability should be configured from the **Storage Partitions** or the **Copy Services** > **Partnerships** panel in the management GUI.

Follow the steps to configure Policy-based High Availability (HA) by using the management GUI:

1. If you do not already have a partnership, refer to .
2. Select a partnership that is ready for use with policy-based replication, and select **Setup policy-based replication**.
3. Choose the High-availability replication type.
4. The management GUI guides you through the required steps to:
   - Configure an IP quorum application.
   - Link storage pools between systems.
   - Create an HA replication policy and a storage partition.
   - Select existing volume groups and mapped hosts to add to the partition.
   - Create new hosts, volume groups, and host-to-volume mappings on both systems.

Use the **Storage Partition Overview** panel to monitor connectivity between the two systems and the IP quorum applications, and the health of the Hosts and Volumes associated with the partition.

## Configuring and monitoring High Availability by using the CLI

It is recommended that Policy-based High Availability (HA) is configured and managed by using the GUI. Configuration can also be performed by using the REST API or the CLI.

Prerequisite to configure your systems to support the High Availability solution:

- If not already configured, follow the link to set up a partnership: “Partnerships” on page 10.
- If not already configured, follow the link to set up a linked storage pool: “Creating Linked Pools” on page 13.
- If not already configured, follow the link to set up an IP quorum application: “IP quorum application configuration” on page 14.

Follow the steps to configure Policy-based High Availability (HA) by using command-line interface:

1. To create a replication policy with the *2-site-ha* topology, enter the following command.

   ```
   mkreplicationpolicy -topology 2-site-ha -location1system <name or ID of one system>
   -location1iogrp 0 -location2system <name or ID of other system> -location2iogrp 0
   ```

   if necessary, use **lspartnership** to get the name or ID of each system. For more information, see **mkreplicationpolicy**.
2. To create a Storage Partition and associate it with the replication policy, enter the following command.

   ```
   # Creating a single-site storage partition that is later configured for HA
   mkpartition -name my_partition
   ...
   chpartition -replicationpolicy <HA policy ID or name> my_partition

   # Creating an HA storage partition
   mkpartition -name ha_partition -replicationpolicy <HA policy ID or name>
   ```

   For more information, see **mkpartition**.
3. To create a logical host object that is maintained across both systems by specifying that the host should be associated with the Storage Partition, enter the following command.

   ```
   mkhost -partition <storage partition ID or name>
   ```

   For more information, see **mkhost**.
4. To create a volume group that is maintained across both systems by specifying that the volume group should be associated with the Storage Partition, enter the following command.

   ```
   mkvolumegroup -partition <storage partition ID or name>
   ```

   For more information, see **mkvolumegroup**.
5. To create a volume in this volume group, enter the following command.

   ```
   mkvolume -volumegroup <volume group ID or name>
   ```

   For more information, see **mkvolume**.
6. To map the created volume to the created host, enter the following command.

   ```
   mkvdiskhostmap -host <host ID or name> <volume ID or name>
   ```

   For more information, see **mkvdiskhostmap**.
7. Hosts can now discover paths to volumes at both locations. Monitor and change the active and preferred management systems that use **lspartition** and **chpartition**. Monitor connectivity between the systems that use **lspartnership**. Monitor the health of the HA solution using **lspartition**.

# Getting started

## Policy-based replication: Asynchronous

*Policy-based replication* uses volume groups and replication policies to automatically deploy and manage replication. Policy-based replication significantly simplifies configuring, managing, and monitoring replication between two systems.

With policy-based replication, you can replicate data between systems with minimal management, significantly higher throughput and reduced latency compared to the remote-copy function. A replication policy has following properties:

- A replication policy can be assigned to one or more volume groups.
- Replication policies cannot be changed after they are created. If changes are required, a new policy can be created and assigned to the associated volume group.
- Each system supports up to a maximum of 32 replication policies.

For more information, refer to Getting started with policy-based replication.

## Policy-based High Availability

Policy-based High Availability (HA) provides a solution for two Storage Systems in different locations where the storage will automatically remain accessible to hosts if there is an event that impacts the infrastructure and makes one of the systems unavailable. Hosts will seamlessly fail over to the other system. The solution ensures that both data and configuration are kept consistent on both systems.

For more information, refer to the “Policy-based High Availability” on page 5 scenario.

## Storage partitions

Storage partitions are used to implement Policy-based High Availability solution. Partitions contain volumes, volume groups, hosts, and host-to-volume mappings.

Within a partition:

- All volumes are in volume groups.
- Mappings can only be created between volumes and hosts in the same partition.

Each partition that is associated with an HA replication policy has two properties - the preferred management system and the active management system.

All configuration actions on a storage partition must be performed on the active management system. The storage partition can be monitored on either system.

The preferred management system is the system that you would like to be the active management system under ideal conditions. In the event of a situation where the active management system and the preferred management system are not the same system, the system will automatically failover the active management system back to the preferred management system when it is able. The preferred management system can be changed by the user.

You can configure additional volumes, volume groups, hosts, and host-to-volume mappings at any time, either by adding to an existing partition or by creating a new one.

**Related concepts**

“Policy-based High Availability” on page 5
Policy-based High Availability (HA) provides a solution for two Storage Systems in different locations where the storage will automatically remain accessible to hosts if there is an event that impacts the

 

infrastructure and makes one of the systems unavailable. Hosts will seamlessly fail over to the other system. The solution ensures that both data and configuration are kept consistent on both systems.

# Partnerships

Partnerships are used to connect systems together to enable migration, data replication, and high-availability solutions.

A system can have partnerships with up to three remote systems. The connectivity for each partnership can be either Fibre Channel or IP. Systems also become indirectly associated with each other through partnerships. If two systems each have a partnership with a third system, those two systems are indirectly associated. A maximum of four systems can be directly or indirectly associated with each other.

A partnership configuration requires actions on both systems involved. This ensures that there is authority to access each system and share data between them.

## Background copy management

Certain types of replication differentiate between foreground host writes and background synchronisation traffic. The background copy rate is specified as a percentage of the partnership link bandwidth that is available to background synchronisation activities. Policy-based replication treats all traffic as background work so the background copy rate should be set to 100% if the system is using only policy-based replication.

HyperSwap, Metro Mirror, and Global Mirror (if supported by your system) use the background copy rate to control synchronisation traffic. Multi-cycling Global Mirror (Global Mirror with Change Volumes) uses only background copy therefore to achieve the best possible recovery point the background copy rate should be set to 100%. If you are using Metro Mirror, HyperSwap or non-cycling Global Mirror on your system, a lower value should be used to ensure that there is sufficient bandwidth to replicate host writes.

## Replication between IBM Storage Virtualize systems

Systems that run IBM Storage Virtualize software are in one of two layers: the replication layer or the storage layer.

- A SAN Volume Controller system is always in the replication layer.
- A FlashSystem is in the storage layer by default, but the system can be configured to be in the replication layer instead.

To create a partnership between systems, both systems must be in the same layer. For more information, including how to change the layer, see System layers.

## Partnership states

The state of the partnership helps determine whether the partnership operates as expected. A partnership can have the following states:

**Configured**
Both the local and remote systems have a partnership that is defined and are running as expected.

**Partial Local**

For the partnership to be fully configured, you must create a partnership from the remote system to the local system.

**Local Stopped**
Indicates that the partnership is defined on both the systems, but the partnership is stopped on the local system.

**Remote Stopped**
Indicates that the partnership is the defined on both the systems, but the partnership is stopped on remote system.

**Partial Local Stopped**
Indicates that only the local system has the partnership that is defined and the partnership is stopped on the local system.

**Local Excluded**
Indicates that both the local system and the remote system have the partnership that is defined, but the local system is excluding the link to the remote system. This state usually occurs when the link between the two systems is compromised by too many errors or slow response times of the partnership.

**Remote Excluded**
Indicates that both the local system and the remote system are defined in a partnership, but the remote system is excluding the link to the local system. This state usually occurs when the link between the two systems is compromised by too many errors or slow response times of the partnership.

**Exceeded**
Indicates that the partnership is unavailable because the network of systems exceeds the number of systems that are allowed in partnerships. To resolve this error, reduce the number of systems that are in partnerships in this network.

**Not Present**
Indicates that the remote system is not visible. This state can be caused by a problem with the connectivity between the local and remote system or if the remote system is unavailable.

For more information, see Creating IP partnership.

# Partnerships using Fibre Channel connectivity

## Environment

See the following pages to understand the Fibre Channel network and port configuration needed to support Fibre Channel partnerships.

- Fibre Channel Zoning
- Planning for more than four fabric ports per node canister
- Long-distance requirements for partnerships

To use partnerships for policy-based replication (asynchronous or high-availability), IP connectivity is required between management IP addresses of partnered systems. The management traffic uses authentication certificates to prevent unauthorized access and ensure secure communications between the systems. Therefore, ensure that valid authentication certificates are installed on both the systems. For more information, see .

## Creating a partnership using Fibre Channel connectivity by using the management GUI

1. Connect to the GUI on either system. Select **Copy Services** > **Partnerships and Remote Copy** and select **Create Partnership**.
2. To create a partnership for policy-based replication (asynchronous or high-availability), ensure that the **Use Policy-Based Replication** checkbox is selected.
3. To create a partnership for Global Mirror or Metro Mirror, nondisruptive system migration, or 3-site partnerships, deselect the **Use Policy-Based Replication** checkbox. The GUI also provides step guidance for configuring 3-site partnerships using 3-Site Orchestrator.
4. If the certificate retrieved from the second system is signed by an authority that is not yet recognized on this system, then from the **Validate certificate** select **Upload File** to upload the root certificate of the certificate authority that signed the partner system's certificate.

### Creating a partnership using Fibre Channel connectivity by using the command-line interface

To create a Fibre Channel partnership between the two systems, use the following command on both systems:

```
mkfcpartnership -linkbandwidthmbits <link_bandwidth_in_mbps> -backgroundcopyrate <percentage>
<remote_system_id | remote_system_name>
```

where **mkfcpartnership** command defines a new partnership created over a Fibre Channel connection. For more information, see **mkfcpartnership**.

## Zoning for partnerships

Inter-system replication (Global Mirror, Metro Mirror, Asynchronous policy-based replication, or Policy-based High Availability (HA)) requires systems to be zoned so they can communicate by using the Fiber Channel network.

### ISL vs FCIP configurations

Where the Fibre Channel fabric spans two physical sites, the inter-site connectivity can be achieved either by using ISLs to connect the sites, or by using Fibre Channel over IP (FCIP) routers to connect the sites by using an IP link.

Where FCIP is used, the configuration must provide guaranteed bandwidth for private (local node to node) traffic either by using separate dedicated links or by implementing QoS.

### Public / Private ports

Best practice is to ensure that cluster ports performing node to node traffic are segregated from those performing inter-system replication, or host I/O. This ensures that if there is an issue with a host, or with inter-system replication, any Fibre Channel credit loss does not impact local system's node to node communication.

### Inter Site Latency

For policy-based High Availability, a maximum of 1 ms RTT (Round-Trip Time) inter site latency is supported.

Both asynchronous policy-based replication and Global Mirror support up to 250 ms RTT with appropriate zoning.

### Policy-based replication (Asynchronous or HA) zoning

Policy-based replication allows a maximum of two I/O groups on the production system to each communicate with an I/O group on the DR system. This is defined by using replication policies.

Zone two ports from each node or canister in the first cluster I/O group with two ports from each canister in the second cluster I/O group. If dual-redundant fabrics are available, zone one port from each node across each fabric to provide the greatest fault tolerance. No other Fibre Channel ports on any node should have remote zones.

### Metro Mirror and Global Mirror (where RTT is less than 80 ms) zoning

For Metro Mirror and Global Mirror configurations where the RTT between systems is less than 80 ms, zone two Fibre Channel ports on each node in the local system to two Fibre Channel ports on each node in the remote system. If dual-redundant fabrics are available, zone one port from each node across each fabric to provide the greatest fault tolerance. No other Fibre Channel ports on any node should have remote zones.

**Optional:** Reducing the number of nodes that are zoned together can reduce the complexity of the intersystem zoning and might reduce the cost of the routing hardware that is required for large installations. Reducing the number of nodes also means that I/O must make extra hops between the nodes in the system, which increases the load on the intermediate nodes and can increase the performance impact, especially for Metro Mirror configurations.

### Global Mirror (where RTT is greater than 80 ms) zoning

If the RTT between systems is greater than 80 ms, stricter configuration requirements apply:

- Use SAN zoning and port masking to ensure that two Fibre Channel ports on each node that is used for replication are dedicated for replication traffic.
- Apply SAN zoning to provide separate intersystem zones for each local-to-remote I/O group pair that is used for replication. See the information about long-distance links for Metro Mirror and Global Mirror partnerships for further details.

**Optional:** As an alternative, choose a subset of nodes in the local system to be zoned to the nodes in the remote system. Minimally, you must ensure that one whole I/O group in the local system has connectivity to one whole I/O group in the remote system. I/O between the nodes in each system is then routed to find a path that is permitted by the configured zoning.

### Host Zoning

For asynchronous policy-based replication, Global Mirror, or Metro Mirror, it is optional to add zoning so the hosts that are visible to the local system can recognize the remote system. This zoning enables a host to examine data in both the local and remote system.

For policy-based High Availability, hosts need to be zoned so they can see the host ports from both systems.

## Creating Linked Pools

With policy-based replication, storage pool links define the pool that is used on the remote system to create the replicated volumes when replication policies exist. Replication policies can only be assigned to volume groups containing volumes in linked storage pools. When linking pools on a stretched topology system, only the pools that are in site 1 are required to be linked.

If the storage pools exist on the production and recovery systems, you can add a link between the pools from either system. If child pools currently exist on a single system only, you can use the management GUI on the partnered system to create and link a child pool in a single step. The management GUI simplifies the process of creating a linked pool on the partnered system. The management GUI automatically displays the properties such as name, capacity, and provisioning policy from the system where the child pools already exist. You can use these values to create the new linked child pool on the partnered system without logging in to the other system.

When you create pool links, you can assign a provisioning policy to the pools that you are linking. Provisioning policies control how capacity is provisioned on all volumes in the linked pool. Pools on each partnered system can be assigned provisioning policies with different capacity savings methods. The system creates two default provisioning policies when the first parent pool is created. You can create more user-defined policies to specify alternative capacity savings. If a provisioning policy is not configured, the system automatically creates fully provisioned volumes.

### Linking pools in topologies with more than two systems

If a pool on one of the systems has existing links to another partnered system, you must add the link from the unlinked system. The existing link between pools for other partnerships is not affected.

### Creating and modifying links between pools

Using the **Pools** panel on the management GUI, you can:

- Create links between existing pools
- Modify links between existing pools
- Create and link child pools, where child pools already exist on one system.

To create linked child pools, use the GUI on the system that does not have the child pools created.

**Note:** In the **Properties** column on the **Pools** page, linked pools are marked with a link icon.

## IP quorum application configuration

The IP quorum application is a Java™ application that runs on a server that is separate from the storage system. A policy-based High Availability partnership requires an IP quorum application to arbitrate in case of a site or system loss. The IP quorum application can be generated on either system in the partnership. The partnership must be created before generating and deploying the IP quorum application.

The maximum number of IP quorum applications that can be deployed on a single system is five. This enables multiple servers to be used to provide redundancy. Only one instance of the IP quorum application per server, per system, is supported. For example, a server can run two IP quorum instances if each instance is connected to a different nonpartnered storage system. Ensure that bandwidth is available to support multiple IP quorum instances.

Do not deploy the IP quorum application on a server that depends on storage that is presented by the system. This action can result in a situation where the nodes need to detect the IP quorum application to process I/O, but cannot because the IP quorum application cannot access storage.

An Ethernet connectivity issue can prevent an IP quorum application from accessing a node that is still online and an event is raised on the system if this occurs.

Use the following support article to understand IP network requirements: https://www.ibm.com/support/pages/node/7013877

**Warning:** For the IP quorum application to be able to connect using TLS 1.3 (SSL Security Levels 6 and 7), the version of Java running the application must also support TLS 1.3.

### Configuring IP Quorum on the storage system

You can configure IP quorum on the storage system by using the GUI or CLI:

**Using the management GUI**
In the management GUI, the Policy-based replication setup procedure includes generating the IP quorum application.

**Using the CLI**
On the CLI, enter the command: **`mkquorumapp -partnersystem <remote system ID or name>`**

### Deploying the IP Quorum Application on the server

Follow these steps to deploy the IP quorum application on the server:

1. Create a separate directory that is dedicated to the IP quorum application.
2. Transfer the IP quorum application from the storage system to the dedicated directory.
3. Use the **`ping`** command on the server to verify that it can establish a connection with the service IP address of each node in both systems.
4. Enter the command **`java -jar ip_quorum.jar`** to initialize the IP quorum application.

**Note:** The IP quorum application needs to be running always.

The options available when running the IP quorum application can be listed with **-help**:

```
1. $ java -jar ip_quorum.jar -help
2. -name (optional) to help identify the IP quorum app instance in the GUI/lsquorum.  Can only
contain 1-20 characters that are A-Z, a-z or 0-9.
3. -debug (optional) run the app in debug mode to display verbose messages on stdout and in the
log file.
4. -emit (optional) display T3 metadata header information
5. -location (optional) default = ip_quorum.log.16845080043810. Specify the full, or relative
path to store your log files. Allowed characters: [a-z A-Z 0-9 .-_/]
6. -rotation (optional) default = 5. Specify the maximum number of log files. Allowed values:
[1-10]
7. -size (optional) default = 5120. Specify the log file size in kb. Allowed values:
[1024-10240]
8. -version / -v (optional) display the vrmf of the system when the app was generated,
timestamp and generation id.
```

For more information on configuring IP quorum on a Linux host, see Helpful resource and publications.

### Reviewing and monitoring the IP Quorum Application

You can monitor and review the IP quorum application from each system by using the management GUI or CLI:

**Using the management GUI**
In the management GUI, select **Settings** > **System** > **IP Quorum**.

**Using the CLI**
Enter the **lsquorum** command. For more information, see **lsquorum**.

The output will display application_type: partnership.

## Configuring quorum

A quorum device is used to break a tie when a SAN fault occurs, when exactly half of the nodes that were previously a member of the system are present. A quorum device is also used to store a backup copy of important system configuration data. Just over 256 MB is reserved for this purpose on each quorum device.

It is possible for a system to split into two groups where each group contains half the original number of nodes in the system. A quorum device determines which group of nodes stops operating and processing I/O requests. In this tie-break situation, the first group of nodes that accesses the quorum device is marked as the owner of the quorum device and as a result continues to operate as the system, handling all I/O requests. If the other group of nodes cannot access the quorum device or finds that the quorum device is owned by another group of nodes, it stops operating as the system and does not handle I/O requests.

A system can have only one active quorum device that is used for a tie-break situation. However, the system uses up to three quorum devices to record a backup of system configuration data to be used in the event of a disaster. The system automatically selects one quorum device to be the active quorum device. The other quorum devices provide redundancy if the active quorum device fails before a system is partitioned. To avoid the possibility of losing all the quorum devices with a single failure, assign quorum disk candidates on multiple storage systems or run IP quorum applications on multiple servers.

### Single site configurations

The normal configuration is to use a managed drive or an MDisk as the quorum device when the system is not configured as a stretched or HyperSwap system. A system automatically assigns quorum disk candidates. When you add new storage to a system or remove existing storage, however, it is a good practice to review the quorum disk assignments. Optionally an IP quorum device can be configured either as an alternative to using quorum disks or to provide additional redundancy.

### Stretched or HyperSwap configurations

To provide protection against failures that affect an entire location, such as a power failure, you can use a configuration that splits a single system across three physical locations.

A stretched or HyperSwap system has system nodes divided between two sites. If a SAN fault causes loss of connectivity between sites or a fault causes a site wide outage then the quorum configuration determines which site continues operating and processing I/O requests. A high availability solution has the active quorum device configured at a third site so that the system will continue to operate after any single-site failure.

Generally, when the nodes in a system are split among sites, configure the system this way:

- Site 1: Half of system nodes + one quorum device
- Site 2: Half of system nodes + one quorum device
- Site 3: Active quorum device

Typically the quorum devices at site 1 and site 2 are quorum disks and the quorum device at site 3 is an IP quorum application. However, the system can be configured to use either quorum disks or IP quorum applications at any site. This configuration ensures that a quorum device is always available, even after a single-site failure.

When you are using an IP quorum application at a third site, you can configure a preference for which site continues operation if there is a loss of connectivity between the two sites. If only one site runs critical applications, you can configure this site as preferred. If a preferred site is configured and a failure causes an outage at the preferred site, the other site wins the tie-break and continues operating and processing I/O requests.

A stretched or HyperSwap system can be configured without a quorum device at a third site. If there is no third site, then quorum must be configured to select a site to always win a tie-break. If there is a loss of connectivity between the sites, then the site that is configured as the winner continues operating and processing I/O requests and the other site stops until the fault is fixed. If there is a site outage at the wining site, then the system stops processing I/O requests until this site is recovered or the manual quorum override procedure is used.

Generally, when the nodes in a system are split between two sites and there is no third site quorum, configure the system this way:

- Site 1: Half of system nodes + one or two quorum devices
- Site 2: Half of system nodes + one quorum device

Typically, the quorum devices at site 1 and site 2 are both quorum disks and are automatically configured by the system. It is possible to configure IP quorum applications as an alternative to using quorum disks. When a winner site has been configured and both sites are operational, there is no active quorum device. The quorum devices at site 1 and site 2 are only used to retain a backup copy of important system configuration data. If a failure results in just the nodes at the winner site continuing operation, then the system automatically selects one of the quorum devices at that site to be the active quorum device to protect against further failures.

## Replication policies

A replication policy defines how replication is configured between systems.

A replication policy defines three key attributes:

**A set of locations**
Defines the I/O groups on the partnered systems that contain a replicated copy of the volume group or storage partition. The location defines *where* data is replicated.

**A topology**
Represents organization of the systems and the type of replication that is completed between each location. The topology defines *how* data is replicated between the locations.

**A recovery point objective (RPO) for asynchronous replication topologies**
Defines a maximum acceptable RPO for the asynchronous replication between locations.

The following rules apply to replication policies:

- An asynchronous replication policy can be assigned to one or more volume groups.
- A high availability replication policy can be assigned to one or more storage partitions.
- A replication policy cannot be assigned to a volume group that is in a storage partition.
- Replication policies cannot be changed after they are created. If changes are required, a new policy can be created and assigned to a volume group.
- Each system supports up to a maximum of 32 replication policies.

### Creating a replication policy

To create a replication policy, use the **Policies** > **Replication Policies** panel in the management GUI. You can also create replication policy by following the **Copy services** > **Partnerships and remote copy** panel in the management GUI and click **Setup policy-based replication**

### Assigning an asynchronous replication policy to a volume group

To assign an asynchronous replication policy to a volume group, use the **Volumes** > **Volume Groups** panel in the management GUI. Select the volume group that needs to be associated with the replication policy, and under the **Policies** tab, click **Assign Replication Policy**.

### Assigning a high availability replication policy to an existing storage partition

To configure a high availability replication policy on a new storage partition, use the **Copy services** > **Partnerships and remote copy** panel in the management GUI and click **Setup policy-based replication**.

To assign a high availability replication policy to an existing storage partition, use the **Storage partitions** panel in the management GUI. Choose a partition that does not already have a replication policy assigned and click **Add High availability replication**.

## Network fabrics

A *fabric* refers collectively to the equipment and configuration that implements a network. A *network fabric* describes the network topology in which components pass data to each other through interconnecting switches.

A network fabric consists of hubs, switches, adapter endpoints, and the connecting cables that support a communication protocol between devices. The system supports both LAN and SAN network fabrics.

## Volume groups

A *volume group* is a container for managing a set of related volumes as a single object. The volume group provides consistency across all volumes in the group.

Volume groups can be used with the following functions:

**Safeguarded Copy function**
One implementation of volume groups is to group volumes to be configured as Safeguarded. Safeguarded copy function is a cyber-resiliency feature that creates immutable copies of data that cannot be changed or manipulated.

A Safeguarded volume group describes a set of source volumes that can span different pools and are backed up collectively with the Safeguarded Copy function. Safeguarded snapshots are supported on the system through an internal scheduler that is defined in the snapshot policy or can be configured with an external snapshot scheduling application such as IBM Copy Services Manager.

**Policy-based High Availability replication**

One implementation of volume groups is to add the volume group to a storage partition that is associated with a high availability replication policy. Volumes in that volume group, hosts, and host-to-volume mappings contained in the partition are automatically configured across both systems associated with the replication policy. For more information, see .

**Asynchronous policy-based replication**

Asynchronous policy-based replication is configured on all volumes in a volume group by assigning an asynchronous replication policy to that volume group. The system automatically replicates the data and configuration for volumes in the group based on the values and settings in the replication policy. As part of asynchronous replication, a recovery volume group is created automatically on the recovery system. Recovery volume groups cannot be created, changed, or deleted. A single replication policy can be assigned to multiple volume groups to simplify replication management. When additional volumes are added to the group, replication is automatically configured for these new volumes.

**Snapshot function**

*Snapshots* are the read only point-in-time copies of a volume group that cannot be directly accessible from the hosts. To access the snapshot contents, you can create a clone or thin clone of a volume group snapshot. You can use the management GUI to configure volume groups to use snapshot policies for multiple volumes for consistent management. Safeguarded snapshot with internal scheduler can be created by using snapshot function.

The volumes in a volume group are supposed to be mutually consistent. This means that volume group only make sense as a group. When a group of thin-clone or clone is populated, it is snapshot function's responsibility to ensure that the images are mutually consistent. When volumes are added or removed from a group, the host applications ensures the volume groups are mutually consistent.

# SANs

A *storage area network (SAN)* is a pool of storage systems that are interconnected to the servers in an enterprise. A SAN administrator is the person responsible for administering the various resources that make up the SAN.

A SAN allows the establishment of direct connections between storage devices and servers. It offers simplified storage management, scalability, flexibility, availability, and improved data access, movement, and backup.

A SAN storage system consists of two to eight system nodes that are arranged in a clustered system. These nodes appear as part of the SAN fabric, along with the host systems, the RAID storage systems, and the storage devices, all connected together to create the SAN. Other devices such as fabric switches might be required to complete the SAN.

There are two types of SAN: redundant and counterpart. A *redundant* SAN consists of a fault tolerant arrangement of two counterpart SANs. A redundant SAN configuration provides two independent paths for each device that is attached to the SAN. A *counterpart* SAN is a non-redundant portion of a redundant SAN and provides all the connectivity of the redundant SAN, but without the redundancy. Each counterpart SAN provides an alternative path for each device that is attached to the SAN.

**Note:** For best availability, use a redundant SAN with the system. Non-redundant SANs are also supported.

## Counterpart SANs

A *counterpart SAN* is a non-redundant portion of a redundant SAN. A counterpart SAN provides all the connectivity of the redundant SAN but without the redundancy. Each counterpart SAN provides an alternate path for each device that is attached to the SAN.

A redundant SAN that consists of more than one counterpart SAN can contain switches from different vendors.

## Redundant SANs

Because of the requirement for high availability, the system is typically installed into a redundant SAN.

Degraded performance can occur when you have a SAN configuration in which any single component might fail and connectivity between the devices within the SAN is maintained. Splitting the SAN into two independent counterpart SANs achieves this normally.