IBM System Storage N series

Data ONTAP 8.1 High Availability and MetroCluster Configuration Guide For 7-Mode

This Release Candidate publication is provided in conjunction with a Data ONTAP 8.1.2 RC code level. Later levels of code and associated publications may be available through the N series support site at www.ibm.com/storage/support/nseries. The N series support site includes the "IBM System Storage N series Data ONTAP Release Model", which describes the different types of Data ONTAP releases.

GA32-1036-03

Contents

Preface

About this guide

This document applies to IBM N series systems running Data ONTAP, including systems with gateway functionality. If the term *7-Mode* is used in the document, it refers to Data ONTAP operating in 7-Mode, which has the same features and functionality found in the prior Data ONTAP 7.1, 7.2, and 7.3 release families.

In this document, the term *gateway* describes IBM N series storage systems that have been ordered with gateway functionality. Gateways support various types of storage, and they are used with third-party disk storage systems—for example, disk storage systems from IBM, HP®, Hitachi Data Systems®, and EMC®. In this case, disk storage for customer data and the RAID controller functionality is provided by the back-end disk storage system. A gateway might also be used with disk storage expansion units specifically designed for the IBM N series models.

The term *filer* describes IBM N series storage systems that either contain internal disk storage or attach to disk storage expansion units specifically designed for the IBM N series storage systems. Filer storage systems do not support using third-party disk storage systems.

Supported features

IBM System Storage N series storage systems are driven by NetApp Data ONTAP software. Some features described in the product software documentation are neither offered nor supported by IBM. Please contact your local IBM representative or reseller for further details.

Information about supported features can also be found on the N series support website (accessed and navigated as described in *Websites* on page 11).

Websites

IBM maintains pages on the World Wide Web where you can get the latest technical information and download device drivers and updates. The following web pages provide N series information:

- A listing of currently available N series products and features can be found at the following web page:
  www.ibm.com/storage/nas/
- The IBM System Storage N series support website requires users to register in order to obtain access to N series support content on the web. To understand how the N series support web content is organized and navigated, and to access the N series support website, refer to the following publicly accessible web page:

www.ibm.com/storage/support/nseries/

This web page also provides links to AutoSupport information as well as other important N series product resources.

- IBM System Storage N series products attach to a variety of servers and operating systems. To determine the latest supported attachments, go to the IBM N series interoperability matrix at the following web page:

  www.ibm.com/systems/storage/network/interophome.html
- For the latest N series hardware product documentation, including planning, installation and setup, and hardware monitoring, service and diagnostics, see the IBM N series Information Center at the following web page:

  publib.boulder.ibm.com/infocenter/nasinfo/nseries/index.jsp

Getting information, help, and service

If you need help, service, or technical assistance or just want more information about IBM products, you will find a wide variety of sources available from IBM to assist you. This section contains information about where to go for additional information about IBM and IBM products, what to do if you experience a problem with your IBM N series product, and whom to call for service, if it is necessary.

Before you call

Before you call, make sure you have taken these steps to try to solve the problem yourself:

- Check all cables to make sure they are connected.
- Check the power switches to make sure the system is turned on.
- Use the troubleshooting information in your system documentation and use the diagnostic tools that come with your system.
- Refer to the N series support website (accessed and navigated as described in *Websites* on page 11) for information on known problems and limitations.

Using the documentation

The latest versions of N series software documentation, including Data ONTAP and other software products, are available on the N series support website (accessed and navigated as described in *Websites* on page 11).

Current N series hardware product documentation is shipped with your hardware product in printed documents or as PDF files on a documentation CD. For the latest N series hardware product documentation PDFs, go to the N series support website.

Hardware documentation, including planning, installation and setup, and hardware monitoring, service, and diagnostics, is also provided in an IBM N series Information Center at the following web page:

publib.boulder.ibm.com/infocenter/nasinfo/nseries/index.jsp

Hardware service and support

You can receive hardware service through IBM Integrated Technology Services. Visit the following web page for support telephone numbers:

www.ibm.com/planetwide/

Firmware updates

IBM N series product firmware is embedded in Data ONTAP. As with all devices, ensure that you run the latest level of firmware. Any firmware updates are posted to the N series support website (accessed and navigated as described in *Websites* on page 11).

Note: If you do not see new firmware updates on the N series support website, you are running the latest level of firmware.

Verify that the latest level of firmware is installed on your machine before contacting IBM for technical support.

How to send your comments

Your feedback helps us to provide the most accurate and high-quality information. If you have comments or suggestions for improving this document, please send them by email to *starpubs@us.ibm.com*.

Be sure to include the following:

- Exact publication title
- Publication form number (for example, GC26-1234-02)
- Page, table, or illustration numbers
- A detailed description of any information that should be changed

HA pair types and requirements

There are four types of High Availability (HA) pairs, each having different advantages and requirements.

Overview of HA pairs

The different types of HA pairs all offer access to storage through two different controllers. Each type has its own benefits and requirements.

What an HA pair is

An HA pair is two storage systems (nodes) whose controllers are connected to each other either directly or, in the case of a fabric-attached MetroCluster, through switches and FC-VI interconnect adapters. In this configuration, one node can take over its partner's storage to provide continued data service if the partner goes down.

You can configure the HA pair so that each node in the pair shares access to a common set of storage, subnets, and tape drives, or each node can own its own distinct set of storage.

The controllers are connected to each other through an HA interconnect. This allows one node to serve data that resides on the disks of its failed partner node. Each node continually monitors its partner, mirroring the data for each other’s nonvolatile memory (NVRAM or NVMEM). The interconnect is internal and requires no external cabling if both controllers are in the same chassis.

Takeover is the process in which a node takes over the storage of its partner. *Giveback* is the process in which that storage is returned to the partner. Both processes can be initiated manually or configured for automatic initiation.

How HA pairs support nondisruptive operations and fault tolerance

HA pairs provide fault tolerance and let you perform nondisruptive operations, including software upgrades and hardware maintenance.

- Fault tolerance

  When one node fails or becomes impaired and a takeover occurs, the partner node continues to serve the failed node’s data.
- Nondisruptive software upgrades or hardware maintenance

  When you halt one node and a takeover occurs (automatically, unless you specify otherwise), the partner node continues to serve data for the halted node while you upgrade or perform maintenance on the node you halted.

The HA pair supplies nondisruptive operation and fault tolerance due to the following aspects of their configuration:

- The controllers in the HA pair are connected to each other either through an HA interconnect consisting of adapters and cable, or, in systems with two controllers in the same chassis, through an internal interconnect.

  The nodes use the interconnect to perform the following tasks:

  - Continually check whether the other node is functioning
  - Mirror log data for each other's NVRAM or NVMEM
  - Synchronize each other's time
- They use two or more disk shelf loops, or third-party storage, in which the following conditions apply:

  - Each node manages its own disks or array LUNs.
  - In case of takeover, the surviving node provides read/write access to the partner's disks or array LUNs, until the failed node becomes available again.

  Note: Disk ownership is established by Data ONTAP or the administrator, rather than by which disk shelf the disk is attached to.
- They own their spare disks, spare array LUNs, or both, and do not share them with the other node.
- They each have mailbox disks or array LUNs on the root volume that do the following tasks:

  - Maintain consistency between the pair
  - Continually check whether the other node is running or whether it has performed a takeover
  - Store configuration information

Related concepts

Where to find procedures for nondisruptive operations with HA pairs on page 207
Managing EXN1000, EXN2000, or EXN4000 unit disk shelves in an HA pair on page 184

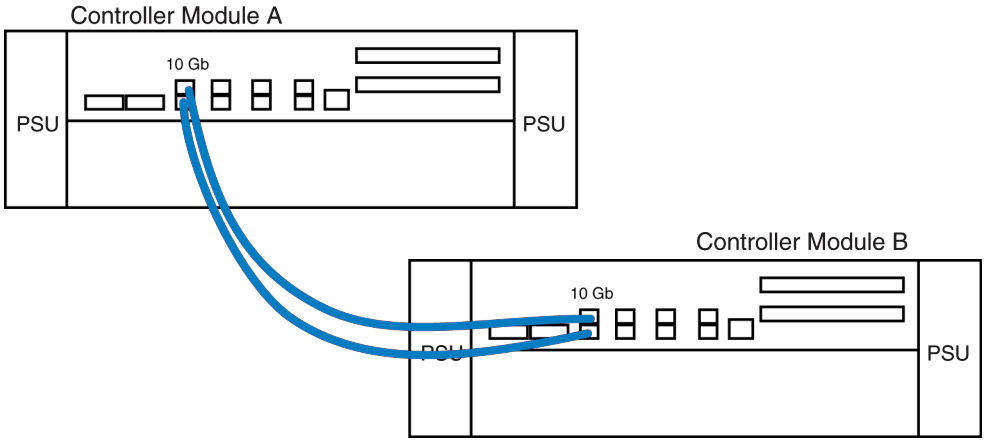
Single-chassis and dual-chassis HA pairs

Depending on the model of the storage system, an HA pair can consist of two controllers in a single chassis, or two controllers in two separate chassis. Some models can be configured either way, while other models can be configured only as a single-chassis HA pair or dual-chassis HA pair.

The following example shows a single-chassis HA pair:

In a single-chassis HA pair, both controllers are in the same chassis. The HA interconnect is provided by the internal backplane. No external HA interconnect cabling is required.

The following example shows a dual-chassis HA pair and the HA interconnect cables:



In a dual-chassis HA pair, the controllers are in separate chassis. The HA interconnect is provided by external cabling.

Interconnect cabling for systems with variable HA configurations

In systems that can be configured either as a single-chassis or dual-chassis HA pair, the interconnect cabling is different depending on the configuration.

The following table describes the interconnect cabling for N6200 series and N7x50T series systems:

If the controller modules in the HA pair are...	The HA interconnect cabling is...
Both in the same chassis	Not required. An internal interconnect is used.
Each in a separate chassis	Required

HA configuration and the HA state PROM value

Some controller modules and chassis automatically record in a PROM whether they are in an HA pair or stand-alone. This record is the *HA state* and must be the same on all components within the stand-alone system or HA pair. The HA state can be manually configured if necessary.

Related tasks

Verifying and setting the HA state on controller modules and chassis on page 143

Table of storage system models and HA configuration differences

The supported storage systems have key differences in their HA configuration, depending on the model.

The following table lists the supported storage systems and their HA configuration differences:

Storage system model	HA configuration (single-chassis, dual-chassis, or either)	Interconnect type (internal InfiniBand, external InfiniBand, or external 10-Gb Ethernet)	Uses HA state PROM value?
N7950T	Single-chassis or dual-chassis	Dual-chassis: External InfiniBand using NVRAM adapter Single-chassis: internal InfiniBand	Yes
N7000 series	Dual-chassis	External InfiniBand using NVRAM adapter	No
N6270	Single-chassis or dual-chassis	Dual-chassis: External 10-Gb Ethernet using onboard ports c0a and c0b These ports are dedicated HA interconnect ports. Regardless of the system configuration, these ports cannot be used for data or other purposes. Single-chassis: Internal InfiniBand	Yes
N6240	Single-chassis or dual-chassis	Dual-chassis: External 10-Gb Ethernet using onboard ports c0a and c0b These ports are dedicated HA interconnect ports. Regardless of the system configuration, these ports cannot be used for data or other purposes. Single-chassis: Internal InfiniBand	Yes

Storage system model	HA configuration (single-chassis, dual-chassis, or either)	Interconnect type (internal InfiniBand, external InfiniBand, or external 10-Gb Ethernet)	Uses HA state PROM value?
N6210	Single-chassis	Internal Infiniband	Yes
N5000 series	Dual-chassis	External InfiniBand using NVRAM adapter	No
N3150, N3220 and N3240	Single-chassis	Internal InfiniBand	Yes
N3400	Single-chassis	Internal InfiniBand	No

Best practices for HA pairs

To ensure that your HA pair is robust and operational, you need to be familiar with configuration best practices.

- Make sure that each power supply unit in the storage system is on a different power grid, so that a single power outage does not affect all power supply units.
- Use interface groups (virtual interfaces) to provide redundancy and improve availability of network communication.
- Follow the documented procedures in the *Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode* when upgrading your HA pair.
- Maintain consistent configuration between the two nodes.
  An inconsistent configuration is often the cause of failover problems.
- Make sure that each node has sufficient resources to adequately support the workload of both nodes during takeover mode.
- If your system supports remote management (through an RLM or Service Processor), make sure that you configure it properly, as described in the *Data ONTAP System Administration Guide for 7-Mode*.
- Follow recommended limits for FlexVol volumes, dense volumes, Snapshot copies, and LUNs to reduce the takeover or giveback time.
  When adding traditional or FlexVol volumes to an HA pair, consider testing the takeover and giveback times to ensure that they fall within your requirements.
- For systems using disks, check for and remove any failed disks, as described in the *Data ONTAP Storage Management Guide for 7-Mode*.
- Multipath HA is required on all HA pairs except for some N3400, N3150, N3220 and N3240 system configurations, which use single-path HA and lack the redundant standby connections.
- To ensure that if takeover becomes disabled you receive prompt notification, configure your system for automatic e-mail notification for the `takeover impossible` EMS messages:
  - `ha.takeoverImpVersion`
  - `ha.takeoverImpLowMem`

- `ha.takeoverImpDegraded`
- `ha.takeoverImpUnsync`
- `ha.takeoverImpIC`
- `ha.takeoverImpHotShelf`
- `ha.takeoverImpNotDef`

- Set the `-cancel-auto-giveback-network-failure` option to `true` if network related failovers are enabled.

Comparison of HA pair types

The different types of HA pairs support different capabilities for data duplication, distance between nodes, and failover.

HA pair type	Data duplication?	Distance between nodes	Failover possible after loss of entire node (including storage)?	Notes
Standard HA pair	No	Up to 500 meters **Note:** SAS configurations are limited to 5 meters between nodes	No	Use this configuration to provide higher availability by protecting against many hardware single-points-of-failure.
Mirrored HA pair	Yes	Up to 500 meters **Note:** SAS configurations are limited to 5 meters between nodes	No	Use this configuration to add increased data protection to the benefits of a standard HA pair.

HA pair type	Data duplication?	Distance between nodes	Failover possible after loss of entire node (including storage)?	Notes
Stretch MetroCluster	Yes	Up to 500 meters (270 meters if operating at 4 Gbps) **Note:** SAS configurations are limited to 5 meters between nodes	Yes	Use this configuration to provide data and hardware duplication to protect against a local disaster (for example, a power outage to one node).
Fabric-attached MetroCluster	Yes	Up to 160 kilometers, depending on switch configuration. See the MetroCluster Compatibility Matrix on the N series support website (accessed and navigated as described in *Websites* on page 11).	Yes	Use this configuration to provide data and hardware duplication to protect against a larger scale disaster, such as the loss of an entire site.

Standard HA pairs

Standard HA pairs provide high availability (HA) by pairing two controllers so that one can serve data for the other in case of controller failure or other unexpected events.

Related references

SPOF analysis for HA pairs on page 208

How Data ONTAP works with standard HA pairs

In a standard HA pair, Data ONTAP functions so that each node monitors the functioning of its partner through a heartbeat signal sent between the nodes. Data from the NVRAM or NVMEM of one node is mirrored by its partner, and each node can take over the partner's disks or array LUNs if the partner fails. Also, the nodes synchronize each other's time.

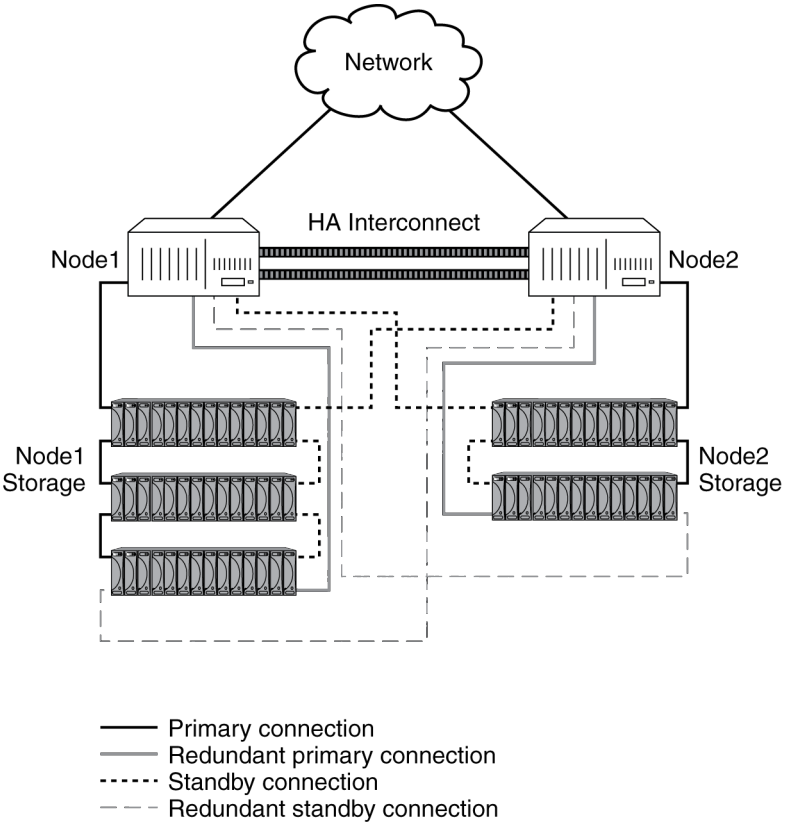
Note: If a node reboots (but a takeover does not occur because the `cf.takeover.on_reboot` option is off or overridden), note that the HA interconnect link comes up prior to Data ONTAP completely loading on the rebooting partner. Commands issued on the surviving controller (that is not rebooting) that check the status of the partner or configuration might indicate that the partner could not be reached. Wait until the partner has fully rebooted and reissue the command.

In some cases (such as the `lun config_check` command) these commands are issued automatically when the interconnect comes up. The resulting error can generate an AutoSupport indicating a configuration problem when in fact the underlying problem is that Data ONTAP has not fully booted.

Standard HA pair diagram

A standard HA pair using native disk storage connects to the data network, has an HA interconnect between the controllers, and has connections to both node's disk shelves.

Standard HA pair



This diagram shows a standard HA pair with native disk shelves and multipath HA.

This diagram shows EXN1000, EXN2000, or EXN4000 unit disk shelves. For information about cabling SAS disk shelves, see the *Universal SAS and ACP Cabling Guide* on the N series support website (accessed and navigated as described in *Websites* on page 11).

Setup requirements and restrictions for standard HA pairs

You must follow certain requirements and restrictions when setting up a new standard HA pair.

The following list specifies the requirements and restrictions to be aware of when setting up a new standard HA pair:

- Architecture compatibility

  Both nodes must have the same system model and be running the same Data ONTAP software and system firmware versions. See the *Data ONTAP Release Notes for 7-Mode* for the list of supported systems.
- Storage capacity

  The number of disks or array LUNs must not exceed the maximum configuration capacity. If your system uses both native disks and third-party storage, the combined total of disks and array LUNs cannot exceed the maximum configuration capacity. In addition, the total storage attached to each node must not exceed the capacity for a single node.

  To determine the maximum capacity for a system using disks, see the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11).

  To determine the maximum capacity for a system using array LUNs or both array LUNs and disks, see the limits information in the *Gateway Limits Reference for Third-Party Storage*.

  Note: After a failover, the takeover node temporarily serves data from all the storage in the HA pair.
- Disks and disk shelf compatibility
  - FC, SATA, and SAS storage are supported in standard HA pairs.

    FC disks cannot be mixed on the same loop as SATA or SAS disks.
  - One node can have only one type of storage and the partner node can have a different type, if needed.
  - Multipath HA is required on all HA pairs except for some N3400, N3150, N3220 and N3240 system configurations, which use single-path HA and lack the redundant standby connections.
- Mailbox disks or array LUNs on the root volume
  - Two disks are required if the root volume is on a disk shelf.
  - One array LUN is required if the root volume is on a storage array.

  The mailbox disks and LUNs are used to do the following tasks:
  - Maintain consistency between the pair
  - Continually check whether the other node is running or whether it has performed a takeover
  - Store configuration information that is not specific to any particular node
- HA interconnect adapters and cables must be installed, unless the system has two controllers in the chassis and an internal interconnect.
- Nodes must be attached to the same network and the Network Interface Cards (NICs) must be configured correctly.
- The same system software, such as SyncMirror, Common Internet File System (CIFS), or Network File System (NFS), must be licensed and enabled on both nodes.

  Note: If a takeover occurs, the takeover node can provide only the functionality for the licenses installed on it. If the takeover node does not have a license that was being used by the partner node to serve data, your HA pair loses functionality after a takeover.

- For an HA pair using third-party storage, both nodes in the pair must be able to see the same array LUNs.

  However, only the node that is the configured owner of a LUN has read and write access to the LUN.

Possible storage configurations in the HA pairs

HA pairs can be configured symmetrically, asymmetrically, as an active/passive pair, or with shared disk shelf stacks.

- Symmetrical configurations

  In a symmetrical HA pair, the nodes each have the same amount of storage.
- Asymmetrical configurations

  In an asymmetrical standard HA pair, one node has more storage than the other. This is supported, as long as neither node exceeds the maximum capacity limit for the node.
- Active/passive configurations

  In this configuration, the passive node has only a root volume, and the active node has all the remaining storage and services all data requests during normal operation. The passive node responds to data requests only if it has taken over the active node.
- Shared loops or stacks

  You can share a loop or stack between the two nodes. This is particularly useful for active/passive configurations, as described in the preceding bullet.

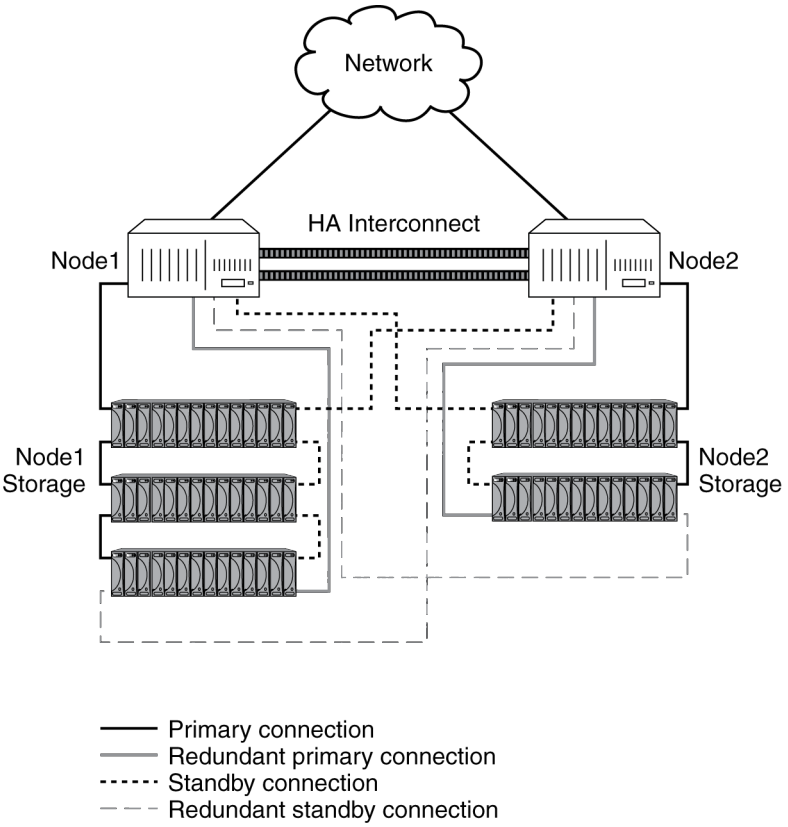
Multipath HA requirement

Multipath HA is required on all HA pairs except for some N3400, N3150, N3220 and N3240 system configurations, which use single-path HA and lack the redundant standby connections. Multipath HA was previously referred to as *Multipath Storage*.

What multipath HA for HA pairs is

Multipath HA provides redundancy for the path from each controller to every disk shelf in the configuration. It is the preferred method for cabling a storage system. An HA pair without multipath HA has only one path from each controller to every disk, but an HA pair with multipath HA has two paths from each controller to each disk, regardless of which node owns the disk.

The following diagram shows the connections between the controllers and the disk shelves for an example HA pair using multipath HA. The redundant primary connections and the redundant standby connections are the additional connections required for multipath HA for HA pairs.



How the connection types are used

A multipath HA configuration uses primary, redundant and standby connections to ensure continued service in the event of the failure of an individual connection.

The following table outlines the connection types used for multipath HA for HA pairs, and how the connections are used.

Connection type	How the connection is used
Primary connection	For normal operation, used to serve data (load-balanced with redundant primary connection).
Redundant primary connection	For normal operation, used to serve data (load-balanced with primary connection).
Standby connection	For normal operation, used for heartbeat information only. After a takeover, assumes role of primary connection.

Connection type	How the connection is used
Redundant standby connection	Not used for normal operation. After a takeover, assumes role of redundant primary connection. If the standby connection is unavailable at takeover time, assumes role of primary connection.

Advantages of multipath HA

Multipath connections in an HA pair reduce single-points-of-failure.

By providing two paths from each controller to every disk shelf, multipath HA provides the following advantages:

- The loss of a disk shelf module, connection, or host bus adapter (HBA) does not require a failover.

  The same storage system can continue to access the data using the redundant path.
- The loss of a single disk shelf module, connection, or HBA does not prevent a successful failover.

  The takeover node can access its partner's disks using the redundant path.
- You can replace modules without having to initiate a failover.

Note: While multipath HA adds value to a stretch MetroCluster environment, it is not necessary in a fabric MetroCluster configuration since multiple paths already exist.

Understanding mirrored HA pairs

Mirrored HA pairs provide high availability through failover, just as standard HA pairs do. Additionally, mirrored HA pairs maintain two complete copies of all mirrored data. These copies are called plexes and are continually and synchronously updated every time Data ONTAP writes to a mirrored aggregate. The plexes can be physically separated to protect against the loss of one set of disks or array LUNs.

Note: Mirrored HA pairs do not provide the capability to fail over to the partner node if one node is completely lost. For example, if power is lost to one entire node, including its storage, you cannot fail over to the partner node. For this capability, use a MetroCluster.

Mirrored HA pairs use SyncMirror. For more information about SyncMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Advantages of mirrored HA pairs

Data mirroring provides additional data protection in the event of disk failures and reduces the need for failover in the event of other component failures.

Mirroring your data protects it from the following problems which would cause data loss without mirroring:

- The failure or loss of two or more disks in a RAID4 aggregate
- The failure or loss of three or more disks in a RAID-DP (RAID double-parity) aggregate
- The failure of an array LUN; for example, because of a double disk failure on the storage array
- The failure of a third-party storage array

The failure of an FC-AL adapter, SAS HBA, disk shelf loop or stack, or disk shelf module does not require a failover in a mirrored HA pair.

Similar to standard HA pairs, if either node in a mirrored HA pair becomes impaired or cannot access its data, the other node can automatically serve the impaired node's data until the problem is corrected.

Setup requirements and restrictions for mirrored HA pairs

The restrictions and requirements for mirrored HA pairs include those for a standard HA pair with these additional requirements for disk pool assignments and cabling.

- You must ensure that your pools are configured correctly:
  - Disks or array LUNs in the same plex must be from the same pool, with those in the opposite plex from the opposite pool.
  - There must be sufficient spares in each pool to account for a disk or array LUN failure.
  - Both plexes of a mirror should not reside on the same disk shelf, as it would result in a single point of failure.

  See the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode* for more information about requirements for setting up SyncMirror with third-party storage.
- You must enable the following licenses on both nodes:
  - cf
  - syncmirror_local
- If you are using third-party storage, paths to an array LUN must be redundant.

Related concepts

Setup requirements and restrictions for standard HA pairs on page 22

Asymmetrically mirrored HA pairs

You can selectively mirror your storage. For example, you could mirror all the storage on one node, but none of the storage on the other node. Takeover will function normally. However, any unmirrored data is lost if the storage that contains it is damaged or destroyed.

Note: You must connect the unmirrored storage to both nodes, just as for mirrored storage. You cannot have storage that is connected to only one node in an HA pair.

Understanding stretch MetroCluster configurations

Stretch MetroCluster configurations provide data mirroring and the additional ability to initiate a failover if an entire site becomes lost or unavailable.

The stretch MetroCluster configuration employs SyncMirror to build a system that can continue to serve data even after complete loss of one of the sites. Data consistency is retained, even when the data is contained in more than one aggregate.

Like mirrored HA pairs, stretch MetroCluster configurations contain two complete copies of the specified data volumes or file systems that you indicated as being mirrored volumes or file systems in your HA pair. These copies are called *plexes* and are continually and synchronously updated every time Data ONTAP writes data to the disks. Plexes are physically separated from each other across different groupings of disks or array LUNs.

Note: You can have both mirrored and unmirrored volumes in a stretch MetroCluster configuration. However, stretch MetroCluster configurations can preserve data only if volumes are mirrored. Unmirrored volumes are lost if the storage where they reside is destroyed.

Unlike mirrored HA pairs, stretch MetroCluster configurations provide the capability to force a failover when an entire node (including the controllers and storage) is destroyed or unavailable.

See the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode* for detailed information about using SyncMirror to mirror data.

Advantages of stretch MetroCluster configurations

MetroCluster configurations provide the same advantages of mirroring as mirrored HA pairs, with the additional ability to initiate failover if an entire site becomes lost or unavailable.

For MetroCluster configuration on filers, the advantages of a stretch MetroCluster configuration are as follows:

- Your data is protected if there is a failure or loss of two or more disks in a RAID 4 aggregate or three or more disks in a RAID-DP aggregate.
- The failure of an FC-AL adapter, loop, or IOM module does not require a failover.

For MetroCluster configurations on gateways, the advantages of a stretch MetroCluster configuration are as follows:

- For RAID 0 aggregates, your data is protected if there is a failure of a storage array or loss of an array LUN on one of the storage arrays in the MetroCluster configuration.

In addition, a MetroCluster configuration provides the `cf forcetakeover -d` command, giving you a single command to initiate a failover if an entire site becomes lost or unavailable. If a disaster occurs at one of the node locations and destroys your data there, your data not only survives on the other node, but can be served by that node while you address the issue or rebuild the configuration.

Related concepts

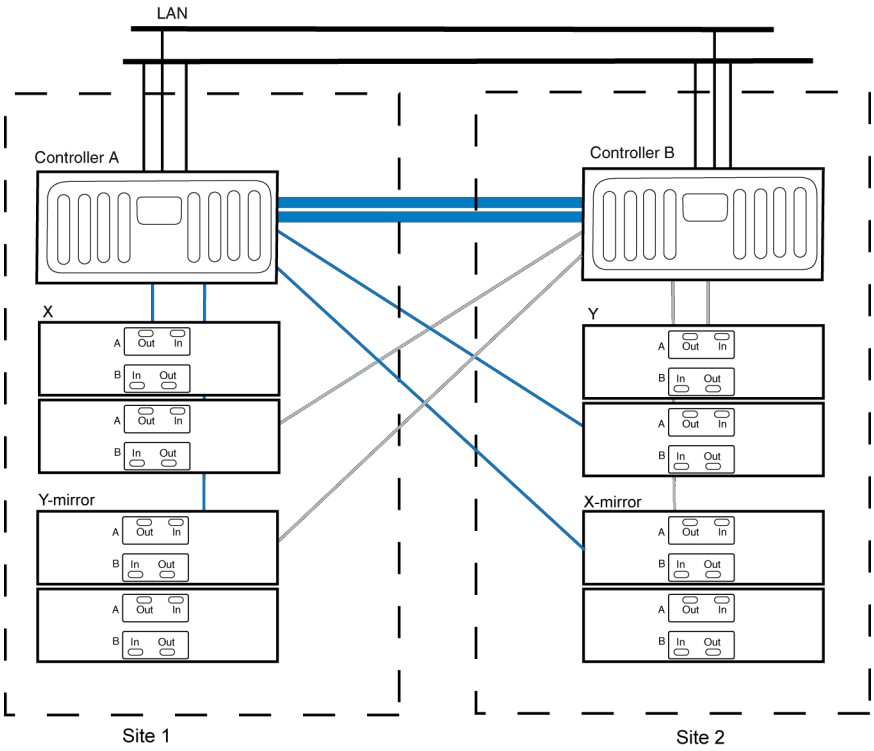
Disaster recovery using MetroCluster configurations on page 196

How filers are connected in stretch MetroCluster configurations

You can configure a stretch MetroCluster configuration so that each controller can access its own storage and its partner's storage, with local storage mirrored at the partner site.

The following figure illustrates the stretch MetroCluster configuration with filers using EXN1000 or EXN2000 unit disk shelves. For an example of a stretch MetroCluster configuration using SAS disk shelves and FibreBridge 6500N bridge, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11). The configuration includes the following connections:

- Connections from each controller to the user network
- The MetroCluster interconnect between the two controllers
- Connections from each controller to its own storage:
  - Controller A to X
  - Controller B to Y
- Connections from each controller to its partner's storage:
  - Controller A to Y
  - Controller B to X
- Connections from each controller to the mirrors of its storage:
  - Controller A to X-mirror
  - Controller B to Y-mirror



Note: This is a simplified figure that does not show disk shelf-to-disk shelf connections.

Related information

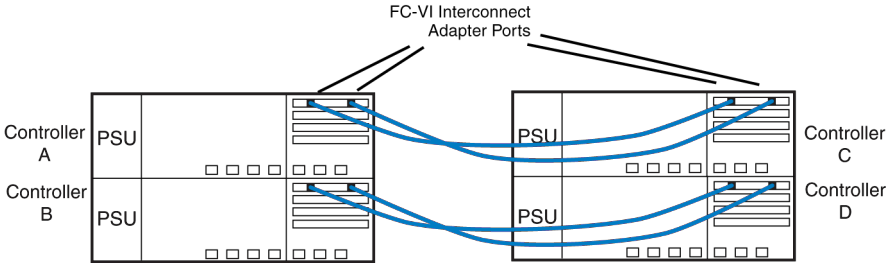
IBM N series support website: www.ibm.com/storage/support/nseries

Stretch MetroCluster configuration on single-enclosure HA pairs

You can configure two stretch MetroCluster configurations between a pair of single-enclosure HA pair systems. In this configuration, the HA pair between the two controllers in each chassis is deactivated, and two separate, side-by-side stretch MetroCluster configurations are formed between the four controllers.

To implement the stretch MetroCluster configuration, you must install an FC-VI adapter in each controller to provide the HA interconnect between the systems. When the FC-VI adapter is installed in the system, the internal InfiniBand interconnect is automatically disabled. This is different from other stretch MetroCluster configurations, which use NVRAM adapters to provide the interconnect.

The following figure shows a stretch MetroCluster configuration on single-enclosure HA pair systems.



Note: The dual-controller N3300 and N3600 systems do not support MetroCluster configurations.

How Data ONTAP works with stretch MetroCluster configurations

Data ONTAP divides storage across physically separated pools of disks.

During configuration, Data ONTAP identifies spare disks and divides them into separate groupings called pools. These pools of disks are physically separated from each other, allowing for high availability of mirrored volumes. When you add a mirrored volume or add disks to one side of a mirrored volume, Data ONTAP determines how much storage you need for the second half of the mirror, and dedicates that storage from a separate pool to the mirrored volume.

Data ONTAP can also be configured to read from both plexes, which in many cases improves read performance.

Note: You can determine which side of the mirrored volume (also called a plex) is read when a data request is received using the raid.mirror_read_plex_pref option. For more information, see the na_options(1) man page.

Configuration variations for stretch MetroCluster configurations

Stretch MetroCluster configurations have asymmetrical and active/passive variations.

The following list describes some common configuration variations that are supported for stretch MetroCluster configurations:

- Asymmetrical mirroring
  You can add storage to one or both nodes that is not mirrored by the other node.

  Attention: Any data contained in the unmirrored storage could be lost if that site experiences a disaster.

  Note: Multiple disk failures in an unmirrored aggregate (three or more disk failures in a RAID-DP aggregate, two or more disk failures in a RAID4 aggregate) or failure of a single array LUN in a RAID 0 aggregate cause the node to panic, resulting in a temporary data service outage while the node reboots, a takeover occurs, or disaster recovery is performed.

  You must mirror the root volumes to enable successful takeover.

Note: You must connect the unmirrored storage to both nodes, just as for mirrored storage. You cannot have storage that is connected to only one node in an HA pair.

- Active/passive MetroCluster configurations
  In this configuration, the remote (passive) node does not serve data unless it has taken over for the local (active) node. Mirroring the passive node's root volume is optional. However, both nodes must have all licenses for a MetroCluster configuration installed so that remote takeover is possible.

Understanding fabric-attached MetroCluster configurations

Like mirrored HA pairs, fabric-attached MetroCluster configurations contain two complete, separate copies of the data volumes or file systems that you configured as mirrored volumes or file systems in your HA pair. The fabric-attached MetroCluster nodes can be physically distant from each other, beyond the distance limit of a stretch MetroCluster configuration.

How fabric-attached MetroCluster configurations use Brocade and Cisco Fibre Channel switches

A MetroCluster configuration for distances greater than 500 meters connects the two nodes by using four Brocade or Cisco Fibre Channel switches in a dual-fabric configuration for redundancy.

Each site has two Fibre Channel switches, each of which is connected through an inter-switch link to a partner switch at the other site. The inter-switch links are fiber optic connections that provide a greater distance between nodes than other HA pairs. For more information about the switches, see the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* for Brocade switches and *Fabric-attached MetroCluster Cisco Switch Configuration Guide* for Cisco switches, available at the N series support website (accessed and navigated as described in *Websites* on page 11).

Each local switch combines with a partner switch to form a fabric. By using four switches instead of two, redundancy is provided to avoid single-points-of-failure in the switches and their connections.

Like a stretch MetroCluster configuration, a fabric-attached MetroCluster configuration employs SyncMirror to build a system that can continue to serve data even after complete loss of one of the nodes and the storage at that site. Data consistency is retained, even when the data is contained in more than one aggregate.

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Advantages of fabric-attached MetroCluster configurations

Fabric-attached MetroCluster configurations provide the same advantages of stretch MetroCluster configurations, while also enabling the physical nodes to be physically distant from each other.

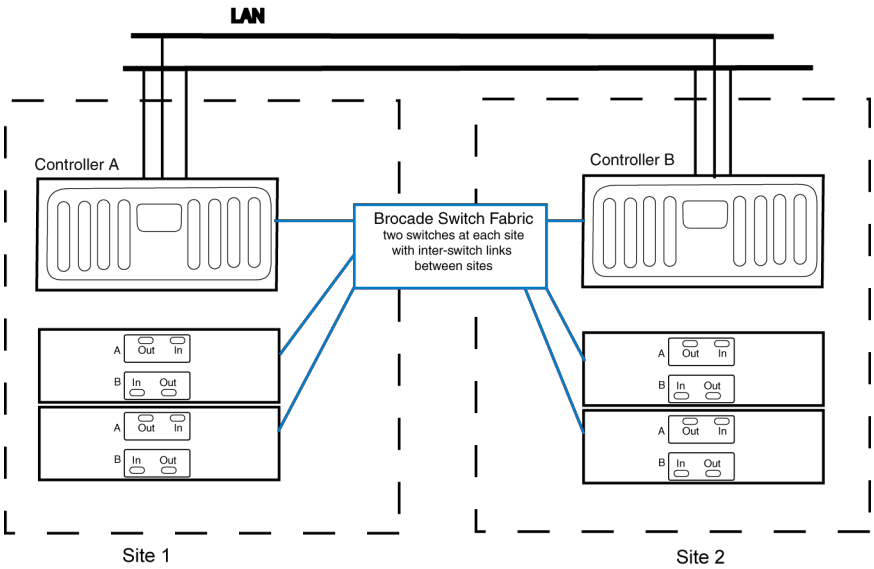
The advantages of a fabric-attached MetroCluster configuration over a stretch MetroCluster configuration include the following:

- The two halves of the configuration can be more than 500 meters apart, which provides increased disaster protection.
- For fabric-attached MetroCluster configurations with filers, disk shelves and nodes are not connected directly to each other, but are connected to a fabric with multiple data routes, ensuring no single point of failure.
- For fabric-attached MetroCluster configurations with gateways, the storage arrays and nodes are not connected directly to each other, but are connected to a fabric with multiple data routes, ensuring no single point of failure.

Fabric-attached MetroCluster configuration with filers

A fabric-attached MetroCluster configuration includes two Brocade or Cisco Fibre Channel switch fabrics that provide long distance connectivity between the nodes. Through the switches, each controller can access its own storage and its partner's storage, with local storage mirrored at the partner site.

The following figure illustrates the fabric-attached MetroCluster configuration with filers.



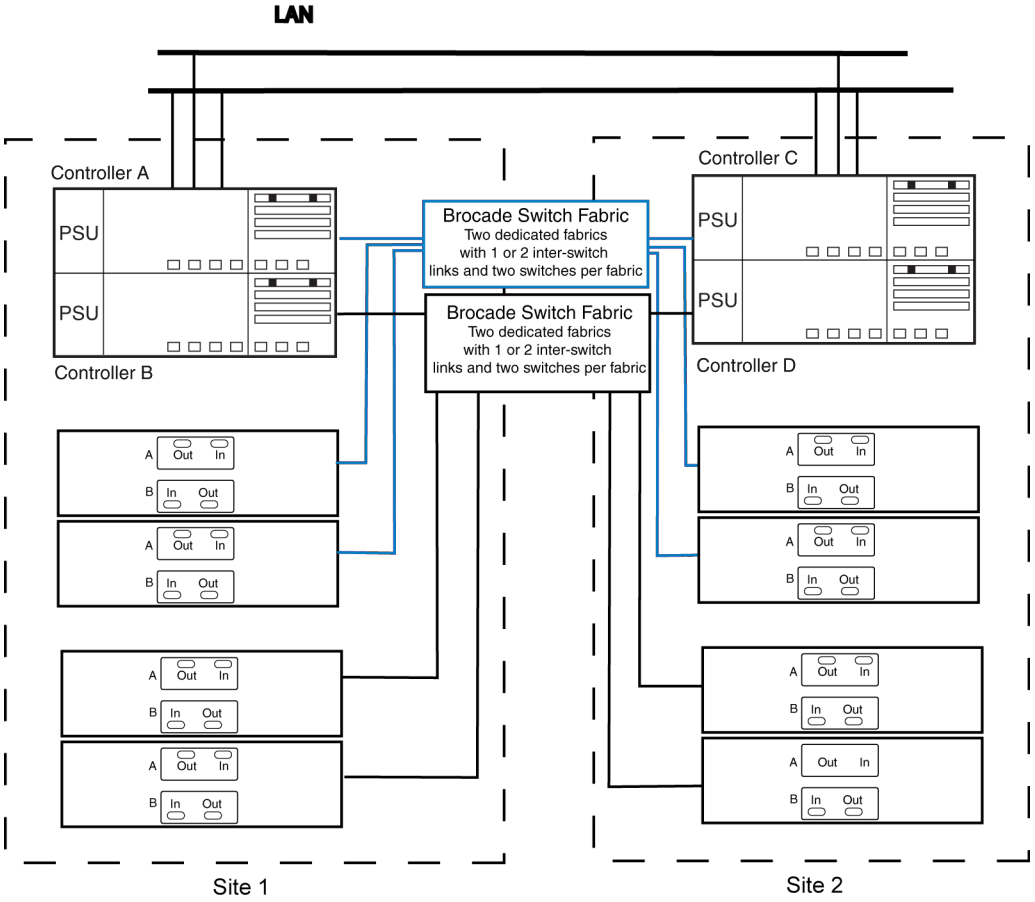
Note: This is a simplified figure that does not show disk shelf-to-disk shelf connections.

Fabric-attached MetroCluster configuration on single-enclosure HA pairs

You can configure a fabric-attached MetroCluster configuration between a pair of single-enclosure HA pair systems. In this configuration, the HA pair between the two controllers in each chassis is deactivated, and two separate, side-by-side MetroCluster configurations are formed between the four controllers.

When the system detects the presence of an FC-VI adapter, which connects the controller to the switch fabric, the internal InfiniBand connection is automatically deactivated.

The following figure shows a fabric-attached MetroCluster configuration on single-enclosure HA pair systems.



Note: This is a simplified figure that does not show disk shelf-to-disk shelf connections.

How Data ONTAP works with fabric-attached MetroCluster configurations

Data ONTAP functions the same way on a fabric-attached MetroCluster configuration as on a stretch MetroCluster configuration.

Related concepts

How Data ONTAP works with stretch MetroCluster configurations on page 31

Configuration limitations for fabric-attached MetroCluster configurations with filers

You must be aware of certain limitations when setting up a new fabric-attached MetroCluster configuration.

The fabric-attached MetroCluster configuration has the following limitations:

- EXN1000, EXN2000, or EXN4000 unit disk shelves using SATA and AT-FCX storage is not supported.
- You cannot use the switches of the MetroCluster configuration to connect Fibre Channel tape devices, or for FC traffic of any kind; you can connect only system controllers and disk shelves to those switches.

  The switches of the MetroCluster configuration either connect EXN1000, EXN2000, or EXN4000 unit disk shelves to the controllers or FibreBridge 6500N bridge (connected to SAS or SATA disk shelves) to the disk shelves.
- You can connect a tape storage area network (SAN) to either of the nodes, but the tape SAN must not use the switches used in a MetroCluster configuration.

Configuration variations for fabric-attached MetroCluster configurations

Fabric-attached MetroCluster configurations support asymmetrical and active/passive configurations.

The following list describes some common configuration variations that are supported for fabric-attached MetroCluster configurations:

- Asymmetrical mirroring

  You can add storage to one or both nodes that is not mirrored by the other node. However, any data contained in the unmirrored storage could be lost if that site experiences a disaster.

  Attention: Multiple disk failures in an unmirrored aggregate (three or more disk failures in a RAID-DP aggregate, two or more disk failures in a RAID4 aggregate) will cause the node to panic, resulting in a temporary data service outage while the node reboots or disaster recovery is performed.

  You must mirror the root volumes to enable successful takeover.

  Note: You must connect the unmirrored storage to both nodes, just as for mirrored storage. You cannot have storage that is connected to only one node in an HA pair.
- Active/passive MetroCluster configurations

  In this configuration, the remote (passive) node does not serve data unless it has taken over for the local (active) node. Mirroring the passive node's root volume is optional. However, both nodes must have all licenses for MetroCluster configuration installed so that remote takeover is possible.

Installing and cabling an HA pair

To install and cable a new standard or mirrored HA pair, you must have the correct tools and equipment and you must connect the controllers to the disk shelves (for filers or gateways using native disk shelves). If it is a dual-chassis HA pair, you must also cable the HA interconnect between the nodes. HA pairs can be installed in either IBM system cabinets or in equipment racks.

The specific procedure you use depends on the following aspects of your configuration:

- Whether you have a standard or mirrored HA pair
- Whether you are using FC or SAS disk shelves

  Note: If your configuration includes SAS Storage Expansion Units, see the *Universal SAS and ACP Cabling Guide* on the N series support website (accessed and navigated as described in *Websites* on page 11) for information about cabling. For cabling the HA interconnect between the nodes, use the procedures in this guide.

Multipath HA is required on all HA pairs except for some N3400, N3150, N3220 and N3240 system configurations, which use single-path HA and lack the redundant standby connections.

System cabinet or equipment rack installation

You need to install your HA pair in one or more IBM system cabinets or in standard telco equipment racks. Each of these options has different requirements.

HA pairs in an equipment rack

Depending on the amount of storage you ordered, you need to install the equipment in one or more telco-style equipment racks.

The equipment racks can hold one or two nodes on the bottom and eight or more disk shelves. For information about how to install the disk shelves and nodes into the equipment racks, see the appropriate documentation that came with your equipment.

HA pairs in a system cabinet

Depending on the number of disk shelves, the HA pair you ordered arrives in a single system cabinet or multiple system cabinets.

The number of system cabinets you receive depends on how much storage you ordered. All internal adapters, such as networking adapters, Fibre Channel adapters, and other adapters, arrive preinstalled in the nodes.

If it comes in a single system cabinet, both the Channel A and Channel B disk shelves are cabled, and the HA adapters are also pre-cabled.

If the HA pair you ordered has more than one cabinet, you must complete the cabling by cabling the local node to the partner node's disk shelves and the partner node to the local node's disk shelves. You must also cable the nodes together by cabling the NVRAM HA interconnects. If the HA pair uses switches, you must install the switches, as described in the accompanying switch documentation. The system cabinets might also need to be connected to each other. See your *System Cabinet Guide* for information about connecting your system cabinets together.

Required documentation

Installation of an HA pair requires the correct documentation.

The following table lists and briefly describes the documentation you might need to refer to when preparing a new HA pair, or converting two stand-alone systems into an HA pair.

Manual name	**Description**
N series Introduction and Planning Guide	This guide describes the physical requirements your site must meet to install IBM N series equipment.
The appropriate system cabinet guide	This guide describes how to install IBM N series equipment into a system cabinet.
The appropriate disk shelf guide	These guides describe how to cable a disk shelf to a storage system.
The appropriate hardware documentation for your storage system model	These guides describe how to install the storage system, connect it to a network, and bring it up for the first time.
Diagnostics Guide	This guide describes the diagnostics tests that you can run on the storage system.
Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode	This guide describes how to upgrade storage system and disk firmware, and how to upgrade storage system software.
Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode	This guide describes, among other topics, SyncMirror technology, which is used for mirrored HA pairs.
Data ONTAP System Administration Guide for 7-Mode	This guide describes general storage system administration.
Data ONTAP Software Setup Guide for 7-Mode	This guide describes how to configure the software of a new storage system for the first time.

Note: If you are installing a gateway HA pair with third-party storage, see the *Gateway Installation Requirements and Reference Guide* for information about cabling gateways to storage arrays, and see the *Gateway Implementation Guide for Third-Party Storage* for information about configuring storage arrays to work with gateways.

Required tools

Installation of an HA pair requires the correct tools.

The following list specifies the tools you need to install the HA pair:

- #1 and #2 Phillips screwdrivers
- Hand level
- Marker

Required equipment

When you receive your HA pair, you should receive the equipment listed in the following table. See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites*) to confirm your storage-system type, storage capacity, and so on.

Required equipment	Details
Storage system	Two of the same type of storage systems
Storage	See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11)
HA interconnect adapter (for controller modules that do not share a chassis) **Note:** When N6200 series systems are in a dual-chassis HA pair, the c0a and c0b 10-GbE ports are the HA interconnect ports. They do not require an HA interconnect adapter. Regardless of configuration, the N6200 series system's c0a and c0b ports cannot be used for data. They are only for the HA interconnect.	InfiniBand (IB) HA adapter (The NVRAM adapter functions as the HA interconnect adapter on N5000 series and later storage systems, except the N6200 series systems)
For EXN1000 or EXN2000 unit disk shelves: FC-AL or FC HBA (FC HBA for Disk) adapters For SAS disk shelves: SAS HBAs, if applicable	Minimum of two FC-AL adapters or two SAS HBAs
Fibre Channel switches	N/A

Required equipment	Details
SFP (Small Form Pluggable) modules	N/A
NVRAM HA adapter media converter	Only if using fiber cabling
Cables (provided with shipment unless otherwise noted)	• One optical controller-to-disk shelf cable per loop • Multiple disk shelf-to-disk shelf cables • Two 4xIB copper cables, or two 4xIB optical cables **Note:** You must purchase longer optical cables separately for cabling distances greater than 30 meters. • Two optical cables with media converters for systems using the IB HA adapter • The N6200 series systems, when in a dual-chassis HA pair, require 10 GbE cables (Twinax or SR) for the HA interconnect.

Preparing your equipment

You must install your nodes in your system cabinets or equipment racks, depending on your installation type.

Installing the nodes in equipment racks

Before you cable your nodes together, you install the nodes and disk shelves in the equipment rack, label the disk shelves, and connect the nodes to the network.

Steps

1. Install the nodes in the equipment rack, as described in the guide for your disk shelf, hardware documentation, or Quick Start guide that came with your equipment.
2. Install the disk shelves in the equipment rack, as described in the appropriate disk shelf guide.
3. Label the interfaces, where appropriate.
4. Connect the nodes to the network, as described in the setup instructions for your system.

Result

The nodes are now in place and connected to the network and power is available.

After you finish

Proceed to cable the HA pair.

Installing the nodes in a system cabinet

Before you cable your nodes together, you must install the system cabinet, nodes, and any disk shelves, and connect the nodes to the network. If you have two cabinets, the cabinets must be connected together.

Steps

1. Install the system cabinets, nodes, and disk shelves as described in the *System Cabinet Guide*.

   If you have multiple system cabinets, remove the front and rear doors and any side panels that need to be removed, and connect the system cabinets together.
2. Connect the nodes to the network, as described in the *Installation and Setup Instructions* for your system.
3. Connect the system cabinets to an appropriate power source and apply power to the cabinets.

Result

The nodes are now in place and connected to the network, and power is available.

After you finish

Proceed to cable the HA pair.

Cabling a standard HA pair

To cable a standard HA pair, you identify the ports you need to use on each node, then you cable the ports, and then you cable the HA interconnect.

About this task

This procedure explains how to cable a configuration using EXN1000, EXN2000, or EXN4000 unit disk shelves.

For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

Note: If you are installing an HA pair between gateway systems with third-party storage, see the *Gateway Installation Requirements and Reference Guide* for information about cabling gateways to storage arrays. See the Gateway implementation guide for your vendor for information about configuring storage arrays to work with gateways.

The sections for cabling the HA interconnect apply to all systems regardless of disk shelf type.

Steps

1. *Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections* on page 41
2. *Cabling Node A to EXN1000, EXN2000, or EXN4000 unit disk shelves* on page 42
3. *Cabling Node B to EXN1000, EXN2000, or EXN4000 unit disk shelves* on page 44
4. *Cabling the HA interconnect (all systems except N6200 series)* on page 46
5. *Cabling the HA interconnect (N6200 series systems in separate chassis)* on page 47

Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections

Before cabling your HA pair, you need to identify which Fibre Channel ports to use to connect your disk shelves to each storage system, and in what order to connect them.

Keep the following guidelines in mind when identifying ports to use:

- Every disk shelf loop in the HA pair requires two ports on the node, one for the primary connection and one for the redundant multipath HA connection.
  A standard HA pair with one loop for each node uses four ports on each node.
- Onboard Fibre Channel ports should be used before using ports on expansion adapters.
- Always use the expansion slots in the order shown in the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) for your platform for an HA pair.
- If using Fibre Channel HBAs, insert the adapters in the same slots on both systems.

See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) to obtain all slot assignment information for the various adapters you use to cable your HA pair.

After identifying the ports, you should have a numbered list of Fibre Channel ports for both nodes, starting with Port 1.

Cabling guidelines for a quad-port Fibre Channel HBA

If using ports on the quad-port, 4-Gb Fibre Channel HBAs, use the procedures in the following sections, with the following additional guidelines:

- Disk shelf loops using ESH4 modules must be cabled to the quad-port HBA first.
- Disk shelf loops using AT-FCX modules must be cabled to dual-port HBA ports or onboard ports before using ports on the quad-port HBA.
- Port A of the HBA must be cabled to the In port of Channel A of the first disk shelf in the loop. Port A of the partner node's HBA must be cabled to the In port of Channel B of the first disk shelf in the loop. This ensures that disk names are the same for both nodes.
- Additional disk shelf loops must be cabled sequentially with the HBA's ports.
  Port A is used for the first loop, port B for the second loop, and so on.
- If available, ports C or D must be used for the redundant multipath HA connection after cabling all remaining disk shelf loops.

- All other cabling rules described in the documentation for the HBA and the *N series Introduction and Planning Guide* must be observed.

Cabling Node A to EXN1000, EXN2000, or EXN4000 unit disk shelves

To cable Node A, you must use the Fibre Channel ports you previously identified and cable the disk shelf loops owned by the node to these ports.

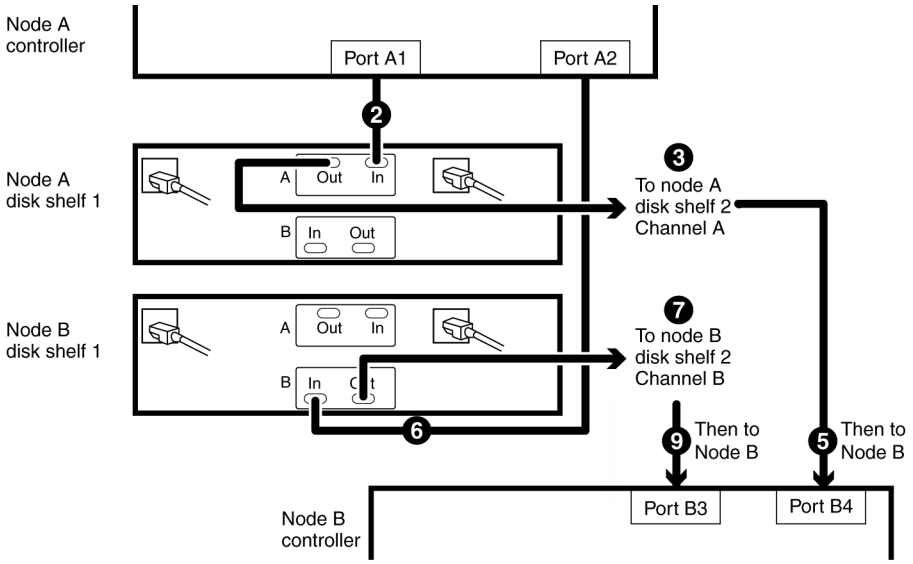
About this task

- This procedure uses multipath HA, which is required on all systems.
- This procedure does not apply to SAS disk shelves.
  For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

Note: You can find additional cabling diagrams in your system's *Installation and Setup Instructions* on the N series support website (accessed and navigated as described in *Websites* on page 11).

Steps

1. Review the cabling diagram before proceeding to the cabling steps.
   - The circled numbers in the diagram correspond to the step numbers in the procedure.
   - The location of the Input and Output ports on the disk shelves vary depending on the disk shelf models.
     Make sure that you refer to the labeling on the disk shelf rather than to the location of the port shown in the diagram.
   - The location of the Fibre Channel ports on the controllers is not representative of any particular storage system model; determine the locations of the ports you are using in your configuration by inspection or by using the *Installation and Setup Instructions* for your model.
   - The port numbers refer to the list of Fibre Channel ports you created.
   - The diagram only shows one loop per node and one disk shelf per loop.
     Your installation might have more loops, more disk shelves, or different numbers of disk shelves between nodes.



2. Cable Fibre Channel port A1 of Node A to the Channel A Input port of the first disk shelf of Node A loop 1.
3. Cable the Node A disk shelf Channel A Output port to the Channel A Input port of the next disk shelf in loop 1.
4. Repeat step 3 for any remaining disk shelves in loop 1.
5. Cable the Channel A Output port of the last disk shelf in the loop to Fibre Channel port B4 of Node B.

   This provides the redundant multipath HA connection for Channel A.
6. Cable Fibre Channel port A2 of Node A to the Channel B Input port of the first disk shelf of Node B loop 1.
7. Cable the Node B disk shelf Channel B Output port to the Channel B Input port of the next disk shelf in loop 1.
8. Repeat step 7 for any remaining disk shelves in loop 1.
9. Cable the Channel B Output port of the last disk shelf in the loop to Fibre Channel port B3 of Node B.

   This provides the redundant multipath HA connection for Channel B.
10. Repeat steps 2 to 9 for each pair of loops in the HA pair, using ports 3 and 4 for the next loop, ports 5 and 6 for the next one, and so on.

Result

Node A is completely cabled.

After you finish

Proceed to cabling Node B.

Cabling Node B to EXN1000, EXN2000, or EXN4000 unit disk shelves

To cable Node B, you must use the Fibre Channel ports you previously identified and cable the disk shelf loops owned by the node to these ports.

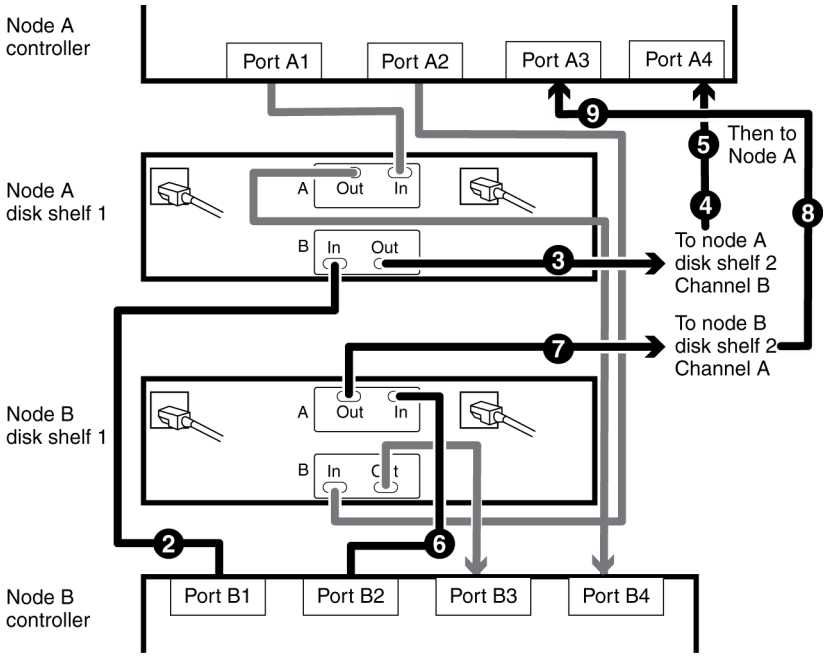
About this task

- This procedure uses multipath HA, required on all systems.
- This procedure does not apply to SAS disk shelves.
  For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

Note: You can find additional cabling diagrams in your system's *Installation and Setup Instructions* on the N series support website (accessed and navigated as described in *Websites* on page 11).

Steps

1. Review the cabling diagram before proceeding to the cabling steps.
   - The circled numbers in the diagram correspond to the step numbers in the procedure.
   - The location of the Input and Output ports on the disk shelves vary depending on the disk shelf models.
     Make sure that you refer to the labeling on the disk shelf rather than to the location of the port shown in the diagram.
   - The location of the Fibre Channel ports on the controllers is not representative of any particular storage system model; determine the locations of the ports you are using in your configuration by inspection or by using the *Installation and Setup Instructions* for your model.
   - The port numbers refer to the list of Fibre Channel ports you created.
   - The diagram only shows one loop per node and one disk shelf per loop.
     Your installation might have more loops, more disk shelves, or different numbers of disk shelves between nodes.



2. Cable Port B1 of Node B to the Channel B Input port of the first disk shelf of Node A loop 1.

   Both channels of this disk shelf are connected to the same port on each node. This is not required, but it makes your HA pair easier to administer because the disks have the same ID on each node. This is true for Step 5 also.

3. Cable the disk shelf Channel B Output port to the Channel B Input port of the next disk shelf in loop 1.
4. Repeat step 3 for any remaining disk shelves in loop 1.
5. Cable the Channel B Output port of the last disk shelf in the loop to Fibre Channel port A4 of Node A.

   This provides the redundant multipath HA connection for Channel B.

6. Cable Fibre Channel port B2 of Node B to the Channel A Input port of the first disk shelf of Node B loop 1.
7. Cable the disk shelf Channel A Output port to the Channel A Input port of the next disk shelf in loop 1.
8. Repeat step 7 for any remaining disk shelves in loop 1.
9. Cable the Channel A Output port of the last disk shelf in the loop to Fibre Channel port A3 of Node A.

   This provides the redundant multipath HA connection for Channel A.

10. Repeat steps 2 to 9 for each pair of loops in the HA pair, using ports 3 and 4 for the next loop, ports 5 and 6 for the next one, and so on.

Result

Node B is completely cabled.

After you finish

Proceed to cable the HA interconnect.

Cabling the HA interconnect (all systems except N6200 series)

To cable the HA interconnect between the HA pair nodes, you must make sure that your interconnect adapter is in the correct slot and connect the adapters on each node with the optical cable.

About this task

This procedure applies to all dual-chassis HA pairs (HA pairs in which the two controller modules reside in separate chassis) except N6200 series systems, regardless of disk shelf type.

Steps

1. See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) to ensure that your interconnect adapter is in the correct slot for your system in an HA pair.

   For systems that use an NVRAM adapter, the NVRAM adapter functions as the HA interconnect adapter.

2. Plug one end of the optical cable into one of the local node's HA adapter ports, then plug the other end into the partner node's corresponding adapter port.

   You must not cross-cable the HA interconnect adapter. Cable the local node ports only to the identical ports on the partner node.

   If the system detects a cross-cabled HA interconnect, the following message appears:

   ```
   HA interconnect port <port> of this appliance seems to be connected to
   port <port> on the partner appliance.
   ```

3. Repeat Step 2 for the two remaining ports on the HA adapters.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

Cabling the HA interconnect (N6200 series systems in separate chassis)

To enable the HA interconnect between N6200 series controller modules that reside in separate chassis, you must cable the onboard 10-GbE ports on one controller module to the onboard GbE ports on the partner.

About this task

This procedure applies to N6200 series systems regardless of the type of attached disk shelves.

Steps

1. Plug one end of the 10-GbE cable to the c0a port on one controller module.
2. Plug the other end of the 10-GbE cable to the c0a port on the partner controller module.
3. Repeat the preceding steps to connect the c0b ports.

   Do not cross-cable the HA interconnect adapter; cable the local node ports only to the identical ports on the partner node.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

Cabling a mirrored HA pair

To cable a mirrored HA pair, you identify the ports you need to use on each node, and then you cable the ports, and then you cable the HA interconnect.

About this task

This procedure explains how to cable a configuration using EXN1000 or EXN2000 unit disk shelves.

For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

Note: If you are installing an HA pair between gateway systems with third-party storage, see the *Gateway Installation Requirements and Reference Guide* for information about cabling gateways to storage arrays. See the gateway *Implementation Guide* for information about configuring storage arrays to work with gateways.

The sections for cabling the HA interconnect apply to all systems regardless of disk shelf type.

Steps

1. *Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections* on page 48
2. *Creating your port list for mirrored HA pairs* on page 49
3. *Cabling the Channel A EXN1000 or EXN2000 unit disk shelf loops* on page 49
4. *Cabling the Channel B EXN1000, EXN2000, or EXN4000 unit disk shelf loops* on page 52
5. *Cabling the redundant multipath HA connection for each loop* on page 54
6. *Cabling the HA interconnect (all systems except N6200 series)* on page 56
7. *Cabling the HA interconnect (N6200 series systems in separate chassis)* on page 56

Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections

Before cabling your HA pair, you need to identify which Fibre Channel ports to use to connect your disk shelves to each storage system, and in what order to connect them.

Keep the following guidelines in mind when identifying ports to use:

- Every disk shelf loop in the HA pair requires two ports on the node, one for the primary connection and one for the redundant multipath HA connection.
  A standard HA pair with one loop for each node uses four ports on each node.
- Onboard Fibre Channel ports should be used before using ports on expansion adapters.
- Always use the expansion slots in the order shown in the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) for your platform for an HA pair.
- If using Fibre Channel HBAs, insert the adapters in the same slots on both systems.

See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) to obtain all slot assignment information for the various adapters you use to cable your HA pair.

After identifying the ports, you should have a numbered list of Fibre Channel ports for both nodes, starting with Port 1.

Cabling guidelines for a quad-port Fibre Channel HBA

If using ports on the quad-port, 4-Gb Fibre Channel HBAs, use the procedures in the following sections, with the following additional guidelines:

- Disk shelf loops using ESH4 modules must be cabled to the quad-port HBA first.
- Disk shelf loops using AT-FCX modules must be cabled to dual-port HBA ports or onboard ports before using ports on the quad-port HBA.
- Port A of the HBA must be cabled to the In port of Channel A of the first disk shelf in the loop. Port A of the partner node's HBA must be cabled to the In port of Channel B of the first disk shelf in the loop. This ensures that disk names are the same for both nodes.
- Additional disk shelf loops must be cabled sequentially with the HBA's ports.
  Port A is used for the first loop, port B for the second loop, and so on.

- If available, ports C or D must be used for the redundant multipath HA connection after cabling all remaining disk shelf loops.
- All other cabling rules described in the documentation for the HBA and the *N series Introduction and Planning Guide* must be observed.

Creating your port list for mirrored HA pairs

After you determine the Fibre Channel ports to use, you create a table identifying which ports belong to which port pool.

About this task

Mirrored HA pairs, regardless of disk shelf type, use SyncMirror to separate each aggregate into two plexes that mirror each other. One plex uses disks in pool 0 and the other plex uses disks in pool 1. You must assign disks to the pools appropriately.

Follow the guidelines for software-based disk ownership in the *Data ONTAP Storage Management Guide for 7-Mode*.

For more information about SyncMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

Step

1. Create a table specifying the port usage; the cabling diagrams in this document use the notation "P1-3" (the third port for pool 1).

   For an N5000 series HA pair that has two mirrored loops, the port list might look like the following example:

Pool 0	**Pool 1**
P0-1: onboard port 0a	P1-1: onboard port 0c
P0-2: onboard port 0b	P1-2: onboard port 0d
P0-3: slot 2 port A	P1-3: slot 4 port A
P0-4: slot 2 port B	P1-4: slot 4 port B

After you finish

Proceed to cable the Channel A loops.

Cabling the Channel A EXN1000 or EXN2000 unit disk shelf loops

To begin cabling of the disk shelves, you cable the appropriate pool ports on the node to the Channel A modules of the disk shelf stack for the pool.

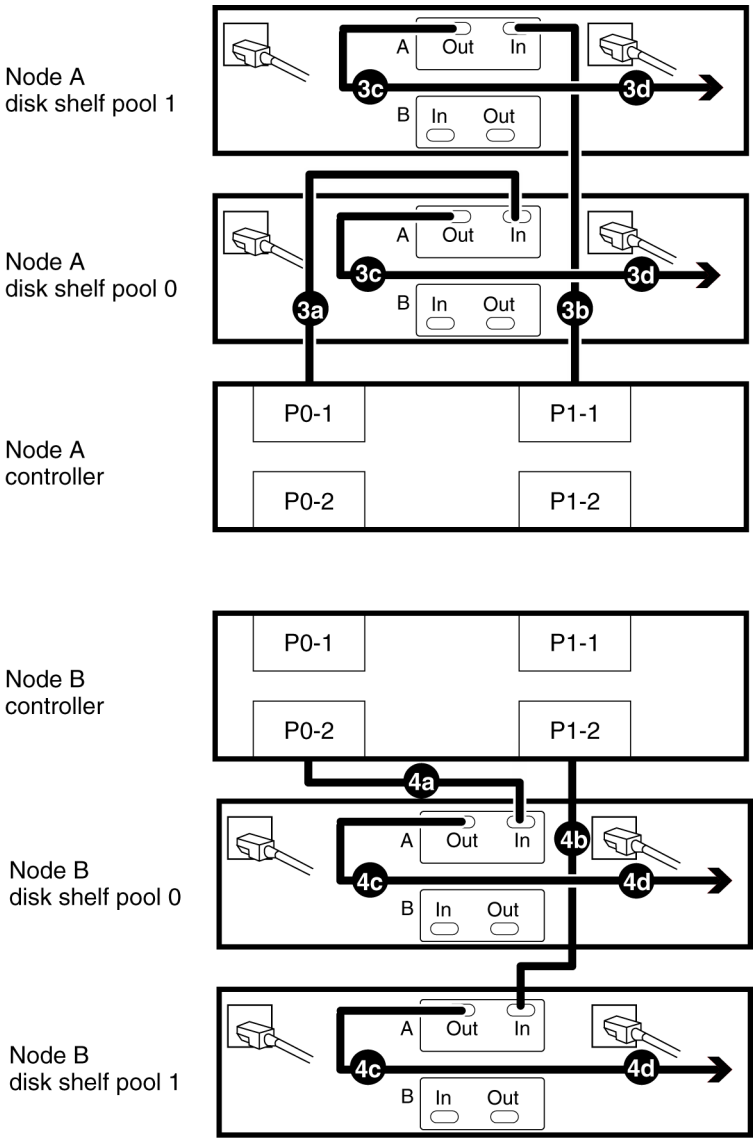
About this task

- This procedure uses multipath HA, required on all systems.

- This procedure does not apply to SAS disk shelves.
  For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

Steps

1. Complete your port list.
2. Review the cabling diagram before proceeding to the cabling steps.
   - The circled numbers in the diagram correspond to the step numbers in the procedure.
   - The location of the Input and Output ports on the disk shelves vary depending on the disk shelf models.
     Make sure that you refer to the labeling on the disk shelf rather than to the location of the port shown in the diagram.
   - The location of the Fibre Channel ports on the controllers is not representative of any particular storage system model; determine the locations of the ports you are using in your configuration by inspection or by using the *Installation and Setup Instructions* for your model.
   - The port numbers refer to the list of Fibre Channel ports you created.
   - The diagram only shows one loop per node and one disk shelf per loop.
     Your installation might have more loops, more disk shelves, or different numbers of disk shelves between nodes.



3. Cable Channel A for Node A.
   a) Cable the first port for pool 0 (P0-1) of Node A to the first Node A disk shelf Channel A Input port of disk shelf pool 0.
   b) Cable the first port for pool 1 (P1-1) of Node A to the first Node A disk shelf Channel A Input port of disk shelf pool 1.
   c) Cable the disk shelf Channel A Output port to the next disk shelf Channel A Input port in the loop for both disk pools.

      Note: The illustration shows only one disk shelf per disk pool. The number of disk shelves per pool might be different for your configuration.

d) Repeat substep 3c, connecting Channel A output to input, for any remaining disk shelves in this loop for each disk pool.
e) Repeat Substep 3a through Substep 3d for any additional loops for Channel A, Node A, using the odd numbered port numbers (P0-3 and P1-3, P0-5, and P1-5, and so on).

4. Cable Channel A for Node B

a) Cable the second port for pool 0 (P0-2) of Node B to the first Node B disk shelf Channel A Input port of disk shelf pool 0.
b) Cable the second port for pool 1 (P1-2) of Node B to the first Node B disk shelf Channel A Input port of disk shelf pool 1.
c) Cable the disk shelf Channel A Output port to the next disk shelf Channel A Input port in the loop for both disk pools.
d) Repeat substep 4c, connecting Channel A output to input, for any remaining disk shelves in each disk pool.
e) Repeat substep 4a through substep 4d for any additional loops on Channel A, Node B, using the even numbered port numbers (P0-4 and P1-4, P0-6, and P1-6, and so on).

After you finish

Proceed to cable the Channel B loops.

Cabling the Channel B EXN1000, EXN2000, or EXN4000 unit disk shelf loops

To provide the mirrored storage, you cable the mirrored pool ports on the node to the Channel B modules of the appropriate disk shelf stack.

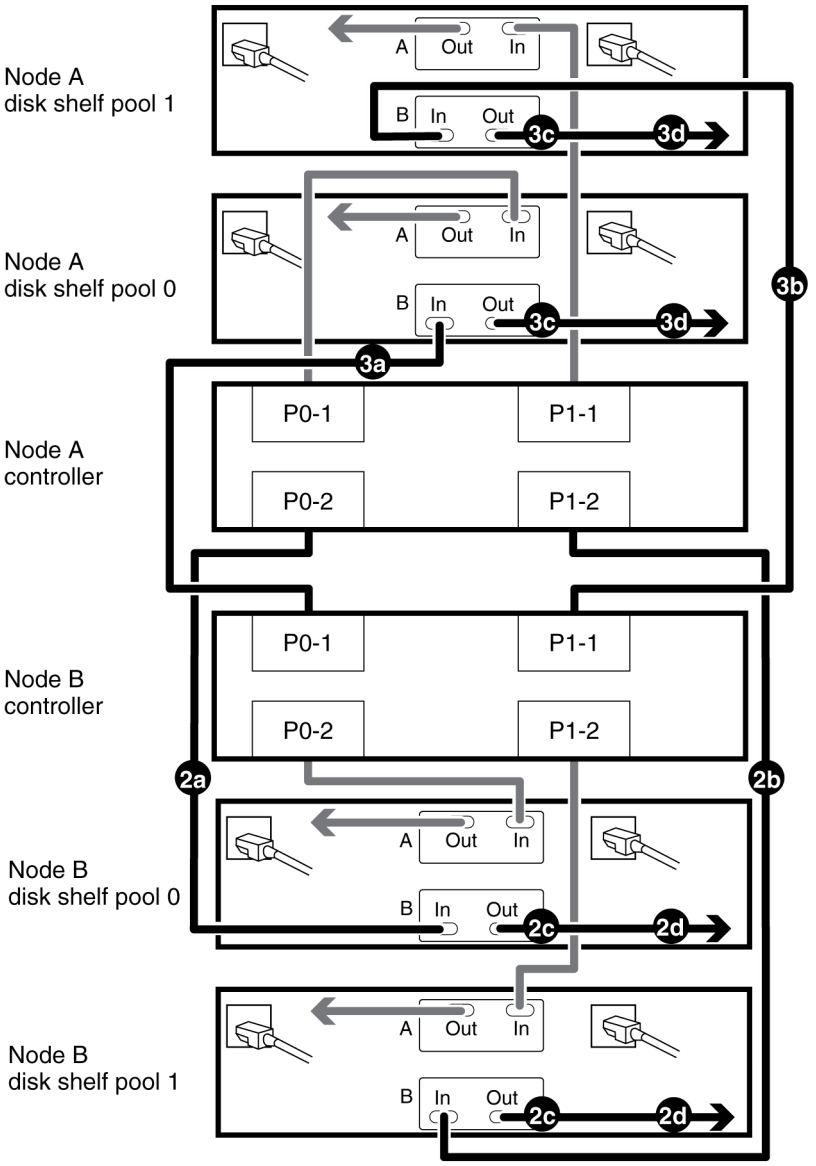
Steps

1. Review the cabling diagram before proceeding to the cabling steps.

- The circled numbers in the diagram correspond to the step numbers in the procedure.
- The location of the Input and Output ports on the disk shelves vary depending on the disk shelf models.

  Make sure that you refer to the labeling on the disk shelf rather than to the location of the port shown in the diagram.
- The location of the Fibre Channel ports on the controllers is not representative of any particular storage system model; determine the locations of the ports you are using in your configuration by inspection or by using the *Installation and Setup Instructions* for your model.
- The port numbers refer to the list of Fibre Channel ports you created.
- The diagram only shows one loop per node and one disk shelf per loop.

  Your installation might have more loops, more disk shelves, or different numbers of disk shelves between nodes.



2. Cable Channel B for Node A.
   a) Cable the second port for pool 0 (P0-2) of Node A to the first Node B disk shelf Channel B Input port of disk shelf pool 0.

      Note: Both channels of this disk shelf are connected to the same port on each node. This is not required, but it makes your HA pair easier to administer because the disks have the same ID on each node.

   b) Cable the second port for pool 1 (P1-2) of Node A to the first Node B disk shelf Channel B Input port of disk shelf pool 1.

c) Cable the disk shelf Channel B Output port to the next disk shelf Channel B Input port in the loop for both disk pools.

Note: The illustration shows only one disk shelf per disk pool. The number of disk shelves per pool might be different for your configuration.

d) Repeat Substep 2c, connecting Channel B output to input, for any remaining disk shelves in each disk pool.

e) Repeat Substep 2a through Substep 2d for any additional loops on Channel B, Node A, using the even numbered port numbers (P0-4 and P1-4, P0-6, and P1-6, and so on).

3. Cable Channel B for Node B.

a) Cable the first port for pool 0 (P0-1) of Node B to the first Node A disk shelf Channel B Input port of disk shelf pool 0.

b) Cable the first port for pool 1 (P1-1) of Node B to the first Node A disk shelf Channel B Input port of disk shelf pool 1.

c) Cable the disk shelf Channel B Output port to the next disk shelf Channel B Input port in the loop for both disk pools.

d) Repeat Substep 3c, connecting Channel B output to input, for any remaining disk shelves in each disk pool.

e) Repeat Substep 3a through Substep 3d for any additional loops for Channel B, Node B, using the odd numbered port numbers (P0-3 and P1-3, P0-5, and P1-5, and so on).

After you finish

Proceed to cable the HA interconnect.

Cabling the redundant multipath HA connection for each loop

To complete the multipath HA cabling for the disk shelves, you must add the final connection for each channel on the final disk shelf in each loop.

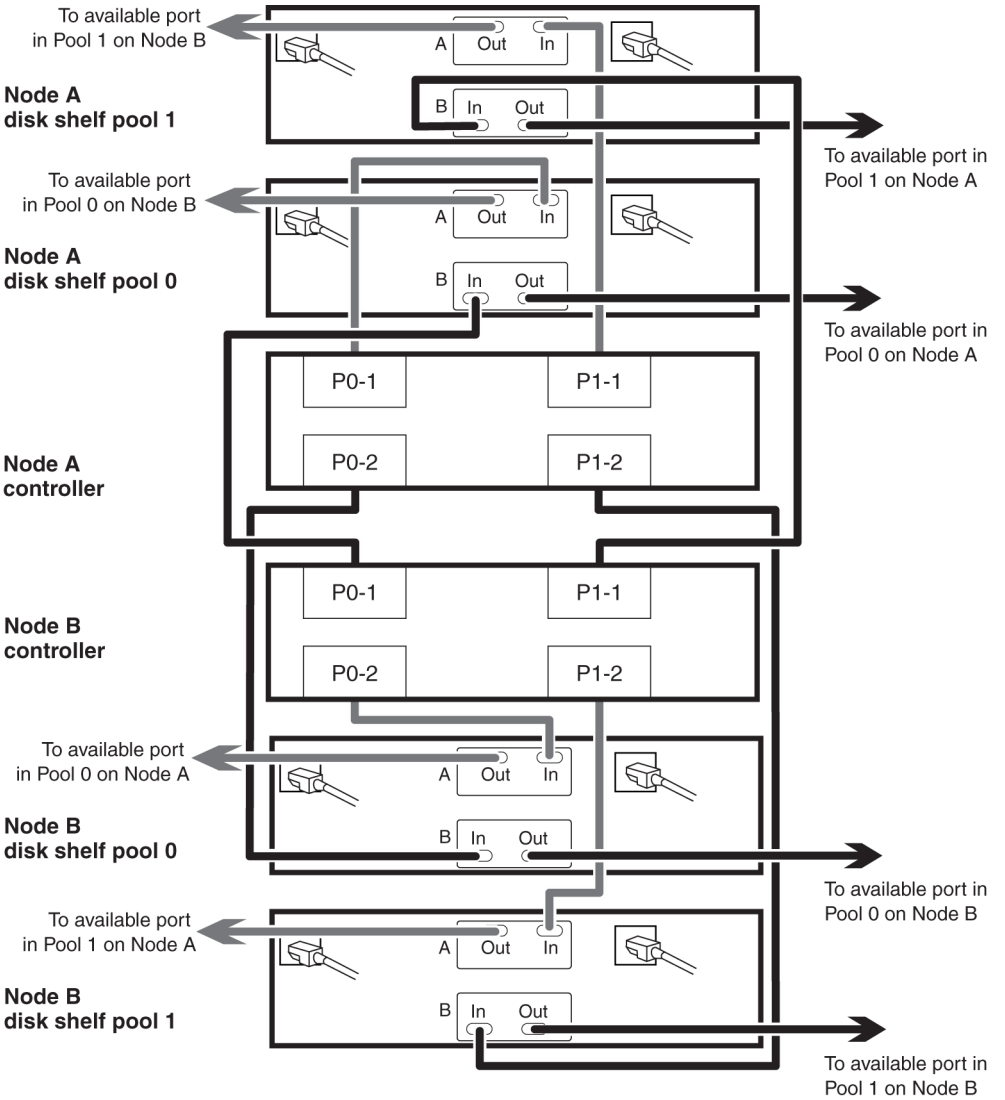
Steps

1. Review the cabling diagram before proceeding to the cabling steps.

- The circled numbers in the diagram correspond to the step numbers in the procedure.
- The location of the Input and Output ports on the disk shelves vary depending on the disk shelf models.

  Make sure that you refer to the labeling on the disk shelf rather than to the location of the port shown in the diagram.
- The location of the Fibre Channel ports on the controllers is not representative of any particular storage system model; determine the locations of the ports you are using in your configuration by inspection or by using the *Installation and Setup Instructions* for your model.
- The port numbers refer to the list of Fibre Channel ports you created.
- The diagram only shows one loop per node and one disk shelf per loop.

  Your installation might have more loops, more disk shelves, or different numbers of disk shelves between nodes.



2. Connect the Channel A output port on the last disk shelf for each loop belonging to Node A to an available port on Node B in the same pool.
3. Connect the Channel B output port on the last disk shelf for each loop belonging to Node A to an available port on Node B in the same pool.
4. Connect the Channel A output port on the last disk shelf for each loop belonging to Node B to an available port on Node B in the same pool.
5. Connect the Channel B output port on the last disk shelf for each loop belonging to Node B to an available port on Node B in the same pool.

Cabling the HA interconnect (all systems except N6200 series)

To cable the HA interconnect between the HA pair nodes, you must make sure that your interconnect adapter is in the correct slot and connect the adapters on each node with the optical cable.

About this task

This procedure applies to all dual-chassis HA pairs (HA pairs in which the two controller modules reside in separate chassis) except N6200 series systems, regardless of disk shelf type.

Steps

1. See the *N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11) to ensure that your interconnect adapter is in the correct slot for your system in an HA pair.

   For systems that use an NVRAM adapter, the NVRAM adapter functions as the HA interconnect adapter.

2. Plug one end of the optical cable into one of the local node's HA adapter ports, then plug the other end into the partner node's corresponding adapter port.

   You must not cross-cable the HA interconnect adapter. Cable the local node ports only to the identical ports on the partner node.

   If the system detects a cross-cabled HA interconnect, the following message appears:

   ```
   HA interconnect port <port> of this appliance seems to be connected to
   port <port> on the partner appliance.
   ```

3. Repeat Step 2 for the two remaining ports on the HA adapters.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

Cabling the HA interconnect (N6200 series systems in separate chassis)

To enable the HA interconnect between N6200 series controller modules that reside in separate chassis, you must cable the onboard 10-GbE ports on one controller module to the onboard GbE ports on the partner.

About this task

This procedure applies to N6200 series systems regardless of the type of attached disk shelves.

Steps

1. Plug one end of the 10-GbE cable to the c0a port on one controller module.
2. Plug the other end of the 10-GbE cable to the c0a port on the partner controller module.
3. Repeat the preceding steps to connect the c0b ports.

   Do not cross-cable the HA interconnect adapter; cable the local node ports only to the identical ports on the partner node.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

Required connections for using uninterruptible power supplies with standard or mirrored HA pairs

You can use a UPS (uninterruptible power supply) with your HA pair. The UPS enables the system to fail over gracefully if power fails for one of the nodes, or to shut down gracefully if power fails for both nodes. You must ensure that the correct equipment is connected to the UPS.

To gain the full benefit of the UPS, you must ensure that all the required equipment is connected to the UPS. The equipment that needs to be connected depends on whether your configuration is a standard or a mirrored HA pair.

For a standard HA pair, you must connect the controller, disks, and any FC switches in use.

For a mirrored HA pair, you must connect the controller and any FC switches to the UPS, as for a standard HA pair. However, if the two sets of disk shelves have separate power sources, you do not have to connect the disks to the UPS. If power is interrupted to the local controller and disks, the controller can access the remote disks until it shuts down gracefully or the power supply is restored. In this case, if power is interrupted to both sets of disks at the same time, the HA pair cannot shut down gracefully.

MetroCluster system installation with filers

You can install a stretch or fabric-attached MetroCluster configuration to provide complete data mirroring and takeover capabilities if a site is lost in a disaster. Fabric-attached MetroCluster configurations provide an HA pair with physically separated nodes at a greater distance than that provided by a stretch MetroCluster configuration.

Note: If you are installing the FibreBridge 6500N bridges as part of your MetroCluster configuration, see the *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11) for cabling procedures.

Related concepts

Disaster recovery using MetroCluster configurations on page 196

Setup requirements and restrictions for stretch MetroCluster configurations with filers on page 62

Setup requirements and restrictions for fabric-attached MetroCluster configurations with filers on page 70

Required documentation, tools, and equipment

Describes the IBM documentation and the tools required to install a MetroCluster configuration.

Required documentation

You must refer to some of the flyers and guides that are required to install a new MetroCluster configuration, or convert two stand-alone systems into a MetroCluster configuration.

IBM hardware and service documentation comes with your hardware and is also available at the N series support website (accessed and navigated as described in *Websites* on page 11).

The following table lists and briefly describes the documentation you might need to refer to when preparing a new MetroCluster configuration, or converting two stand-alone systems into a MetroCluster configuration:

Manual name	Description
The appropriate system cabinet guide	Describes how to install IBM N series equipment into a system cabinet.
N series Introduction and Planning Guide	Describes the physical requirements your site must meet to install IBM N series equipment.

Manual name	Description
The appropriate disk shelf guide	Describes how to cable a disk shelf to a storage system.
The appropriate hardware documentation for your storage system model	Describes how to install the storage system, connect it to a network, and bring it up for the first time.
Diagnostics Guide	Describes the diagnostics tests that you can run on the storage system.
Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode	Describes how to upgrade the storage system and disk firmware, and how to upgrade the storage system software.
Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode	Describes, among other topics, SyncMirror technology, which is used for mirrored HA pair.
Data ONTAP System Administration Guide for 7-Mode	Describes general storage system administration.
Data ONTAP Software Setup Guide for 7-Mode	Describes how to configure the software of a new storage system for the first time.
Fabric-attached MetroCluster Brocade Switch Configuration Guide	Describes how to configure Brocade switches for a fabric-attached MetroCluster configuration.
Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges	Describes how to install FibreBridge 6500N bridges as part of your MetroCluster configuration. You can find this document on the N series support website (accessed and navigated as described in *Websites* on page 11).
Fabric-attached MetroCluster Cisco Switch Configuration Guide	Describes how to configure Cisco switches for a fabric-attached MetroCluster configuration.

Required tools

Lists the tools you need to install the HA pair.

The following list specifies the tools you need to install the MetroCluster configuration:

- #1 and #2 Phillips screwdrivers
- Hand level
- Marker

Required equipment

You should receive the required equipment, such as storage system, HA interconnect adapter and so on to install a MetroCluster configuration.

See the appropriate hardware and service guide at the N series support website (accessed and navigated as described in *Websites* on page 11) to confirm your storage system type, storage capacity, and so on.

Note: For fabric-attached MetroCluster configurations, use the information in the appropriate hardware and service guide labeled for MetroCluster configurations. For stretch MetroCluster configurations, use the information in the appropriate hardware and service guide labeled "for HA Environments."

Required equipment	Stretch MetroCluster configuration	Fabric-attached MetroCluster configuration
Storage system	Two of the same type of storage systems.	
Storage	See the *IBM System Storage N series Introduction and Planning Guide* on the N series support website (accessed and navigated as described in *Websites* on page 11).	
FibreBridge 6500N bridges (if you are attaching SAS disk shelves)	Two FibreBridges are required for each stack of SAS disk shelves.	
HA interconnect adapter	InfiniBand adapter (Required only for systems that do not use an NVRAM5 or NVRAM6 adapter, which functions as the HA interconnect adapter.) FC-VI adapter (Required only for the N6000 series dual-controller systems.) **Note:** When the FC-VI adapter is installed in an N6000 series system, the internal InfiniBand interconnect is automatically deactivated.	FC-VI adapter

Required equipment	**Stretch MetroCluster configuration**	**Fabric-attached MetroCluster configuration**
FC-AL or FC HBA (FC HBA for Disk) adapters	Two or four Fibre Channel HBAs. These HBAs are required for 4-Gbps MetroCluster operation. Onboard ports can be used for 2-Gbps operation. **Note:** The ports on the Fibre Channel HBAs are labeled 1 and 2. However, the software refers to them as A and B. You see these labeling conventions in the user interface and system messages displayed on the console.	
Fibre Channel switches	N/A	Two pairs of Brocade or Cisco switches **Note:** The Fibre Channel switches must be of the same type. A mixture of switch types (such as Brocade 300 and Brocade 5100 switches) is not allowed.
SFP (Small Form Pluggable) modules	N/A	Two or four long-distance for inter-switch links, depending on whether you are using dual inter-switch links. The type of SFP needed depends on the distance between sites. One short-distance for each switch port used.
NVRAM adapter media converter	Only if using fiber cabling.	N/A

Required equipment	Stretch MetroCluster configuration	Fabric-attached MetroCluster configuration
Cables (provided with shipment unless otherwise noted)	• Four SC/LC (standard connector to low-profile connector) controller-to-disk shelf cables or FibreBridge 6500N • Two SC/LC IB HA adapter cables • Four SC/LC or LC/LC cables **Note:** For information about required cables, see the MetroCluster Compatibility Matrix on the N series support website (accessed and navigated as described in *Websites* on page 11).	• LC/LC controller-to-switch cables • LC/LC (for EXN2000) disk shelf-to-switch cables or FibreBridge 6500N • Two LC/LC inter-switch link cables, not provided in the shipment • Multiple disk shelf-to-disk shelf cables

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Setup requirements and restrictions for stretch MetroCluster configurations with filers

You must follow certain requirements and restrictions when setting up a new stretch MetroCluster configuration with a filer.

The requirement and restrictions for stretch MetroCluster configurations include those for a standard HA pair and those for a mirrored HA pair. In addition, the following requirements apply:

- Starting with Data ONTAP 8.1, stretch MetroCluster configurations support SAS disk shelves when used with FibreBridge 6500N bridges.
- Your storage system must meet all the compatibility requirements for FibreBridge 6500N bridges in the MetroCluster Compatibility Matrix on the N series support website (accessed and navigated as described in *Websites* on page 11).
- SAS, SATA, and Fibre Channel storage is supported on stretch MetroCluster configurations, but both plexes of the same aggregate must use the same type of storage.
- Stretch MetroCluster configuration using SAS disk shelves is supported up to 5 meters.

  For stretch MetroCluster configurations using SAS disk shelves having distance greater than five meters require the use of the FibreBridge 6500N bridges. Each stack of SAS disk shelves requires two FibreBridges.
- Stacks of SAS disk shelves can be added to a MetroCluster configuration that has EXN1000, EXN2000, or EXN4000 unit disk shelves.

- A stack of SAS disk shelves can contain shelves of SAS disk drives and shelves of SATA disk drives, but each SAS disk shelf can only contain SAS or SATA disk drives; you cannot mix SAS and SATA disk drives in the same disk shelf.

  To know about the number and the type of disk shelves supported in a stack, see the MetroCluster Compatibility Matrix on the N series support website (accessed and navigated as described in *Websites* on page 11).
- For stretch MetroCluster configurations using SAS disk shelves, each stack requires two Fibre Channel ports on each controller.
- For the number of SAS disk shelves and the types of SAS disk shelves supported in a stack, see the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11).
- Stretch MetroCluster configurations are not supported on N3400 systems.
- The following distance limitations dictate the default speed you can set:
  - If the distance between the nodes is 150m and you have an 8-Gb FC-VI adapter, the default speed is set to 8-Gb.

    If you want to increase the distance to 270m or 500m, you can set the default speed to 4-Gb or 2-Gb, respectively.
  - If the distance between nodes is between 150m and 270m and you have an 8-Gb FC-VI adapter, you can set the default speed to 4-Gb.
  - If the distance between nodes is between 270m and 500m and you have an 8-Gb FC-VI or 4-Gb FC-VI adapter, you can set the default speed to 2-Gb.
- The following licenses must be enabled on both nodes:
  - cf
  - syncmirror_local
  - cf_remote

Note: See the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11) for more information about hardware and firmware requirements for this configuration.

Related concepts

Setup requirements and restrictions for standard HA pairs on page 22

Setup requirements and restrictions for mirrored HA pairs on page 27

Converting an HA pair to a fabric-attached MetroCluster configuration

With the correct hardware, you can reconfigure an HA pair with EXN1000, EXN2000, or EXN4000 unit disk shelves to a fabric-attached MetroCluster configuration.

About this task

- If you are upgrading an existing HA pair to a MetroCluster configuration, you must upgrade disk firmware to the latest version.

  After upgrading disk firmware, you must power-cycle the affected disk drives to ensure that they work correctly in a fabric-attached MetroCluster configuration. You can download the latest disk firmware from the N series support website (accessed and navigated as described in *Websites* on page 11).
- If you are upgrading an N6000 series system, the resulting upgraded system can only have one controller in each chassis.

  If you have a chassis with two controllers, you must move one controller to a new chassis to form the partner node of the MetroCluster configuration. You must also obtain and install the FC-VI interconnect card on both systems.

Note:

- For details about this conversion process, see the *Best Practices for MetroCluster Design and Implementation*, at the N series support website (accessed and navigated as described in *Websites* on page 11).
- If you are converting an HA pair that has SAS disk shelves to a fabric-attached MetroCluster configuration, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* at the N series support website (accessed and navigated as described in *Websites* on page 11).

Steps

1. Update Data ONTAP, storage system firmware, and disk firmware, as described in the *Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode*, making sure to shut down the nodes to the boot prompt.
2. Remove any ATA drives in the configuration.

   ATA drives are not supported in a MetroCluster configuration.
3. Move the NVRAM adapter and FC-VI adapter to the correct slots for your model, as shown by the appropriate hardware and service guide at the N series support website (accessed and navigated as described in *Websites* on page 11).
4. Determine your switch and general configuration by completing the planning worksheet.

5. Set up and configure the local switches, and verify your switch licenses, as described in the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* and *Fabric-attached MetroCluster Cisco Switch Configuration Guide*.

   Note: The configuration and firmware requirements for Brocade and Cisco switches in a MetroCluster environment are different from the requirements for switches used in SAN environments. Always refer to MetroCluster documentation when installing and configuring your switches for a MetroCluster configuration.

6. Cable the local node.
7. Install the Data ONTAP licenses in the following order:
   a) cf
   b) syncmirror_local
   c) cf_remote
8. Configure the local node depending on the type of HA pair:

If you are converting a...	Then...
Standard HA pair	Set up mirroring and configure the local node.
Stretch MetroCluster configuration	Configure the local node.

9. Transport the partner node, disk shelves, and switches to the remote location.
10. Set up the remote node, disk shelves, and switches.

After you finish

Configure the MetroCluster configuration.

Related concepts

Configuring an HA pair on page 135
Disaster recovery using MetroCluster configurations on page 196

Related tasks

Cabling Node A on page 77
Cabling Node B on page 82
Disabling the change_fsid option in MetroCluster configurations on page 142

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Cabling a stretch MetroCluster configuration

The process of cabling a stretch MetroCluster configuration is the same as a mirrored HA pair. However, your systems must meet the requirements for a stretch MetroCluster configuration.

About this task

Note: If you are installing the FibreBridge 6500N bridge as part of your MetroCluster configuration, see the *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11) for cabling procedures.

Related concepts

Configuring an HA pair on page 135
Setup requirements and restrictions for stretch MetroCluster configurations with filers on page 62
Disaster recovery using MetroCluster configurations on page 196

Related tasks

Cabling a mirrored HA pair on page 47

Cabling a stretch MetroCluster configuration between single-enclosure HA pair systems

If you are configuring a stretch MetroCluster configuration between single-enclosure HA pair systems (for example, N6000 series systems), you must configure FC-VI interconnect adapter connections between the controllers.

About this task

Some storage systems support two controllers in the same chassis. You can configure two dual-controller systems into a pair of MetroCluster configurations. In such a configuration, the internal InfiniBand connections between the controllers are automatically deactivated. Therefore, the two controllers in the chassis are no longer in an HA pair with each other. Each controller is connected through FC-VI connections to another controller of the same type, so that the four controllers form two independent MetroCluster configurations.

Steps

1. Connect port A of the FC-VI adapter on the top controller of the local site to port A of the corresponding FC-VI adapter at the remote site.

2. Connect port B of the FC-VI adapter on the top controller of the local site to port B of the corresponding FC-VI adapter at the remote site.
3. Repeat steps 1 and 2 for connecting the FC-VI adapter on the bottom controller.
4. Cable the disk shelf loops for the stretch MetroCluster configuration formed by the top controllers as described in the procedure for cabling a mirrored HA pair.
5. Cable the disk shelf loops for the stretch MetroCluster configuration formed by the bottom controllers as described in the procedure for cabling a mirrored HA pair.

Related concepts

Stretch MetroCluster configuration on single-enclosure HA pairs on page 30

Related tasks

Cabling a mirrored HA pair on page 47

Changing the default configuration speed of a stretch MetroCluster configuration

The distance between your nodes and the FC-VI adapter speed dictates the default configuration speed of your stretch MetroCluster configuration. If the distance between nodes is greater than the supported default configuration speed, you must change the default configuration speed.

Before you begin

The stretch MetroCluster system's default configuration speed must conform to the stretch MetroCluster system setup requirements and restrictions.

About this task

- You enter the commands at the boot environment prompt, which can be CFE> or LOADER>, depending on your storage system model.
- You must perform these steps at both the nodes if they are configured at different speeds.

Steps

1. At the storage console prompt, halt the storage system by entering the following command:

   ```
   halt
   ```

2. Reset the configuration speed by entering the following commands.

If you want to set the speed to...	Then...

4 Gb	**a.** Enter the following command:
	`setenv ispfcvi-force-4G-only? True`
	b. If you previously modified the speed to 2 Gb, ensure that the 2-Gb port speed is not set by entering the following command:
	`unsetenv ispfcvi-force-2G-only?`
	c. Verify that your system is unconfigured for 2 Gb by entering the following command:
	`printenv ispfcvi-force-2G-only?`
	The storage system console displays output similar to the following:
	`Variable Name Value`
	`------------------- --------------`
	`ispfcvi-force-2G-only? *** Undefined ***`
	d. Verify that your storage system is configured for 4 Gb by entering the following command:
	`printenv ispfcvi-force-4G-only?`
	The storage system console displays output similar to the following:
	`Variable Name Value`
	`------------------- --------------`
	`ispfcvi-force-4G-only? true`

If you want to set the speed to...	Then...

2 Gb	**a.** Enter the following command:
	setenv ispfcvi-force-2G-only? True
	b. If you previously modified the default speed to 4 Gb, ensure that the 4-Gb speed is not set by entering the following command:
	unsetenv ispfcvi-force-4G-only?
	c. Verify that your storage system is unconfigured for 4 Gb by entering the following command:
	printenv ispfcvi-force-4G-only?
	The storage system console displays output similar to the following:
	`Variable Name Value`
	`------------------- ------------------`
	`ispfcvi-force-4G-only? *** Undefined ***`
	d. Verify that your storage system is configured for 2 Gb by entering the following command:
	printenv ispfcvi-force-2G-only?
	If your storage system is configured correctly, the system console displays output similar to the following:
	`Variable Name Value`
	`-------------------- --------------------`
	`ispfcvi-force-2G-only? true`

3. Boot the storage system by entering the following command:

   boot_ontap

Resetting a stretch MetroCluster configuration to the default speed

If you modified the default configuration speed in a stretch MetroCluster configuration using an FC-VI adapter, you can reset the speed to the default configuration speed by using the `unsetenv` command at the boot environment prompt.

About this task

- The boot environment prompt can be CFE> or LOADER>, depending on your storage system model.

- The steps require you to unset the previously configured speed, but because the speed is set to default automatically, you do not need to set the default speed explicitly.

Steps

1. At the storage prompt, halt the system by entering the following command:
   halt
2. Reset the configuration speed.

If you want to reset the speed from...	Then...

4 Gb	**a.** Enter the following command:
	unsetenv ispfcvi-force-4G-only?
	b. Verify that your system is unconfigured for 4 Gb by entering the following command:
	printenv ispfcvi-force-4G-only?
	The system console displays output similar to the following:
	`Variable Name Value`
	`------------------------------`
	`ispfcvi-force-4G-only? *** Undefined ***`

2 Gb	**a.** Enter the following command: **unsetenv ispfcvi-force-2G-only?** **b.** Verify that your system is unconfigured for 2 Gb by entering the following command: **printenv ispfcvi-force-2G-only?**

3. Boot the storage system by entering the following command:
   boot_ontap

Setup requirements and restrictions for fabric-attached MetroCluster configurations with filers

You must follow certain requirements and restrictions when setting up a new fabric-attached MetroCluster configuration.

The setup requirements for a fabric-attached MetroCluster configuration with filers include those for standard and mirrored HA configurations, with the following exceptions:

Note: See the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11) for more information about hardware and firmware requirements for this configuration.

Node requirements

- The nodes must be one of the following system models configured for mirrored volume use; each node in the pair must be the same model.
  - N5000 series systems, except for the N5500 and N5200 systems
  - N6000 series systems
  - N6200 series systems
  - N7000 series
  - N7x50T series systems
- Each node requires a 4-Gbps FC-VI (Fibre Channel-Virtual Interface) adapter; the slot position is dependent on the controller model.

  Note: For information about supported cards and slot placement, see the *N series Introduction and Planning Guide* on the N series support website (accessed and navigated as described in *Websites* on page 11).

  The FC-VI adapter is also called a VI-MC or VI-MetroCluster adapter.

- The 8-Gbps FC-VI adapter is supported only on the N6200 series and N7x50T series systems.
- If you want to convert a stretch MetroCluster configuration to a fabric-attached MetroCluster configuration, and you have modified the default configuration speed of FC-VI adapters by setting the boot environment variables to `True`, you must reset it to default configuration speed before the conversion.

Disk and disk shelf requirements

- The only Fibre Channel disk shelves supported are EXN1000, EXN2000, or EXN4000 unit.
- Starting with Data ONTAP 8.1, fabric-attached MetroCluster configurations also support SAS disk shelves when used with FibreBridge 6500N.
- Each stack of SAS disk shelves requires two FibreBridges.
- Your storage system must meet all the compatibility requirements for FibreBridge 6500N in the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11).
- EXN1000, EXN2000, or EXN4000 unit disk shelves using SATA drives and AT-FCX storage is not supported.
- You can connect a maximum of two EXN1000, EXN2000, or EXN4000 unit disk shelves to each loop.
- A stack of SAS disk shelves can contain shelves of SAS disk drives and shelves of SATA disk drives, but each SAS disk shelf can only contain SAS or SATA disk drives; you cannot have SAS and SATA disk drives in the same disk shelf.

  For more information about the number and the types of disk shelves supported in a stack, see the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11).
- Stacks of SAS disk shelves can be added to a fabric-attached MetroCluster configuration that has EXN1000 or EXN2000 unit or EXN4000 unit disk shelves.

- For the number of SAS disk shelves and the types of SAS disk shelves supported in a stack, see the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11).

Capacity limits

The maximum capacity for a system configured in a fabric-attached MetroCluster configuration is the smallest of the following limits:

- The maximum storage capacity for the node.

  Note: For the maximum storage capacity, see the appropriate hardware and service guide on the N series support website (accessed and navigated as described in *Websites* on page 11).
- 840 Fibre Channel disks (60 disk shelves).
- In a MetroCluster configuration using both FC and SAS disk shelves), the total disk drives count must not exceed storage capacity.

Fibre Channel switch requirements

Note: For the most up-to-date switch information, including supported switches and firmware downloads, see the N series support website (accessed and navigated as described in *Websites* on page 11).

- Each site of the MetroCluster configuration requires two switches.
- Switches must be a supported Brocade or Cisco model supplied by IBM.

  Customer-supplied switches are not supported.
- The two switches within the fabric must be the same model and must be licensed for the same number of ports.
- All four switches for a particular fabric-attached MetroCluster configuration must support the same maximum speed.
- Switches must be running the correct firmware version.
- Configurations using SAS storage and FibreBridge 6500N bridges supports Brocade or Cisco switch models supplied by IBM.

  See the *MetroCluster Compatibility Matrix* on the N series support website (accessed and navigated as described in *Websites* on page 11) for more information about supported switch models.

License requirements

The following licenses must be installed.

- cf
- syncmirror_local
- cf_remote

Related concepts

Setup requirements and restrictions for standard HA pairs on page 22
Setup requirements and restrictions for mirrored HA pairs on page 27

Cabling a fabric-attached MetroCluster configuration

You cable the fabric-attached MetroCluster configuration so that the controller and the disk shelves at each site are connected to Brocade or Cisco switches. In turn, the switches at one site are connected through inter-switch links to the switches at the other site.

Before you begin

To cable a fabric-attached MetroCluster configuration, you must be familiar with HA pairs, the Brocade or Cisco command-line interface, and synchronous mirroring. You must also be familiar with the characteristics of fabric-attached MetroCluster configurations. You must also have the following information:

- Correct Brocade or Cisco licenses for each switch
- Unique domain IDs for each of the switches

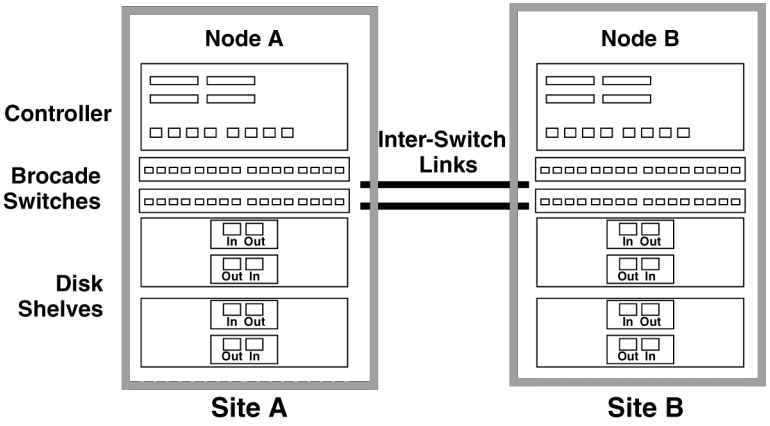
  Note: You can use the switch numbers (1, 2, 3, and 4) as the switch Domain ID.
- Ethernet IP address for both the switches and nodes

  Note: The switches ship with a default IP address (10.77.77.77), which you can use if the switches are not attached to a network.
- Ethernet subnetmask
- Gateway address

About this task

A fabric-attached MetroCluster configuration involves two nodes at physically separated sites. To differentiate these nodes in this documentation, the guide refers to the two nodes as Node A and Node B.

Note: If you are using SAS disk shelves, the disk shelves connect to FibreBridge 6500N bridges. To see an example of a cabled fabric-attached MetroCluster system with FibreBridges and SAS disk shelves, see the *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* at *www.ibm.com/storage/support/nseries*.



Steps

1. *Planning the fabric-attached MetroCluster installation* on page 74
2. *Configuration differences for fabric-attached MetroCluster configurations on single-enclosure HA pairs* on page 76
3. *Configuring the switches* on page 76
4. *Cabling Node A* on page 77
5. *Cabling Node B* on page 82
6. *Assigning disk pools* on page 86
7. *Verifying disk paths* on page 88

Related concepts

Setup requirements and restrictions for fabric-attached MetroCluster configurations with filers on page 70

Configuring an HA pair on page 135

Disaster recovery using MetroCluster configurations on page 196

Related tasks

Disabling the change_fsid option in MetroCluster configurations on page 142

Planning the fabric-attached MetroCluster installation

You must fill out the fabric-attached MetroCluster configuration worksheet to record specific cabling information about your fabric-attached MetroCluster configuration. You must identify several pieces

of information that you use during configuration procedures. Recording this information can reduce configuration errors.

Step

1. Fill in the following tables.

   Each site has two Brocade Fibre Channel switches. Use the following table to record the configured names, IP addresses, and domain IDs of these switches.

Switch number...	At site...	Is named...	IP address...	Domain ID...
1	A			
2	A			
3	B			
4	B			

   In addition to on-board ports, each site has an FC-VI adapter and two Fibre Channel HBAs that connect the node to the switches. Use the following table to record which switch port these adapters are connected to.

This adapter...	At site...	Port 1 of this adapter is...		Port 2 of this adapter is...	
		Cabled to switch...	Switch port...	Cabled to switch...	Switch port...
FC-VI adapter	A	1		2	
	B	3		4	
FC HBA 1	A	1		2	
	B	3		4	
FC HBA 2	A	1		2	
	B	3		4	

   Disk shelves or FibreBridge 6500N bridges (if you are using SAS disk shelves) at each site connect to the Fibre Channel switches. Use the following table to record which switch port the disk shelves or FibreBridge 6500N bridges are connected to.

   Note: If you are using SAS or SATA disk shelves, the FibreBridges connect to the disk shelves. Each FibreBridge 6500N bridge needs to be connected to only one switch through one of the FC ports on the FibreBridge bridge. For cabling examples, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* at *www.ibm.com/storage/support/nseries*.

Disk shelf or FibreBridge 6500N...	At site...	Belonging to...	Connects to switches...	On switch port...
1	A	Node A Pool 0	1 and 2	
2				
3		Node B Pool 1		
4				
5	B	Node B Pool 0	3 and 4	
6				
7		Node A Pool 1		
8				

Configuration differences for fabric-attached MetroCluster configurations on single-enclosure HA pairs

When configuring a fabric-attached MetroCluster configuration between single-enclosure HA pair (systems with two controllers in the same chassis), you get two separate MetroCluster configurations.

A single-enclosure HA pair can be connected to another HA pair to create two separate fabric-attached MetroCluster configurations. The internal InfiniBand connection in each system is automatically deactivated when the FC-VI card is installed in the controller.

You must cable each fabric-attached MetroCluster configuration separately by using the normal procedures for each and assign the storage appropriately.

Related concepts

Fabric-attached MetroCluster configuration on single-enclosure HA pairs on page 33

Configuring the switches

To configure the switches, you must refer to the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* for your Brocade switch model. The Brocade switch configuration for a MetroCluster configuration is different than the one used for a SAN configuration.

Step

1. To configure your Brocade switches, see the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* for your switch model at the N series support website (accessed and navigated as described in *Websites* on page 11).

   Note: The configuration and firmware requirements for Brocade switches in a MetroCluster environment are different from the requirements for switches used in SAN environments.

Always refer to MetroCluster documentation, such as the MetroCluster Compatibility Matrix, when installing and configuring your MetroCluster switches.

After you finish

Proceed to configure Node A.

Related information

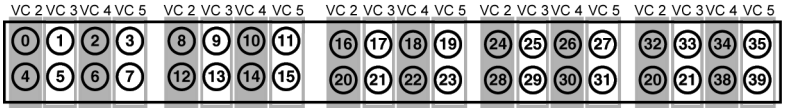
IBM N series support website: www.ibm.com/storage/support/nseries

Virtual channel rules

You must follow the correct virtual channel rules on the Brocade switches.

Use the switch virtual channel (VC) rules when cabling the switch. In this case, switch traffic is distributed across VCs to avoid bottlenecks. The FC-VI and inter-switch links are cabled to ports in one VC, and the disk shelf and controller connections are cabled to ports in another VC.

Virtual channel	Ports
2	0, 4, 8, 12, 16, 20, 32, 36
3	1, 5, 9, 13, 17, 21, 33, 37
4	2, 6, 10, 14, 18, 22, 26, 30, 34, 38
5	3, 7, 11, 15, 19, 23, 27, 31, 35, 39



Related information

Fabric-attached MetroCluster Systems: Brocade Switch Configuration Guide

Cabling Node A

To cable the local node (Node A), you need to attach the controller and the disk shelves to the switches, connect the HA interconnect to the switches, and ensure that the disk shelves in the configuration belong to the correct pools.

About this task

If you are using SAS disk shelves, the SAS disk shelves connect to the FibreBridge 6500N bridges and the bridges connect to the switches.

Steps

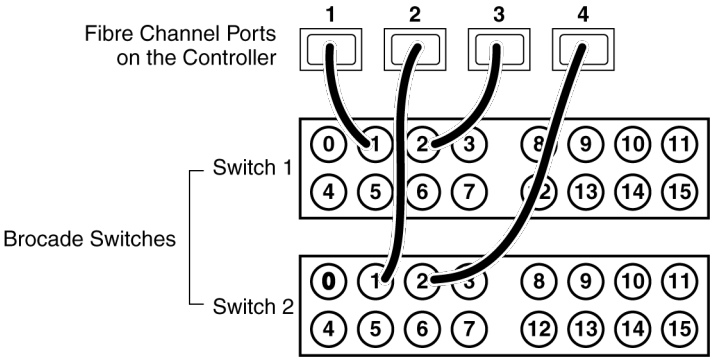
1. *Cabling the controller* on page 78

2. *Cabling the shelves* on page 79
3. *Cabling the FC-VI adapter and inter-switch link* on page 81

Cabling the controller

You can use this procedure to cable the Fibre Channel ports on the controller to the Brocade switches.

About this task



Steps

1. Determine which Fibre Channel ports on your system that you want to use and create a list showing the order you want to use them.

   Note: The numbers in the example refer to the preferred order of usage, not the port ID. For example, Fibre Channel port 1 might be port e0a on the controller.

2. Cable the first two Fibre Channel ports of Node A to the same numbered ports on Switch 1 and Switch 2, for example, port 1.

   They must not go to ports in the virtual channel that you have reserved for the FC-VI and inter-switch link connections. In the example, we are using virtual channel 2 for the FC-VI and inter-switch link. Virtual channel 2 includes ports 0, 4, 8, and 12.

3. Cable the second two Fibre Channel ports of Node A to the same numbered ports on Switch 1 and Switch 2, for example, port 2.

   They must not go to ports in the virtual channel that you have reserved for the FC-VI and inter-switch link connections. In the example, ports 0, 4, 8, and 12 are excluded.

   Note: The switches in the example are 16-port switches.

After you finish

Proceed to cable disk shelves to the switches.

Related concepts

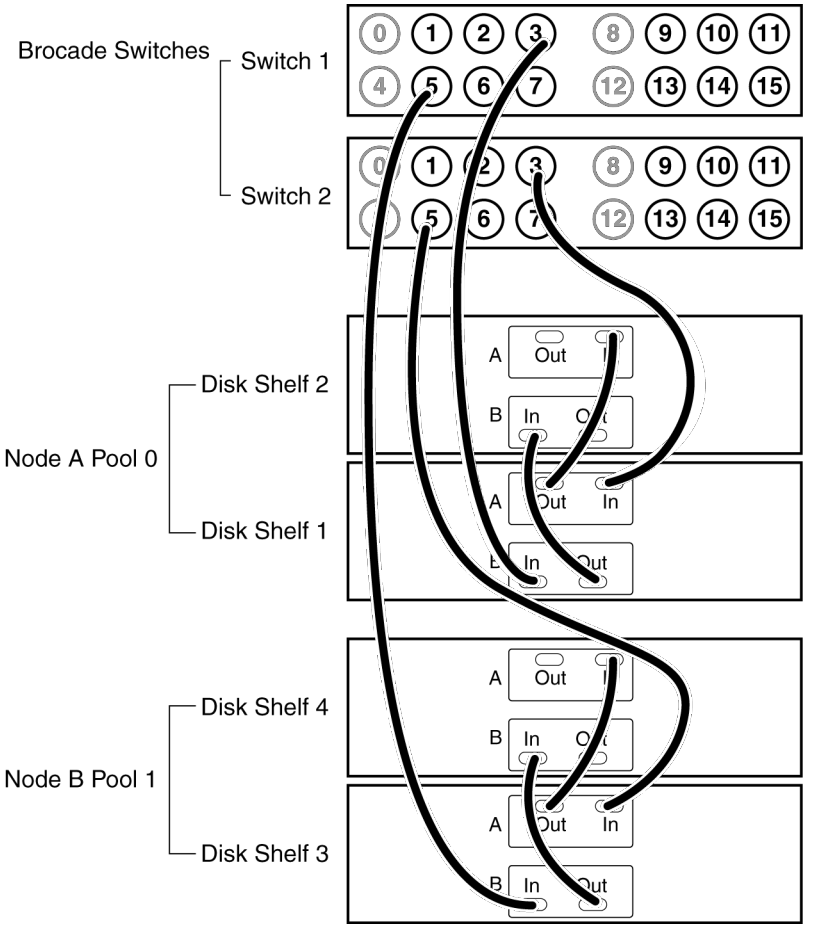
Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections on page 41

Cabling the shelves

You must cable the EXN1000, EXN2000, or EXN4000 unit disk shelf loops on Node A directly to the switches.

About this task

To cable SAS disk shelves and FibreBridge 6500N, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11).



Note: You can cable a maximum of two disk shelves on each loop.

Steps

1. Connect the Node A pool 0 disk shelves to the switches by completing the following substeps:
   a) Connect the Input port of the A module on disk shelf 1 to any available port on Switch 2 other than ports 0, 4, 8, and 12.

      In the example, switch port 3 is used.
   b) Connect the Input port of the B module on disk shelf 1 to the same port on Switch 1.

      In the example, switch port 3 is used.
   c) Connect disk shelf 1 to disk shelf 2 by connecting the Output ports of the module of disk shelf 1 to the Input ports of the corresponding module of the next disk shelf.
   d) If your disk shelf modules have terminate switches, set them to Off on all but the last disk shelf in the disk pool, then set the terminate switches on the last disk shelf to On.

   Note: ESH4 modules are self-terminating and therefore do not have a terminate switch.
2. Connect the Node B pool 1 disk shelves to the switches by completing the following substeps:
   a) Connect the Input port of the module Channel A on disk shelf 3 to any available port on Switch 2 other than ports 0, 4, 8, and 12.

      The example uses switch port 5.
   b) Connect the Input port of the module Channel B on disk shelf 3 to the same port on Switch 1.

      The example uses switch port 5.
   c) Connect disk shelf 3 to disk shelf 4 by connecting the Output ports of the module of disk shelf 3 to the Input ports of the corresponding module of the next disk shelf.
   d) If your disk shelf modules have terminate switches, set them to Off on all but the last disk shelf in the disk pool, then set the terminate switches on the last disk shelf to On.
3. If you have more than one loop, connect the other loops in the same manner.

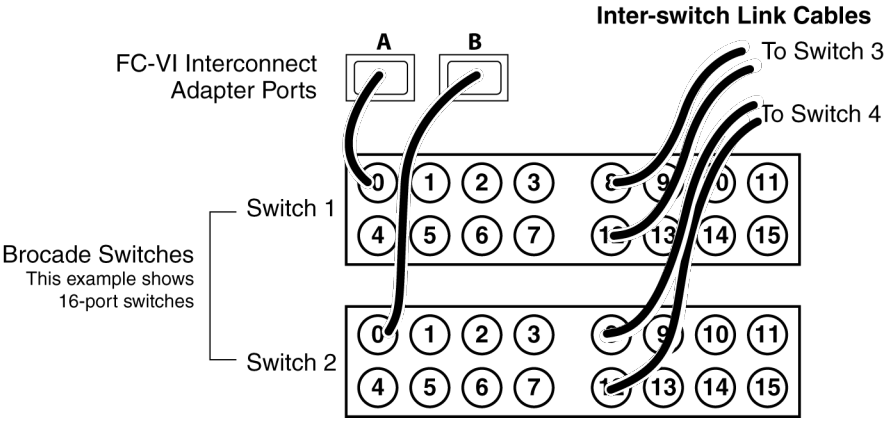
After you finish

Proceed to cable the FC-VI adapter and inter-switch connections.

Cabling the FC-VI adapter and inter-switch link

Describes how to cable the HA interconnect and inter-switch link on Node A.

About this task



Steps

1. Using the ports in the virtual channel you have selected for the FC-VI and inter-switch link connections, connect one port of the FC-VI adapter on switch 1 and the second port to the same port on switch 2.

   In the example we are using virtual channel 2, including ports 0, 4, 8, and 12, for the FC-VI and inter-switch link connections.

   Note: There should be one FC-VI adapter connection for each switch. Make sure that you have the FC-VI adapter in the correct slot for your system, as shown in the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11).

2. Connect an inter-switch link cable to a port in the selected virtual channel on each switch, or, if using a dual inter-switch link, connect two cables in the selected virtual channel.

   In the example we are using virtual channel 2, which includes ports 0, 4, 8, and 12, and are using ports 8 and 12 on switch 1 and switch 2 for the inter-switch links.

   Note: If using dual inter-switch links, traffic isolation must be configured on the switches.

After you finish

Proceed to cable Node B.

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Cabling Node B

To cable the remote node (Node B), you need to attach the controller and the disk shelves to the switches, connect the HA interconnect to the switches, and ensure that the disk shelves in the configuration belong to the correct pools.

About this task

If you are using SAS disk shelves, the SAS disk shelves connect to the FibreBridge 6500N bridge and the bridges connect to the switches.

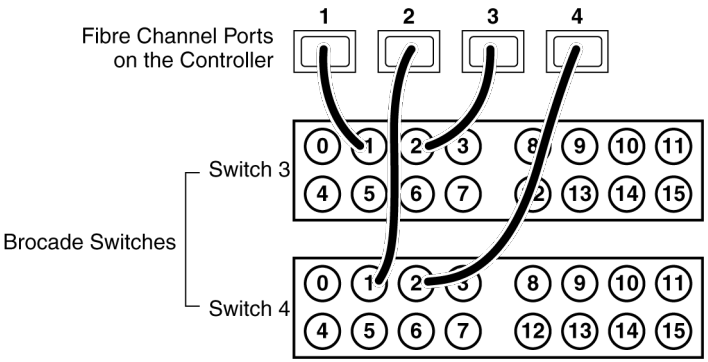
Steps

1. *Cabling the controller* on page 82
2. *Cabling the shelves* on page 83
3. *Cabling the FC-VI adapter and inter-switch link* on page 85

Cabling the controller

You can use this procedure to cable the Fibre Channel ports on the controller to the Brocade switches.

About this task



Steps

1. Determine which Fibre Channel ports on your system that you want to use and create a list showing the order you want to use them.

   Note: The numbers in the example refer to the preferred order of usage, not the port ID. For example, Fibre Channel port 1 might be port e0a on the controller.

2. Cable the first two Fibre Channel ports of Node B to the same numbered ports Switch 3 and Switch 4, for example, port 1.

They must go to ports in the virtual channel that you have reserved for the FC-VI and inter-switch link connections. In the example, we are using virtual channel 2 for the FC-VI and inter-switch link. Virtual channel 2 includes ports 0, 4, 8, and 12.

3. Cable the second two Fibre Channel ports of Node B to the same numbered ports Switch 3 and Switch 4, for example, port 2.

   They must not go to ports in the virtual channel that you have reserved for the FC-VI and inter-switch link connections. In the example, ports 0, 4, 8, and 12 are excluded.

After you finish

Proceed to cable disk shelves to the switches.

Related concepts

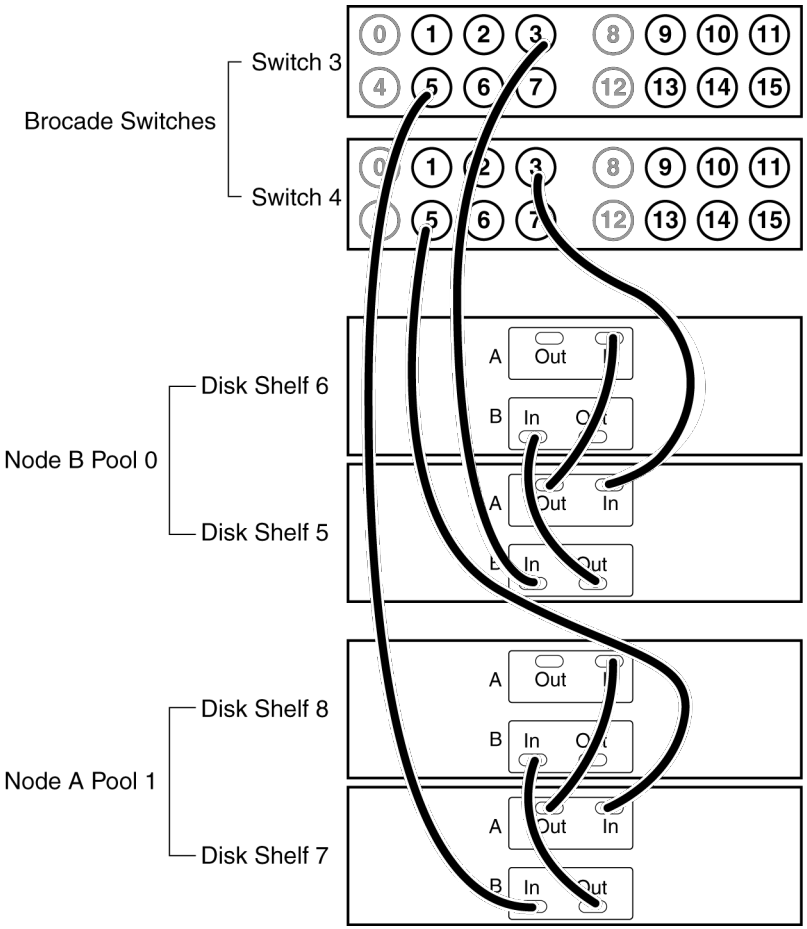
Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections on page 41

Cabling the shelves

You must cable the EXN1000, EXN2000, or EXN4000 unit disk shelf loops on Node B directly to the switches.

About this task

To cable SAS disk shelves and FibreBridge 6500N bridges, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11).



Note: You can cable a maximum of two disk shelves on each loop.

Steps

1. Connect the Node B pool 0 disk shelves to the switches by completing the following substeps:
   a) Connect the Input port of the A module on disk shelf 5 to any available port on Switch 4 that is not in the virtual channel reserved for the FC-VI and inter-switch link connections.

      The example uses switch port 3.
   b) Connect the Input port of the B module on disk shelf 5 to the same port on Switch 3.

      The example uses switch port 3.
   c) Connect disk shelf 5 to disk shelf 6 by connecting the Output ports of the module of disk shelf 5 to the Input ports of the corresponding module of the next disk shelf.
   d) If your disk shelf modules have terminate switches, set them to Off on all but the last disk shelf in the disk pool, then set the terminate switches on the last disk shelf to On.

Note: ESH4 modules are self-terminating and therefore do not have a terminate switch.

2. Connect the Node B pool 1 disk shelves to the switches by completing the following substeps:
   a) Connect the Input port of the module Channel A on disk shelf 7 to any available port on Switch 4 that is not in the virtual channel reserved for the FC-VI and inter-switch link connections.

      The example uses switch port 5.
   b) Connect the Input port of the module Channel B on disk shelf 7 to the same port on Switch 3.

      The example uses switch port 5.
   c) Connect disk shelf 7 to disk shelf 8 by connecting the Output ports of the module of disk shelf 7 to the Input ports of the corresponding module of the next disk shelf.
   d) If your disk shelf modules have terminate switches, set them to Off on all but the last disk shelf in the disk pool, then set the terminate switches on the last disk shelf to On.
3. If you have more than one loop, connect the other loops in the same manner.

After you finish

Proceed to cable the FC-VI adapter and inter-switch connections.

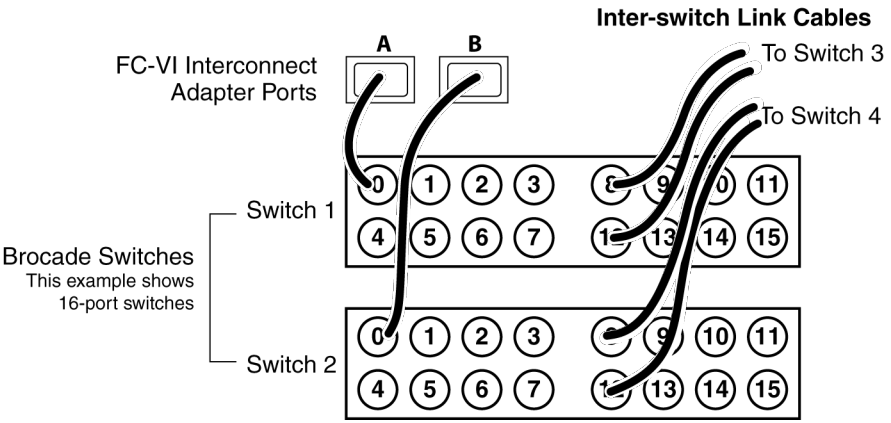
Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Cabling the FC-VI adapter and inter-switch link

You must cable the HA interconnect and inter-switch link on Node B.

About this task



Steps

1. Connect one port of the FC-VI adapter to a port in the virtual channel that you have reserved for the FC-VI and inter-switch link connections.

   In the example, port 0 on switch 1 and port 0 on switch 2 is used.

   Note: There should be one FC-VI adapter connection for each switch. Make sure that you have the FC-VI adapter in the correct slot for your system, as shown in the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 11).

2. Connect an inter-switch link cable to a port in the selected virtual channel on each switch, or if using dual inter-switch links, connect two cables in the selected virtual channel.

   The example uses virtual channel 2, which includes ports 0, 4, 8, and 12, and uses port 8 and port 12 on switch 1 and switch 2 for the inter-switch links.

   Note: If using dual inter-switch links, traffic isolation must be configured on the switches.

After you finish

Proceed to assign disks to disk pools.

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

Assigning disk pools

You must assign the attached disk shelves to the appropriate pools.

About this task

You can explicitly assign disks on the attached disk shelves to the appropriate pool with the `disk assign` command. Using wildcards in the command enables you to assign all the disks on a disk shelf with one command.

The following table shows the pool assignments for the disk shelves in the example used in this section.

Disk shelf...	At site...	Belongs to...	And is assigned to that node's...
Disk shelf 1	Site A	Node A	Pool 0
Disk shelf 2			
Disk shelf 3		Node B	Pool 1
Disk shelf 4			

Disk shelf...	At site...	Belongs to...	And is assigned to that node's...
Disk shelf 5	Site B	Node B	Pool 0
Disk shelf 6			
Disk shelf 7		Node A	Pool 1
Disk shelf 8			

Note: Pool 0 always contains the disks that are local to (at the same site as) the storage system that owns them.

Pool 1 always contains the disks that are remote to the storage system that owns them.

Steps

1. Boot Node A into Maintenance mode, if you haven't already.
2. Assign the local disks to Node A pool 0 by entering the following command at the console:

   disk assign switch2:port3.*-p *0*
   This indicates that the disks attached to port 3 of switch 2 are assigned to pool 0. The asterisk (*) indicates that all disks attached to the port are assigned.
3. Assign the remote disks to Node A pool 1 by entering the following command at the console:

   disk assign switch4:port5.* -p *1*
   This indicates that the disks attached to port 5 of switch 4 are assigned to pool 1. The asterisk (*) indicates that all disks attached to the port are assigned.
4. Boot Node B into Maintenance mode, if you haven't already.
5. Assign the local disks to Node B pool 0 by entering the following command at the console:

   disk assign switch4:port3.* -p *0*
   This indicates that the disks attached to port 3 of switch 4 are assigned to pool 0. The asterisk (*) indicates that all disks attached to the port are assigned.
6. Assign the remote disks to Node B pool 1 by entering the following command at the console:

   disk assign switch2:port5.* -p 1
   This indicates that the disks attached to port 5 of switch 2 are assigned to pool 1. The asterisk (*) indicates that all disks attached to the port are assigned.

After you finish

Proceed to verify the disk paths on the system.

Verifying disk paths

You must use this procedure to verify your disk paths for configurations using EXN1000, EXN2000, or EXN4000 unit or SAS shelves.

Steps

1. Boot Node A into normal mode, if necessary.
2. Enter the following command to confirm that your aggregates and volumes are operational and mirrored:

   `aggr status`

   See the *Data ONTAP Storage Management Guide for 7-Mode* for information about the `aggr status` command.
3. Repeat steps 1 and 2 on Node B.

Setting preferred primary port in a MetroCluster configuration

After cabling a MetroCluster configuration, you can define the primary port between the four switches to carry the traffic. You can set the preferred primary port in a shared-switches configuration by using the `ic primary set` command.

About this task

By default, the primary port is 0.

Steps

1. To set the primary port, enter the following command:

   `ic primary set 0|1 [-r]`

   You must use the `-r` option for the primary port to take effect immediately.

   Note: If the FC-VI link between the preferred primary ports fails, the fabric-attached MetroCluster configuration fails over to the secondary FC-VI link. When the primary link is restored, the fabric-attached MetroCluster configuration again uses the primary link. The failover temporarily causes the logs to be unsynchronized.

Example

If you want to set the preferred primary port to 1, enter the following command:

`ic primary set 1 -r`

In a shared-switches configuration, if you have set the preferred primary port of FMC1 (FMC1-1 and FMC1-2) to 1, the preferred primary port for FMC2 (FMC2-1 and FMC2-2) is set to 0.

2. To view the current primary port, enter the following command:

   ```
   ic primary show
   ```

   It displays the primary port.

Removing the preferred primary port in a fabric-attached MetroCluster configuration

You can remove the port that you assigned as primary port in a fabric-attached MetroCluster configuration.

Step

1. To remove the primary port, enter the following command:

   ```
   ic primary unset
   ```

Required connections for using uninterruptible power supplies with MetroCluster configurations

You can use a UPS (Uninterruptible Power Supply) with your MetroCluster configurations. The UPS enables the system to fail over gracefully if power fails for one of the nodes, or to shut down gracefully if power fails for both nodes. You must ensure that the correct equipment is connected to the UPS.

The equipment that you need to connect to the UPS depends on how widespread a power outage you want to protect against. Always connect both controllers, any Fibre Channel switches in use, and any inter-switch link infrastructure (for example, a Dense Wavelength Division Multiplexing, or DWDM) to the UPS.

You can leave the disks on the regular power supply. In this case, if power is interrupted to one site, the controller can access the other plex until it shuts down or power is restored. If, however, power is interrupted to both sites at the same time and the disks are not connected to the UPS, the MetroCluster configuration cannot shut down gracefully.

Setting up a shared-switches configuration

In a shared-switches configuration, two fabric-attached MetroCluster configurations share the same four switches and the ISLs between them.

Shared-switches configuration are cost-effective because you can use the four Brocade switches between two MetroCluster configurations.

Related concepts

Requirements for a shared-switches configuration with third-party storage on page 97

Requirements for a shared-switches configuration

There are certain requirements that you must have when creating a shared-switches configuration.

- Fabric-attached MetroCluster configuration
- Brocade 5100 switch
- SAS disk shelves with FibreBridge 6500N

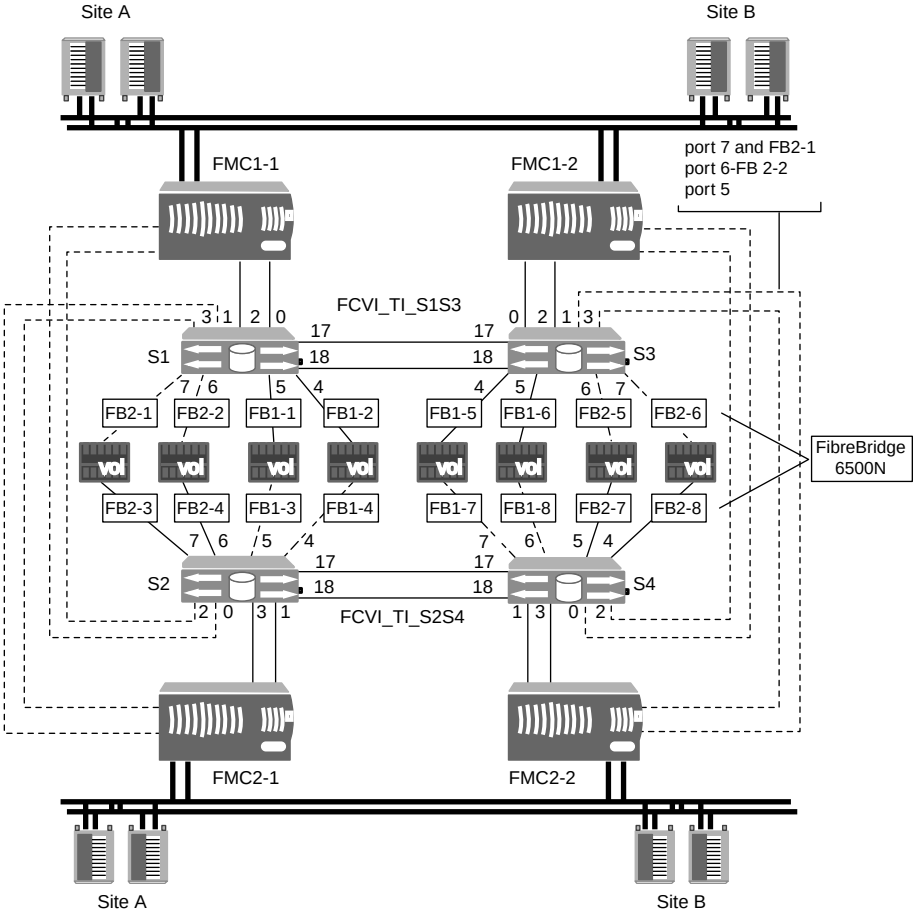
  To know how to install and configure FibreBridge 6500N as part of your MetroCluster configuration, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges* on the N series support website (accessed and navigated as described in *Websites* on page 11).
- A minimum two ISLs per fabric

Cabling the FC-VI adapter and ISL in a shared-switches configuration

Cabling in a shared-switches configuration involves cabling the FC-VI and HBA ports to the Brocade switches, and cabling the FibreBridge 6500N bridges to the Brocade switches and the SAS disk shelves.

About this task

In a shared-switches configuration, four nodes join together to form two fabric-attached MetroCluster configurations. In the illustration, FMC1-1 and FMC1-2 form one fabric-attached MetroCluster configuration. FMC1, and FMC2-1 and FMC2-2 form another fabric-attached MetroCluster configuration FMC2.



Steps

1. Connect the ports of the FC-VI adapter to primary and secondary switches for both the storage systems in the MetroCluster configuration.

   In the illustration, the FC-VI adapter from storage system FMC1-1 connects to primary switch S1 and secondary switch S2 through port 0.

   The FC-VI adapter from storage system FMC2-1 connects to primary switch S2 and secondary switch S1 through port 1.

2. Connect the ports of the HBA to the primary and secondary switches.

   In the illustration, the HBA from storage system FMC1-1 connects to primary switch S1 and secondary switch S2 through port 2.

   The HBA from storage system FMC2-1 connects to primary switch S2 and secondary switch S1 through port 3.

3. If you are using the FibreBridge 6500N bridge, complete the following steps:

   For details about installing FibreBridge 6500N bridge as part of your MetroCluster configuration, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges*, available at the N series support website (accessed and navigated as described in *Websites* on page 11).

   a) Connect the FibreBridge 6500N bridge to the switches.

      In the illustration, the FibreBridge bridge FB1-1 of the FMC1-1 connects to switch S1 through port 5.

   b) Connect the FibreBridge 6500N bridge to the disk shelves.

4. Repeat steps 1 through 3 at another site.

5. When cabling at both the sites are complete, connect an ISL cable to a port on each switch.

   In the illustration, the fabric containing switches S1 and S3 has two ISLs and the fabric containing switches S2 and S4 also has two ISLs.

   The switch S1 connects to switch S3 and switch S2 connects to switch S4 through ports 17 and 18.

After you finish

Configure the traffic isolation zone on the switches. For more information, see the *Fabric-attached MetroCluster Brocade Switch Configuration Guide*.

After cabling the shared-switches configuration, you must define the preferred primary port.

Related information

IBM N series support website: www.ibm.com/storage/support/nseries

MetroCluster configurations with third-party storage

You can install a stretch or fabric-attached MetroCluster configuration to provide complete data mirroring and takeover capabilities if a site is lost in a disaster. Gateways that use only third-party storage are supported for MetroCluster configurations (no native disks).

MetroCluster configurations are supported only for Data ONTAP operating in 7-Mode.

Fabric-attached MetroCluster configurations provide an HA pair with physically separated nodes at a greater distance than that provided by a stretch MetroCluster configuration.

Installing a MetroCluster configuration with third-party storage involves planning, connecting the gateways and other devices in the MetroCluster, and testing to ensure that the MetroCluster configuration is set up correctly.

Related concepts

Understanding stretch MetroCluster configurations on page 28
Understanding fabric-attached MetroCluster configurations on page 32

Planning a MetroCluster configuration with third-party storage

Creating a detailed plan for your MetroCluster configuration helps you understand the unique requirements for a MetroCluster configuration with third-party storage. A plan also helps you communicate with other people involved in the installation. Installing a MetroCluster configuration involves connecting and configuring a number of devices, which might be done by different people.

Related concepts

Implementation overview for a MetroCluster configuration with third-party storage on page 94
Requirements for a MetroCluster configuration with third-party storage on page 95
Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104
Planning zoning for a MetroCluster configuration with third-party storage on page 105

Implementation overview for a MetroCluster configuration with third-party storage

Implementing a MetroCluster configuration with third-party storage involves planning your implementation, installing hardware, connecting multiple devices, configuring Data ONTAP, and testing the MetroCluster configuration to ensure it is operating correctly.

The following tasks must be completed, in the order shown, to set up a MetroCluster configuration with third-party storage. Storage array configuration is performed by the storage array administrator or the storage array vendor. Zoning is often performed by a switch administrator.

1. Planning your MetroCluster implementation.
2. Setting up the storage array to present array LUNs to Data ONTAP and configuring the parameters that Data ONTAP requires to work with Data ONTAP.
3. Installing the FC-VI adapter on each gateway, if it is not installed already.

   Note: New gateways that are ordered for a MetroCluster configuration are shipped with the correct FC-VI adapter. If you are configuring a MetroCluster on existing gateways, you might need to install a new adapter or move an existing adapter to another slot. The *N series Introduction and Planning Guide* contains information about the required FC-VI adapter for your platform and the slot to install it in.

4. Connecting the local gateway to the fabric.
5. Connecting the remote gateway to the fabric.
6. Connecting the switched fabric.
7. Connecting the storage array to the fabric.
8. Configuring zoning.
9. Assigning array LUNs to specific gateways.
10. Configuring Data ONTAP features.
11. Testing the MetroCluster.

Related concepts

Connecting devices in a MetroCluster configuration with third-party storage on page 107
Setting up Data ONTAP after connecting devices in a MetroCluster configuration with third-party storage on page 124
Testing a MetroCluster configuration with third-party storage on page 124

Related tasks

Connecting the local gateways in a MetroCluster configuration on page 108
Connecting the remote gateways in a MetroCluster configuration on page 111
Connecting the switch fabric in a MetroCluster configuration with third-party storage on page 114

Connecting the fabric and storage array in a MetroCluster configuration with third-party storage on page 116

Configuring zoning in a MetroCluster configuration with third-party storage on page 122

Requirements for a MetroCluster configuration with third-party storage

There are some unique requirements for setting up a MetroCluster configuration with third-party storage.

The *Gateway Interoperability Matrix* contains the latest policies for a MetroCluster configurations with third-party storage and includes information about supported storage arrays, switches, and gateways. The *Gateway Interoperability Matrix* is the final authority for information about requirements and restrictions for MetroCluster configurations with third-party storage.

Requirements for gateways

- Only gateways (not filers) can be deployed in a MetroCluster configuration with third-party storage.
- The gateways in a MetroCluster configuration must use only third-party storage.

  Gateways using native disks cannot be deployed in a MetroCluster configuration, even if they also use third-party storage.
- The gateway platform must be identified in the *Gateway Interoperability Matrix* as supported for MetroCluster configurations.
- A system with a two controllers in the same enclosure, such as an N6200 series system, requires an FC-VI adapter.

  The *N series Introduction and Planning Guide* contains information about adapter requirements for different models.

Requirements for storage arrays

- The storage arrays must be identified in the *Gateway Interoperability Matrix* as supported for MetroCluster configurations.
- The storage arrays in the MetroCluster configuration must be symmetric, which means the following:
  - The two storage arrays must be in the same vendor family.

    The *Gateway Implementation Guide for Third-Party Storage* contains details about storage array families.
  - A pool of array LUNs for the mirrored storage must be created on each of the two storage arrays.

    The array LUNs must be the same size.
  - Disk types (FC, SATA, or SAS) used for mirrored storage must be the same on both storage arrays.
  - Storage arrays that provide tiered configurations (for example, Hitachi) must use the same tiered topology on each side of the MetroCluster configuration.
  - The root volume must be mirrored for successful takeover to occur.

Requirements for Fibre Channel switches

- The switches and switch firmware must be identified in the *Gateway Interoperability Matrix* as supported for MetroCluster configurations.
- Each fabric must have two switches.
- Each gateway must be connected to storage using redundant components so that there is redundancy in case of device failure.

The *Gateway Interoperability Matrix* contains the latest information about switch requirements and supported switches. The configuration and firmware requirements for switches in a MetroCluster environment might differ from those in other gateway configurations.

Zoning requirements

- Single-initiator zoning is recommended.
  Single-initiator zoning limits each zone to a single gateway FC initiator port. It also improves discovery and boot time because the gateway FC initiators do not attempt to discover each other.
- FC-VI ports on the FC adapters must be zoned end-to-end across the fabric.
  The *Gateway Interoperability Matrix* contains specific guidelines for FC-VI.

Requirements for ISLs

Data ONTAP supports using one or two ISLs, depending on the configuration. The *Gateway Interoperability Matrix* contains information about configuration if you have one or two ISLs.

For specific information about Traffic Isolation zoning for FC-VI traffic only, see the *Brocade's Zoning Feature* section in the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* on the N series support website (accessed and navigated as described in *Websites* on page 11). Regular zoning guidelines apply for gateways and storage arrays.

SyncMirror requirements

- SyncMirror is required for a MetroCluster configuration.
- SyncMirror with a MetroCluster configuration requires the following licenses, which must be installed in the order shown:
  - cf
  - syncmirror_local
  - cf_remote
- Two separate storage arrays are required for the mirrored storage.

  Note: Using two different storage arrays for mirroring is optional in non MetroCluster configurations.
- Two sets of LUNs are required—one set for the aggregate on the local storage array (pool0) and another set at the remote storage array for the mirror of the aggregate (the other plex of the aggregate, pool1).

The *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode* provides more details about planning for setting up SyncMirror for a MetroCluster configuration with third-party storage.

Requirements for a shared-switches configuration with third-party storage

There are specific requirements for creating a shared-switches configuration with third-party storage.

You need the following to create a shared-switches MetroCluster configuration with third-party storage:

- Four Brocade 5100 switches
- Two storage arrays at each site

  One storage array at one site is used for the first MetroCluster configuration and the other storage array at the site is used for the second MetroCluster configuration.

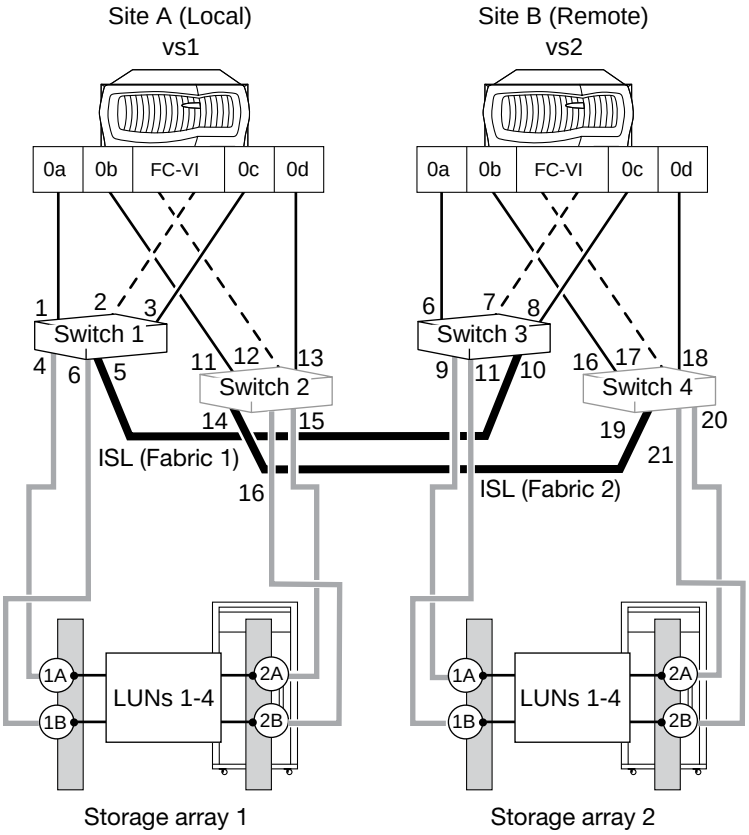
  Note: A single storage array at each site is supported if the target ports on the storage array and the array LUNs are not shared between the two MetroCluster configurations.
- A minimum of two ISLs per fabric

Recommended fabric-attached MetroCluster configuration with third-party storage

When configuring a fabric-attached MetroCluster configuration with gateways, you should follow the best practice recommendations for how to connect the hardware components. The *Gateway Interoperability Matrix* contains information about the hardware components that are supported for a gateway MetroCluster configuration.

The following illustration shows the components and best practice configuration of a fabric-attached gateway MetroCluster configuration. The fabric-attached MetroCluster configuration shown provides the same single-point-of-failure protections that are available for all mirrored HA pairs.

Note: Gateway FC initiator port names differ between some platforms. Your platform's port names might not match the port names shown in the illustration.



The following sections describe the connections for the fabric-attached MetroCluster configuration in the previous illustration.

Gateway interconnections

The gateways in the sample MetroCluster illustration are configured to be an HA pair. They are interconnected by connecting ports A and B on each system's FC-VI adapter to alternate switch fabrics.

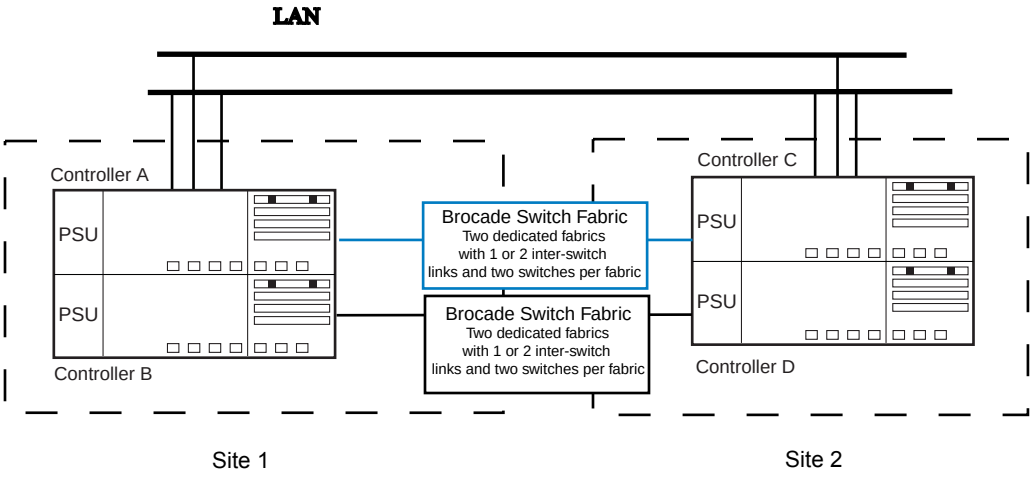
Gateway	Switch	Fabric
vs1: Port A vs1: Port B	Switch 2, Port 12 Switch 1, port 2	2 1
vs2: Port A vs2: Port B	Switch 4, port 17 Switch 3, port 7	2 1

If your gateways are dual-enclosure systems with two controllers in the same chassis (such as a N6200 series system), you connect the top controller in one system to the top controller in the other

system through the fabric. You connect the bottom controllers between the fabric in the same way. These two separate connections result in two separate fabric-attached MetroCluster configurations.

Note: The internal InfiniBand connection in each system is automatically deactivated when the FC-VI adapter is installed in the controller.

The following illustration shows the connections of the gateway controllers to the switch fabric for dual-enclosure systems with two controllers in the same chassis.



Inter-Switch Link connections (ISLs)

Fabric-attached MetroCluster configurations use a switched fabric to connect the local half of the configuration to the remote half of the configuration. The sample MetroCluster illustration shows the following:

- Switches 1 and 3 are connected to each other (Fabric 1).
  The first fabric in the MetroCluster configuration begins from Switch 1 on Site A (local) and is completed by connecting the ISL cable to the first switch on Site B (remote)—Switch 3.
- Switches 2 and 4 are also connected (Fabric 2).
  The second fabric is created using Switch 2 on Site A (local), connected through a second ISL cable, to the second switch on Site B (remote)—Switch 4.

The following table lists the ISLs in this configuration.

ISL connection	Switch	Fabric
Port 5 on switch 1	Port 10 on switch 3	Fabric 1
Port 14 on switch 2	Port 19 on switch 4	Fabric 2

Gateway-to-switch connections

The best practice connections in the recommended MetroCluster illustration eliminate a single point of failure in the following ways:

- FC initiator ports on the same Fibre Channel controller chip (for example, port 0a and 0b) connect to alternate fabrics.
- Multiple paths and zones ensure that FC initiator ports on the same controller chips access the array LUN from different gateways and switch fabrics.

The following table lists the connections from the gateway to the switch.

Gateway port	Switch
vs1: FC port 0a	Switch 1: Port 1
vs1: FC port 0b	Switch 2: Port 11
vs1: FC port 0c	Switch 1: Port 3
vs1: FC port 0d	Switch 2: Port 13
vs2: FC port 0a	Switch 3: Port 6
vs2: FC port 0b	Switch 4: Port 16
vs2: FC port 0c	Switch 3: Port 8
vs2: FC port 0d	Switch 4: Port 18

Storage array-to-switch connections

Best practice connections from the storage array to the switch are as follows:

- Ports 1A and 2A on each storage array connect to alternate fabrics.
- Ports 1B and 2B on each storage array connect to alternate fabrics.
- Gateways are configured to access any array LUN on two storage array paths (1A and 2A or 1B and 2B).

Storage array port	Switch	Fabric
Array 1: Port 1A	Switch 1: Port 4	1
Array 1: Port 2A	Switch 2: Port 15	2
Array 1: Port 1B	Switch 1: Port 6	1
Array 1: Port 2B	Switch 2: Port 16	2
Array 2: Port 1A	Switch 3: Port 9	1
Array 2: Port 2A	Switch 4: Port 20	2

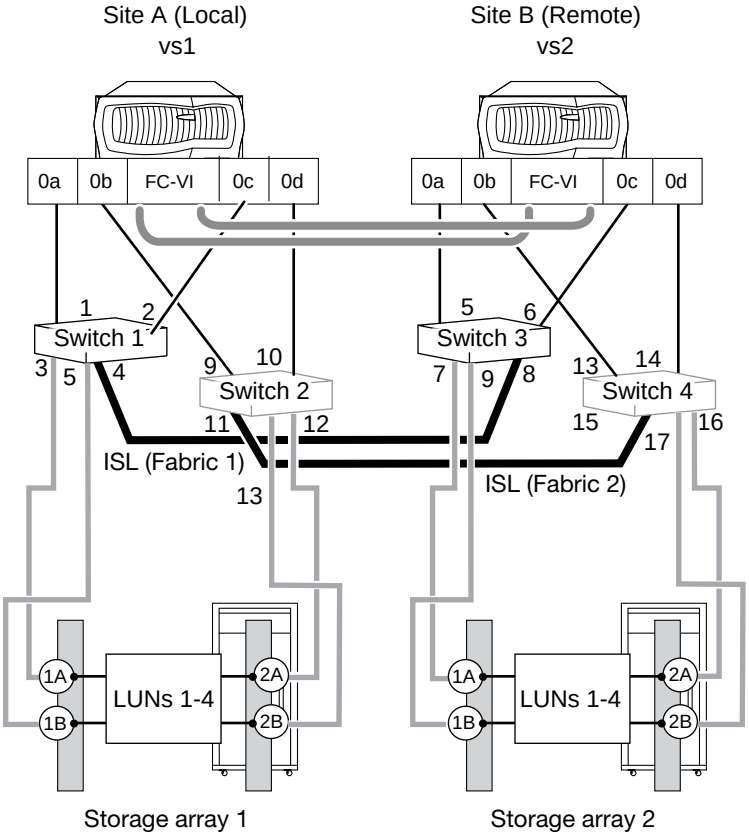
Storage array port	Switch	Fabric
Array 2:Port 1B	Switch 3: Port 11	1
Array 2:Port 2B	Switch 4: Port 21	2

Recommended stretch MetroCluster configuration with third-party storage

When configuring a stretch MetroCluster configuration with gateways, you should follow the best practice recommendations for how to connect the hardware components. The *Gateway Interoperability Matrix* contains information about the hardware components that are supported for a gateway MetroCluster configuration.

The following illustration shows the components and best practice for a stretch MetroCluster configuration with third-party storage.

Note: Gateway FC initiator port names differ on some gateway platforms. Gateway FC initiator port names differ between some platforms. Your platform's port names might not match the port names shown in the illustration. The *N series Introduction and Planning Guide* contains information that helps you determine in which slot to install the FC-VI adapter for your platform.



Gateway interconnections

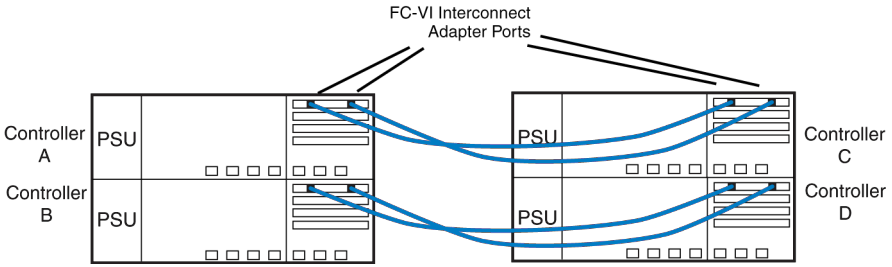
In a stretch MetroCluster configuration, the gateways are connected directly.

The sample illustration shows the connections between a dual-enclosure HA pair for systems that can contain only one controller in the enclosure.

For gateways that are dual-enclosure systems with two controllers in the same chassis, such as a N6000 series system, you install a gateway at each site. To implement the stretch MetroCluster configuration, you must install an FC-VI adapter in each controller to provide the HA interconnect between the systems.

Note: When the FC-VI adapter is installed in the system, the internal InfiniBand interconnect is automatically disabled.

The following illustration shows the FC-VI port connections between HA pair systems such as the N6000 series. Each enclosure contains two controllers. You configure two separate MetroCluster configurations among the four controllers.



In the previous illustration, one MetroCluster configuration is between Controller A and Controller C.

From Controller A	To Controller C
FC-VI port A	FC-VI port A
FC-VI port B	FC-VI port B

In the previous illustration, the other MetroCluster configuration is between Controller B and Controller D.

From Controller B	To Controller D
FC-VI port A	FC-VI port A
FC-VI port B	FC-VI port B

Some older gateways use NVRAM to connect the systems in a MetroCluster configuration. The following table lists the NVRAM connections on gateways that connect to each other through NVRAM.

vs1	**vs2**
NVRAM port L02 Ph2	NVRAM port L02 Ph2
NVRAM port L01 Ph1	NVRAM port L01 Ph1

Inter-Switch Link connections (ISLs)

Stretch MetroCluster configurations use a switched fabric to connect the local half of the configuration to the remote half of the configuration. In the stretch MetroCluster illustration, Switches 1 and 3 are connected to each other (Fabric 1). Switches 2 and 4 are also connected (Fabric 2).

- The first fabric in the MetroCluster configuration begins from Switch 1 on Site A (local) and is completed by connecting the ISL cable to the first switch on Site B (remote)—Switch 3.
- The second fabric is created using Switch 2 on Site A (local), connected through a second ISL cable, to the second switch on Site B (remote)—Switch 4.

ISL Connection	**Switch**	**Fabric**
Port 4 on switch 1	Port 8 on switch 3	Fabric 1
Port 11 on switch 2	Port 15 on switch 4	Fabric 2

Gateway-to-switch connections

The best practice connections in the previous illustration eliminate a single-point-of-failure in the following ways:

- FC initiator ports on the same Fibre Channel controller chip (for example, port 0a and 0b) connect to alternate fabrics.
- Multiple paths and zones ensure that FC initiator ports on the same controller chips access the array LUN from different gateways and switch fabrics.

The following table lists the connections from the gateway to the switch.

Gateway system port	**Switch**
vs1: FC port 0a	Switch 1: Port 1
vs1: FC port 0b	Switch 2: Port 9
vs1: FC port 0c	Switch 1: Port 2
vs1: FC port 0d	Switch 2: Port 10
vs2: FC port 0a	Switch 3: Port 5
vs2: FC port 0b	Switch 4: Port 13
vs2: FC port 0c	Switch 3: Port 6

Gateway system port	Switch
vs2: FC port 0d	Switch 4: Port 14

Storage array-to-switch connections

The best practice connections from the storage array to the switch are as follows:

- Ports 1A and 2A on each storage array connect to alternate fabrics.
- Ports 1B and 2B on each storage array connect to alternate fabrics.
- Gateways are configured to access any array LUN on two storage array paths (1A and 2A or 1B and 2B).

Storage array port	Switch	Fabric
Array 1: Port 1A	Switch 1: Port 3	1
Array 1: Port 2A	Switch 2: Port 12	2
Array 1: Port 1B	Switch 1: Port 5	1
Array 1: Port 2B	Switch 2: Port 13	2
Array 2: Port 1A	Switch 3: Port 7	1
Array 2: Port 2A	Switch 4: Port 16	2
Array 2: Port 1B	Switch 3: Port 9	1
Array 2: Port 2B	Switch 4: Port 17	2

Related concepts

Connecting devices in a MetroCluster configuration with third-party storage on page 107

Cabling guidelines for a MetroCluster configuration with third-party storage

There are a number of cabling guidelines that you need to review before connecting the devices in a MetroCluster configuration with third-party storage.

Cabling guidelines for a dual-enclosure HA pair with multiple controllers in the same enclosure

Some gateways, for example, an N6000 series system, support two controllers in the same enclosure. For a MetroCluster configuration, you must have two such systems. You configure the two systems into a pair of MetroCluster configurations by connecting the FC-VI adapter between the two top controllers and then connecting the FC-VI adapter between the bottom controllers.

In such a configuration, the internal InfiniBand connections between the controllers in the same enclosure are automatically deactivated. Therefore, the two controllers in the enclosure are no longer in an HA pair with each other. Each controller is connected through FC-VI connections to another

controller of the same type, so the four controllers form two independent MetroCluster configurations.

Cabling guidelines for partner-to-partner connections

The cabling for the MetroCluster partners is different on each node of the HA pair. Use the following guidelines to plan your cabling for partner-to-partner connections.

- Each node in the HA pair must have different name than the other so you can distinguish them.
- One node needs to be identified as being in Site A and the other identified as being in Site B.
  For example, the local partner could be Gateway 1, Site A, and the remote partner could be Gateway 2, Site B.
- Each port on the node must be connected to the same fabric.
  For example, if Port A of the FC-VI adapter on the local node is connected to Switch 2 and Port A of the FC-VI adapter on the remote node is connected to Switch 4, then Switch 2 and Switch 4 must be connected by the ISL, thereby connecting the nodes to the same fabric.

Cabling guidelines for gateway-to-switch connections

- Fibre Channel ports on the same channel controller chip cannot be connected to the same switch. One port must be connected to one switch and the other port to the other switch. For example, if onboard port 0a is connected to Switch 3, you cannot connect onboard port 0b to Switch 3; port 0b must be connected to Switch 4.

  Note: Connecting both onboard ports of the pair to the same switch port number can simplify cabling and management of the MetroCluster configuration. For example, if port 0a is connected to Port 1, Switch 1, connect port 0b to Port 1, Switch 2.
- All switches within a fabric must be the same switch model and have the same number of licensed ports.

Cabling guidelines for Inter-Switch connections

You can connect an Inter-Switch Link (ISL) to any available switch port.

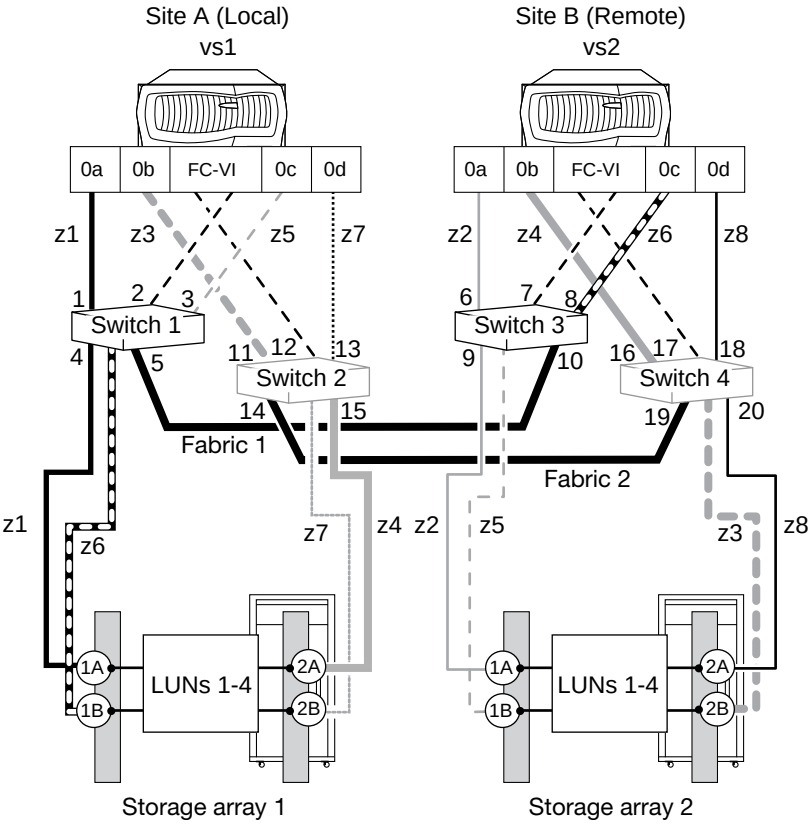
Cabling guidelines for gateway-to-storage array connections

Cabling between the gateways and the storage arrays must be redundant and conform to the required number of paths (two).

Planning zoning for a MetroCluster configuration with third-party storage

Switch zoning defines paths between connected nodes based on the node's unique WWN. Sketching out the zoning plan enables you to correct errors before zones are configured, and makes it easier to communicate the zoning information to the person who configures the switches.

You can use the following example as a reference when determining your zoning plan. The example shows single-initiator zoning for a fabric-attached MetroCluster configuration. The lines in the following example represent zones rather than connections; each line is labeled with its zone number.



In the sample illustration, four array LUNs are allocated on each storage array for the MetroCluster configuration. LUNs of equal size are provisioned on the storage arrays at both sites, which is a SyncMirror requirement. Each FC initiator port on each gateway has a path to each gateway LUN on the storage arrays. The ports on the storage array are redundant, and are configured as follows:

- Storage Array 1
  - Ports 1A and 2A are a redundant port pair.
  - Ports 1B and 2B are a redundant port pair.
  - In each port pair, both ports can access LUNs 1 through 4 because they are alternate paths.
- Storage Array 2
  - Ports 1A and 2A are a redundant port pair.
  - Ports 1B and 2B are a redundant port pair.
  - In each port pair, both ports can access LUNs 1 through 4 because they are alternate paths.

Switches are zoned so that there are only two paths to each array LUN, one unique path from each gateway FC initiator port through each switch. If there are multiple connections between a gateway and the switch, the best practice recommendation is to put each connection into a separate zone.

The following table shows the zones for this example:

Zone	Gateway FC initiator port	Storage array port
Switch 1		
z1	vs1:Port 0a	Storage array 1:Port 1A
z5	vs1:Port 0c	Storage array 2:Port 1B
Switch 2		
z3	vs1:Port 0b	Storage array 2:Port 2B
z7	vs1:Port 0d	Storage array 1:Port 2A
Switch 3		
z2	vs2:Port 0a	Storage array 2:Port 1A
z6	vs2:Port 0c	Storage array 1:Port 1B
Switch 4		
z8	vs2:Port 0d	Storage array 2:Port 2A
z4	vs2:Port 0b	Storage array 1:Port 2B

Related tasks

Configuring zoning in a MetroCluster configuration with third-party storage on page 122
Testing zoning of FC-VI ports in a MetroCluster configuration with third-party storage on page 125

Connecting devices in a MetroCluster configuration with third-party storage

You should plan the cabling required for your MetroCluster configuration before you start connecting the devices. It is helpful to have a port-to-port connectivity diagram to use for reference while you are connecting the devices in the gateway MetroCluster configuration.

Note: Alternative terms for FC-VI adapter are VI-MC adapter and VI-MetroCluster adapter.

Steps

1. *Connecting the local gateways in a MetroCluster configuration* on page 108
2. *Connecting the remote gateways in a MetroCluster configuration* on page 111
3. *Connecting the switch fabric in a MetroCluster configuration with third-party storage* on page 114
4. *Connecting the fabric and storage array in a MetroCluster configuration with third-party storage* on page 116

5. *Setting up a shared-switches configuration* on page 118
6. *Setting preferred primary port in a MetroCluster configuration* on page 121
7. *Removing the preferred primary port in a fabric-attached MetroCluster configuration* on page 122
8. *Configuring zoning in a MetroCluster configuration with third-party storage* on page 122
9. *Changing the configuration speed of a stretch MetroCluster configuration* on page 123

Related concepts

Requirements for a MetroCluster configuration with third-party storage on page 95
Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104
Setting up Data ONTAP after connecting devices in a MetroCluster configuration with third-party storage on page 124
Testing a MetroCluster configuration with third-party storage on page 124

Connecting the local gateways in a MetroCluster configuration

The first stage in connecting the devices in a gateway MetroCluster configuration is to connect a local gateway to the fabric.

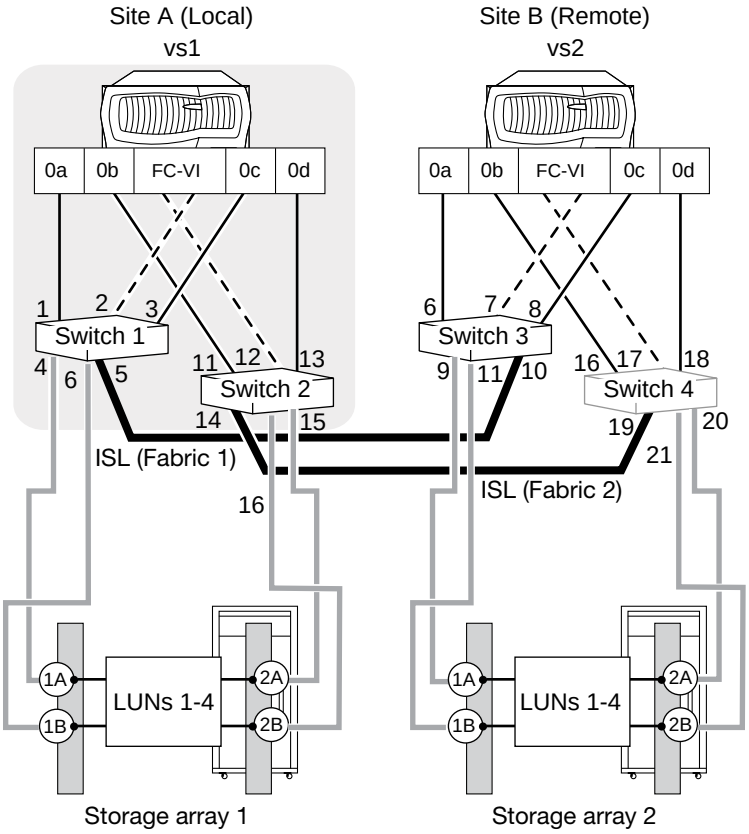
Before you begin

Before you begin this procedure, you should have completed the following:

- Planned for the MetroCluster configuration
- Created storage array LUNs for Data ONTAP and presenting them to Data ONTAP

About this task

Use the following fabric-attached MetroCluster illustration as a reference for connecting the local gateways in a MetroCluster configuration. Differences for stretch MetroCluster configurations are noted in the procedure.



Note: A port pair consists of two FC initiator ports that are used to access the same set of array LUNs. For example, gateway FC initiators 0a and 0d are accessing the same LUNs. See the *Gateway Installation Requirements and Reference Guide* for specific information about port pairs for your platform.

Steps

1. On the local gateway, locate the NVRAM or FC-VI interconnect module.
2. Take one of the following actions to connect the local gateway to the remote gateway.

For a...	Then...
Fabric-attached configuration	**a.** Cable FC-VI Port A to one fabric (Port A to Switch 2 in the previous illustration). **b.** Cable FC-VI Port B to the alternate fabric (Port B to Switch 1 in the previous illustration). **c.** For dual-enclosure systems, repeat substeps a and b for the bottom controllers.

For a...	Then...
Stretch configuration with an FC-VI adapter	Connect the cables between the FC-VI ports directly (point-to-point).
Dual-enclosure HA pair systems with an FC-VI adapter	**a.** Connect port A of the FC-VI adapter on the top controller of the local site (vs1) to port A of the corresponding FC-VI adapter at the remote site (vs2). **b.** Connect port B of the FC-VI adapter on the top controller of the local site (vs1) to port B of the corresponding FC-VI adapter at the remote site (vs2). **c.** Repeat substeps a and b for connecting the FC-VI adapter on the bottom controller.
Stretch configuration with an NVRAM adapter	**a.** Install a copper-to-fiber converter in L01 Ph1, and cable L01 Ph1 on the local node to L01 Ph1 on the remote node. **b.** Install a copper-to-fiber converter in L02 Ph2, and cable L02 Ph2 on the local node to L02 Ph2 on the remote node.

3. Connect the gateway FC initiator ports to the switched fabric:

   a) Identify one of the two FC initiator port pairs on the gateway.
   b) Cable one port of the FC initiator port pair to one fabric.
   c) Cable the other port of the pair to the alternate fabric.

      In the sample illustration, these are ports 0a and 0b.
   d) Cable one port of the FC initiator port pair to one fabric, then cable the other port of the pair to the alternate fabric.
   e) Identify the other onboard FC initiator port pair on the gateway.
   f) Cable one port of the pair to one fabric.
   g) Cable another port of the pair to the alternate fabric.

      In the sample illustration, these are ports 0c and 0d.

4. Optional: Connect the gateway to a tape backup device through a separate FC initiator port or SCSI tape adapter.

5. Connect a console cable to the console port on the gateway, then connect the console cable to the adapter.

   Use the RJ-45 to DB-9 adapter that is included with your system.

6. Install the cable management tray:

   a) Pinch the arms of the tray and fit the holes in the arms through the motherboard tray pins.
   b) Push the cables into the cable holders, thread the adapter cables through the top rows of the cable holders, and then thread the port cables through the lower cable holders.

7. Connect the gateway to the Ethernet network by plugging the network cable into the networking port.

If you are connecting more than one network cable to the network, connect to the ports sequentially. Use the cable management tray to direct all the cabling from your system.

8. Optional: Connect the remote management device from the back of the gateway to the network using an Ethernet cable.

   The network switch port for the remote management device connection must negotiate down to 10/100 or autonegotiate.

9. If applicable, turn on any tape backup devices.

10. For each power supply on the gateway, take the following steps:

    a) Ensure that the power switch is in the Off (0) position.
    b) Connect the socket end of the power cord to the power plug on the power supply.
    c) Secure the power cord with the retaining adjustable clip on the power supply.
    d) Plug the other end of the power cord into a grounded electrical outlet.

       Note: To obtain power supply redundancy, you must connect the second power supply to a separate AC circuit.

11. Start a communications program.

    You must use some form of communications program to perform initial network setup and gateway configuration. You can start a communications program through the remote management device or through the console after connecting to the serial port.

After you finish

Next connect the remote gateway to the fabric.

Related concepts

Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104

Related tasks

Connecting the remote gateways in a MetroCluster configuration on page 111

Connecting the remote gateways in a MetroCluster configuration

After connecting the local gateways to the fabric, you need to connect the remote systems to the fabric.

Before you begin

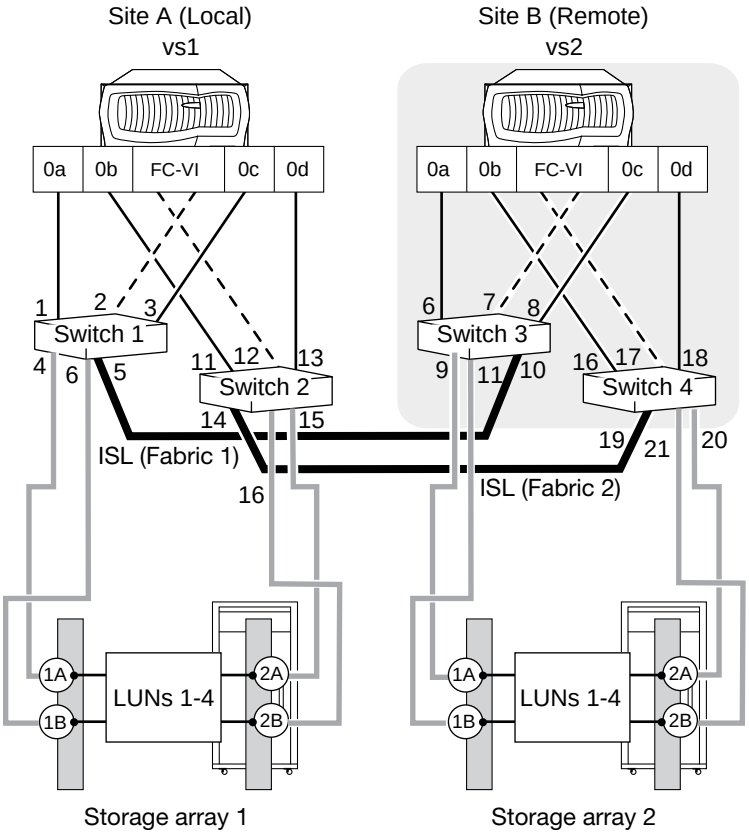
Before you begin this procedure you should have completed the following:

- Connected the local gateways to the fabric

About this task

Use the following illustration of a fabric-attached MetroCluster configuration as a reference for connecting the remote gateways. Differences for stretch MetroCluster configurations are noted in the procedure.

Note: A port pair consists of two FC initiator ports that are used to access the same set of array LUNs. For example, gateway FC initiators 0a and 0d are accessing the same LUNs. See the *Gateway Installation Requirements and Reference Guide* for specific information about port pairs.



Steps

1. Connect the local gateway to the remote gateway.

For a...	Then...
Fabric-attached configuration	**a.** Cable FC-VI Port A to one fabric (Port A to Switch 4 in the sample illustration). **b.** Cable FC-VI Port B to the alternate fabric (Port B to Switch 3 in the sample illustration).

For a...	Then...
Stretch configuration with an FC-VI card	This is a point-to-point connection.

2. Connect the gateway FC initiator ports to the switched fabric:
   a) Identify one of the two FC initiator port pairs on the gateway.
   b) Cable one port of the FC initiator port pair to one fabric.
   c) Cable the other port of the pair to the alternate fabric.

      In the sample illustration, these are ports 0a and 0b.
   d) Identify the other FC initiator port pair on the gateway.
   e) Cable one port of the pair to one fabric.
   f) Cable another port of the pair to the alternate fabric.

      In the sample illustration, these are ports 0c and 0d.
3. Optional: Connect the gateway to a tape backup device through a separate FC initiator port or SCSI tape adapter.
4. Connect a console cable to the console port on the gateway.

   Use the RJ-45 to DB-9 adapter that is included with your system. Connect the console cable to the adapter.
5. Install the cable management tray.
   a) Pinch the arms of the tray and fit the holes in the arms through the motherboard tray pins.
   b) Push the cables into the cable holders, thread the adapter cables through the top rows of the cable holders, and then thread the port cables through the lower cable holders.
6. Connect the gateway to the Ethernet network by plugging the network cable into the networking port.

   If you are connecting more than one network cable to the network, connect to the ports sequentially. Use the cable management tray to direct all the cabling from your system.
7. Optional: Connect the remote management device from the back of the gateway to the network using an Ethernet cable.

   The network switch port for the remote management device connection must negotiate down to 10/100 or autonegotiate.
8. If applicable, turn on any tape backup devices.
9. For each power supply on the gateway, take the following steps:
   a) Ensure that the power switch is in the Off (0) position.
   b) Connect the socket end of the power cord to the power plug on the power supply.
   c) Secure the power cord with the retaining adjustable clip on the power supply.
   d) Plug the other end of the power cord into a grounded electrical outlet.

Note: To obtain power supply redundancy, you must connect the second power supply to a separate AC circuit.

10. Start a communications program.

You must use some form of communications program to perform initial network setup and gateway configuration. You can start a communications program through the remote management device or through the console after connecting to the serial port.

After you finish

Next connect the switch fabric in the MetroCluster configuration.

Related concepts

Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104

Related tasks

Connecting the switch fabric in a MetroCluster configuration with third-party storage on page 114

Connecting the switch fabric in a MetroCluster configuration with third-party storage

Connecting the switch fabric involves connecting the ISL cables and applying power to the switches.

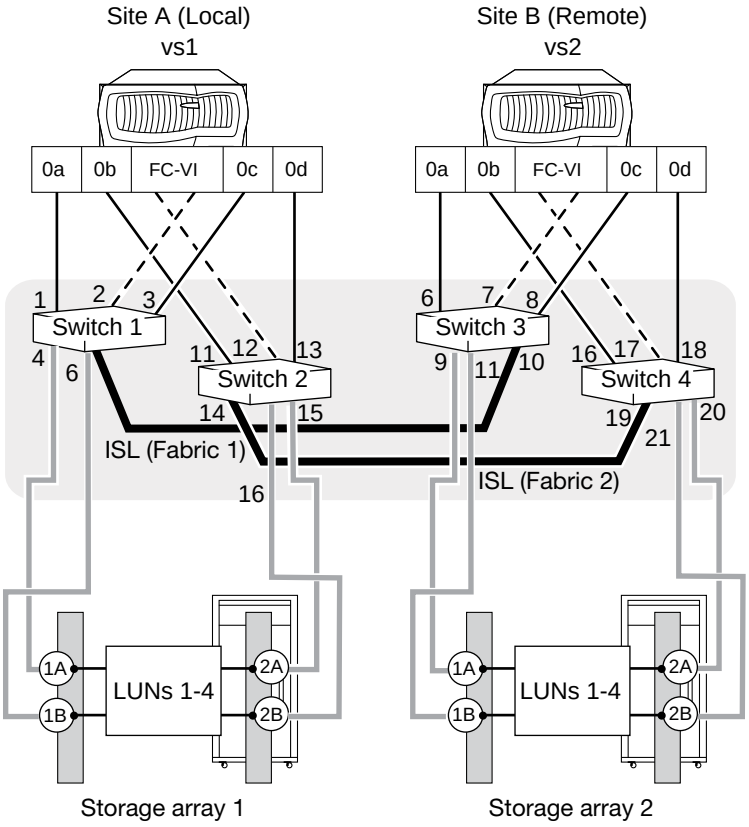
Before you begin

Before you begin this procedure you should have completed the following:

- Connected the local and remote gateways to the fabric.

About this task

The following illustration provides a reference for connecting the switch fabric in this procedure.



Steps

1. Connect the switched fabric:
   a) If you are configuring a new switch fabric, set the domain ID to a unique domain ID for each switch in a fabric.

      See your switch documentation for more details.
   b) Connect an ISL cable to a switch on one fabric and to another switch on the same fabric.

      In the sample illustration, fabric 1, Switch 1, Port 5 connects to Fabric 1, Switch 3, Port 10.
   c) Connect an ISL cable on a switch on the alternate fabric to another switch on the alternate fabric.

      In the sample illustration, Fabric 2, Switch 2, Port 14 connects to Fabric 2, Switch 4, Port 19.

   Note: You must install a long distance SFP adapter in each port that you use to connect an ISL cable. You might also need to install an additional switch license to provide ISL support.
2. Make sure that all switch IDs are set, then turn on each switch 10 minutes apart from one another.

After you finish

Next connect the fabric and the storage arrays.

Related concepts

Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104

Related tasks

Connecting the fabric and storage array in a MetroCluster configuration with third-party storage on page 116

Connecting the fabric and storage array in a MetroCluster configuration with third-party storage

All storage arrays, regardless of model, must be configured to allow Data ONTAP to access a specific LUN on two (primary and secondary) storage array ports.

Before you begin

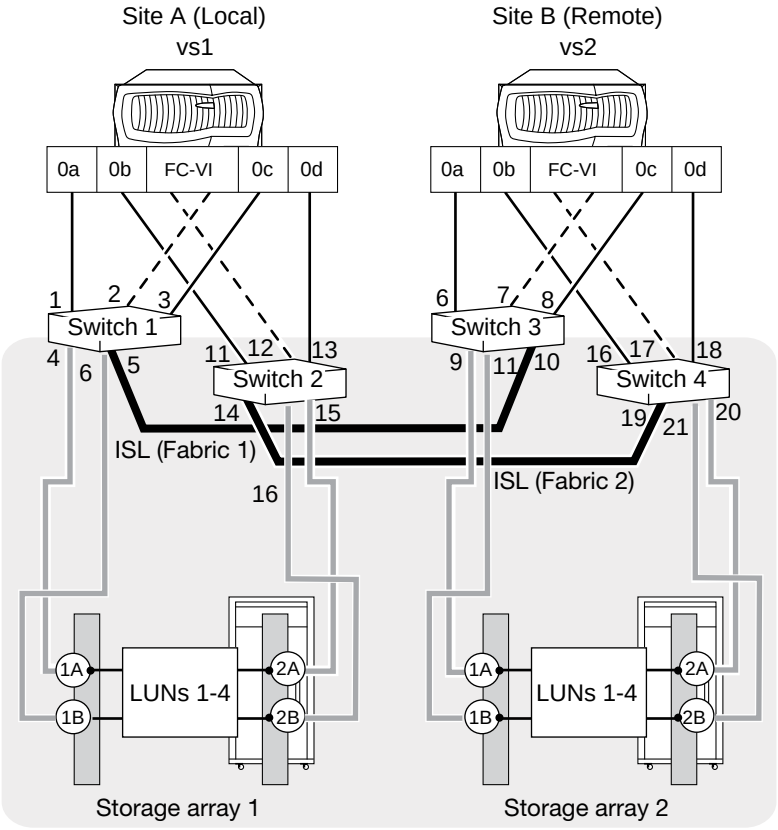
Before you begin this procedure you should have completed the following:

- Connected the local and remote gateways to the fabric.
- Connected the switch fabric and powering up the switches.

Also, check the documentation for your storage array to verify that the ports you plan to use do not access more than the number of array LUNs and host groups that are supported for that storage array model.

About this task

The following illustration provides a reference for connecting the switched fabric and the storage arrays.



Steps

1. Connect the ports on the storage array at Site A.
   a) Connect storage array controller 1A to any port on one fabric.

      In the sample illustration, this is Switch 1, Port 4, Fabric 1.
   b) Connect controller 2A to any port on the alternate fabric.

      In the sample illustration this is Switch 2, Port 15, Fabric 2.
   c) Connect storage array controller 1B to any port on one fabric.

      In the sample illustration, this is Switch 1, Port 6, Fabric 1.
   d) Connect controller 2B to any port on the alternate fabric.

      In the sample illustration this is Switch 2, Port 16, Fabric 2.
   e) Connect additional controller ports and fabric, as required by your MetroCluster configuration.
2. Connect the ports on the storage array at Site B:
   a) Connect storage array controller 1A to any port on one fabric.

In the sample illustration, this is Switch 3, Port 9, Fabric 1.

b) Connect controller 2A to any port on the alternate fabric.

In the sample illustration, this is Switch 4, Port 20, Fabric 2.

c) Connect additional controller ports and fabric, as required by your MetroCluster configuration.

d) Connect storage array controller 1B to any port on one fabric.

In the sample illustration, this is Switch 3, Port 11, Fabric 1.

e) Connect controller 2B to any port on the alternate fabric.

In the sample illustration, this is Switch 4, Port 21, Fabric 2.

After you finish

Next configure zoning.

Related concepts

Recommended fabric-attached MetroCluster configuration with third-party storage on page 97
Recommended stretch MetroCluster configuration with third-party storage on page 101
Cabling guidelines for a MetroCluster configuration with third-party storage on page 104

Related tasks

Configuring zoning in a MetroCluster configuration with third-party storage on page 122

Setting up a shared-switches configuration

In a shared-switches configuration, two fabric-attached MetroCluster configurations share the same four switches and the ISLs between them.

Shared-switches configuration are cost-effective because you can use the four Brocade switches between two MetroCluster configurations.

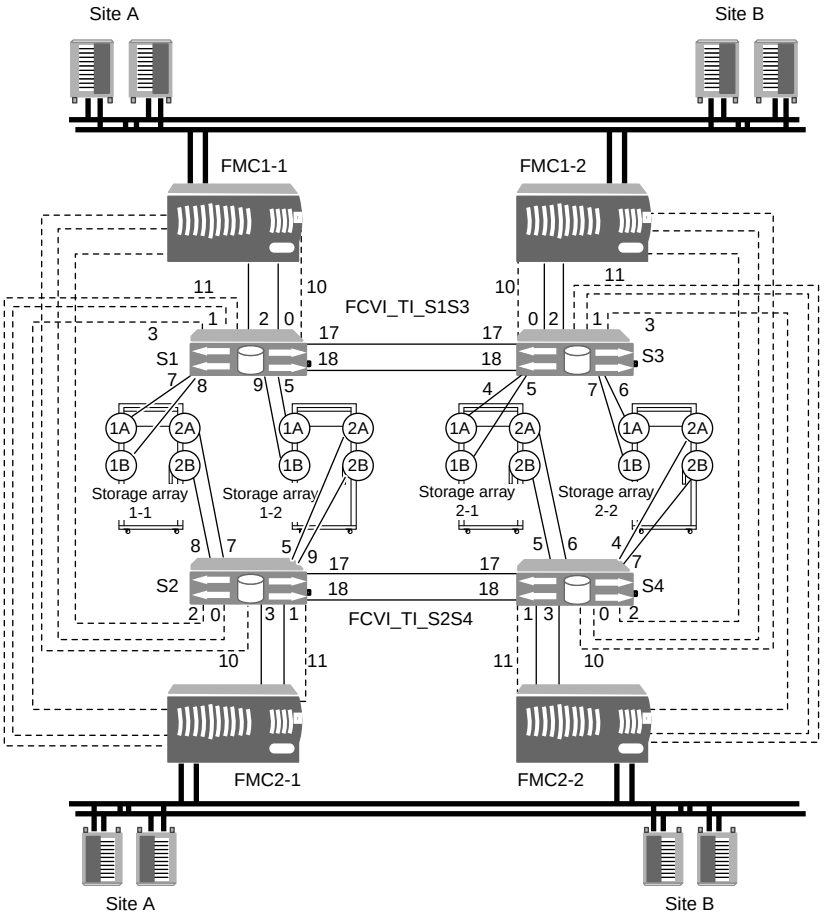
Cabling the FC-VI adapter and ISL in a shared-switches configuration with third-party storage

Cabling in a shared-switches MetroCluster configuration with third-party storage involves cabling the FC-VI and HBA ports to the switches, and cabling the switches to the storage arrays.

About this task

In a shared-switches configuration, four nodes join together to form two fabric-attached MetroCluster configurations. In the illustration, the MetroCluster configurations are as follows:

- FMC1-1 and FMC1-2 form one fabric-attached MetroCluster configuration named FMC1.
- FMC2-1 and FMC2-2 form another fabric-attached MetroCluster configuration named FMC2.



Steps

1. Connect the ports of the FC-VI adapter to primary and secondary switches for both gateways in the MetroCluster configuration.

   The FC-VI adapter connections for the sample configuration are shown in the following table.

Gateway FC-VI card...	**Connects to these switches...**
FMC 1-1	Primary switch S1 and secondary switch S2 through port 0
FMC 1-2	Primary switch S3 and secondary switch S4 through port 1
FMC 2-1	Primary switch S2 and secondary switch S1 through port 1

Gateway FC-VI card...	Connects to these switches...
FMC 2-2	Primary switch S3 and secondary switch S4 through port 1

2. Connect the gateway FC initiator ports to the primary and secondary switches.

   The FC initiator port connections in the illustration are shown in the following table.

gateway FC initiator ports	Connect to these switches...
FMC 1-1	Primary switch S1 and secondary switch S2 through port 2 and port 10
FMC 1-2	Primary switch S3 and secondary switch S4 through port 2 and port 10
FMC 2-1	Primary switch S2 and secondary switch S1 through port 3 and port 11
FMC 2-2	Primary switch S3 and secondary switch S4 through port 3 and port 11

3. Connect the storage arrays to the switches.

   The following table shows the switch-to-storage-array connections for FMC1.

Storage array and port	Switch and port
1-1: Port 1A	Switch 1: Port 7
1-1: Port 1B	Switch 1: Port 8
1-1: Port 2A	Switch 2: Port 7
1-1: Port 2B	Switch 2: Port 8
1-2: Port 1A	Switch 1: Port 5
1-2: Port 1B	Switch 1: Port 9
1-2: Port 2A	Switch 2: Port 5
1-2: Port 2B	Switch 2: Port 9

4. Repeat steps 1 through 3 at another site.

   The following table shows the switch-to-storage-array connections for FMC2.

Storage array and port	Switch and port
2-1: Port 1A	Switch 3: Port 4
2-1: Port 1B	Switch 3: Port 5

Storage array and port	Switch and port
2-1: Port 2A	Switch 4: Port 6
2-1: Port 2B	Switch 4: Port 5
2-2: Port 1A	Switch 3: Port 6
2-2: Port 1B	Switch 3: Port 7
2-2: Port 2A	Switch 4: Port 4
2-2: Port 2B	Switch 4: Port 7

5. When cabling at both the sites is complete, connect an inter-switch link cable to a port on each switch.

   In the illustration, the fabric containing switches S1 and S3 has two ISLs and the fabric containing switches S2 and S4 also has two ISLs.

   Switch S1 connects to switch S3 and switch S2 connects to switch S4 through ports 17 and 18.

After you finish

After you finish cabling the devices, you need to configure traffic isolation on the switches. This configuration is required when you are using dual inter-switch links. See the *Fabric-attached MetroCluster Brocade Switch Configuration Guide* or the *Fabric-attached MetroCluster Cisco Switch Configuration Guide* for more information.

Related tasks

Setting preferred primary port in a MetroCluster configuration on page 88

Setting preferred primary port in a MetroCluster configuration

After cabling a MetroCluster configuration, you can define the primary port between the four switches to carry the traffic. You can set the preferred primary port in a shared-switches configuration by using the `ic primary set` command.

About this task

By default, the primary port is 0.

Steps

1. To set the primary port, enter the following command:

   `ic primary set 0|1 [-r]`

   You must use the `-r` option for the primary port to take effect immediately.

   Note: If the FC-VI link between the preferred primary ports fails, the fabric-attached MetroCluster configuration fails over to the secondary FC-VI link. When the primary link is

restored, the fabric-attached MetroCluster configuration again uses the primary link. The failover temporarily causes the logs to be unsynchronized.

Example

If you want to set the preferred primary port to 1, enter the following command:

```
ic primary set 1 -r
```

In a shared-switches configuration, if you have set the preferred primary port of FMC1 (FMC1-1 and FMC1-2) to 1, the preferred primary port for FMC2 (FMC2-1 and FMC2-2) is set to 0.

2. To view the current primary port, enter the following command:

```
ic primary show
```

It displays the primary port.

Removing the preferred primary port in a fabric-attached MetroCluster configuration

You can remove the port that you assigned as primary port in a fabric-attached MetroCluster configuration.

Step

1. To remove the primary port, enter the following command:

```
ic primary unset
```

Configuring zoning in a MetroCluster configuration with third-party storage

Switch zoning defines paths between connected nodes based on the node's unique WWN. Single-initiator zoning is recommended for a gateway configuration.

Before you begin

Before you begin this procedure you should have completed the following:

- Determined zoning for each switch.
- Connected the local and remote gateways to the fabric.
- Connected the switch fabric and powering up the switches.
- Connected the fabric and storage array.

About this task

For a MetroCluster configuration with third-party storage, the FC-VI ports on the FC adapters must be zoned end-to-end across the fabric.

Steps

1. Gather the WWPNs for the FC-VI ports on the gateway FC-VI card.

The WWPNs for the FC-VI ports are needed when you set up the zones for the FC-VI ports. The WWPNs for the FC-VI ports are available from the following sources:

- The switch
- Data ONTAP `sysconfig -M` output
  Gather WWPNs from entries such as the following: `!Qlogic 2352 FCVI Cluster Interconnect Adapter<adapter_WWPN>`

2. Zone FC-VI port “a” on the local gateway to the FC-VI port “a” on the remote gateway.
3. Zone FC-VI port “b” on the local gateway to the FC-VI port “b” on the remote gateway.

After you finish

- If necessary, change the configuration speed of your stretch MetroCluster configuration.
- Set up Data ONTAP on the gateways so that they can access storage on the storage arrays, and so you can take advantage of Data ONTAP features.

Related concepts

Planning zoning for a MetroCluster configuration with third-party storage on page 105

Related tasks

Testing zoning of FC-VI ports in a MetroCluster configuration with third-party storage on page 125

Changing the configuration speed of a stretch MetroCluster configuration

If the distance between nodes in a stretch MetroCluster configuration is greater than the supported default configuration speed, you must change the default configuration speed. If you modified the default configuration speed in a stretch MetroCluster configuration that uses an FC-VI adapter, you can reset the speed to the default configuration speed.

The default maximum speed between nodes in a stretch MetroCluster configuration depends on the speed at which the FC-VI adapter operates (for example, 4G or 8G). The maximum supported distance is included in the *Gateway Interoperability Matrix*.

The procedures for changing the configuration speed of a stretch MetroCluster configuration are the same for MetroCluster configurations with filers and MetroCluster configurations with gateways.

Related tasks

Changing the default configuration speed of a stretch MetroCluster configuration on page 67
Resetting a stretch MetroCluster configuration to the default speed on page 69

Setting up Data ONTAP after connecting devices in a MetroCluster configuration with third-party storage

After connecting the devices in the MetroCluster configuration, you need to set up the gateways to use the storage on the storage array. You also need to set up any desired Data ONTAP features.

Perform the following tasks after connecting the devices in your MetroCluster configuration with gateways.

1. Set up the gateways as described in the *Data ONTAP Software Setup Guide for 7-Mode* for gateways using only third-party storage.

   Note: You should install the SyncMirror license before you assign array LUNs to your gateway. Otherwise, Data ONTAP specifies the local pool (pool0) for the all the array LUNs and you have to unassign the LUNs for the remote location after the SyncMirror license is installed and reassign them with a pool parameter of pool1.

2. Create one or more mirrored aggregates.
   See the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.
3. Test the MetroCluster configuration, as described in the *Data ONTAP High Availability and MetroCluster Configuration Guide for 7-Mode*.
4. Verify network and protocol setup, as described in the *Data ONTAP File Access and Protocols Management Guide for 7-Mode*.
5. Configure other Data ONTAP features as desired, for example, data protection features.
   See the relevant Data ONTAP guide for the feature you want to set up.

Testing a MetroCluster configuration with third-party storage

It is important to test that your MetroCluster configuration is properly set up before putting it into a production environment.

Steps

1. *Testing zoning of FC-VI ports in a MetroCluster configuration with third-party storage* on page 125
2. *Verifying proper setup at MetroCluster sites* on page 125
3. *Simulating a disaster recovery in a MetroCluster configuration* on page 126

Testing zoning of FC-VI ports in a MetroCluster configuration with third-party storage

To test zoning of the FC-VI ports, you need to check that you have not crossed the FC-VI “a” ports and “b” ports in your zones.

Related tasks

Configuring zoning in a MetroCluster configuration with third-party storage on page 122

Verifying proper setup at MetroCluster sites

After installing a MetroCluster configuration with third-party storage, you need to test the paths, FC ports, and switch configuration at each MetroCluster site to ensure that they are set up correctly.

Steps

1. Enter the following command to display path information from each gateway to the array LUNs:

   `storage show disk -p`

   You should see two paths to each array LUN.

2. To test FC initiator ports, enter the following commands for each FC initiator port in the MetroCluster configuration:

   a) Enter the following command:

      `priv set advanced`

   b) Enter the following command to display the state of the HBA port:

      `fcadmin link_state`

   c) Enter the following command to take an FC port offline to simulate a port failure or a cable pull:

      `fcadmin offline` ***`portname`***

      Example

      `fcadmin offline 0a`

   d) Enter the following command to display disk path information:

      `storage show disk -p`

      The display should show only one path.

   e) Enter the following command to bring the FC port online:

      `fcadmin online` ***`portname`***

      Example

      `fcadmin online 0a`

   f) Enter the following command to verify that both paths are online:

```
storage show disk -p
```

3. Simulate a switch failure or a storage array controller failure for each switch fabric.
   a) Use the following commands to take all FC ports offline on both gateways that are attached to one fabric:

      ```
      fcadmin offline portname
      ```

      Example

      ```
      fcadmin offline 0a
      fcadmin offline 0b
      fcadmin offline 0c
      fcadmin offline 0d
      ```

      If your gateway has more than four FC initiator ports, take the remaining initiator ports offline.
   b) Enter the following command to verify that all HBAs are disabled:

      ```
      fcadmin link_state
      ```
   c) Enter the following command to initiate a takeover from site A:

      ```
      cf takeover
      ```
   d) When site B is in takeover mode, enter the following command to initiate a giveback from Site A:

      ```
      cf giveback
      ```

Simulating a disaster recovery in a MetroCluster configuration

Testing for proper setup of a MetroCluster configuration with third-party storage includes simulating a disaster recovery. Simulating a site failure and recovery involves disabling and degrading mirrors.

About this task

You should never simulate disaster site failure and recovery in production environments without prior planning and downtime.

Steps

1. Disable the HA interconnect between the gateways.
2. Power down one of the gateways.

   The system you power down simulates the site of the disaster.
3. On the surviving system, complete the following steps:
   a) Enter the following command to activate a forced, manual takeover:

      ```
      cf forcetakeover -d
      ```
   b) Enter the following command to validate that the forced takeover has occurred:

```
cf status
```

c) Enter the following commands to validate the status of the aggregates and of the volumes that they include:

```
aggr status
```

```
vol status
```

The display should show an online status for aggregates and volumes, but mirrors, which are disabled, should be displayed as “degraded.”

4. Reconnect the HA interconnect.
5. On the surviving system, enter the following commands:

a) Enter the following command to rejoin the aggregates:

```
aggr mirror surviving_aggregate -v victim_aggregate
```

Note: Data ONTAP uses parentheses to indicate the degraded aggregate. For example, aggr0_b indicates the aggregate that is not degraded, and aggr0_b(1) is degraded.

b) Enter the following command to validate that forced takeover has occurred:

```
cf status
```

c) Enter the following command to validate aggregate status:

```
aggr status
```

The display should show an online and mirrored status for aggregates.

d) Enter the following command to validate volume status:

```
vol status
```

The display should show an online and mirrored status for volumes.

6. Power up the gateway that you powered down to simulate the disaster.
7. On the surviving system, enter the following command to reactivate the site that simulated the disaster:

```
cf giveback
```

8. To validate that the HA pair, aggregates, and mirrors are online and operational, enter the following commands at both sites:

```
cf status
```

```
aggr status
```

```
vol status
```

The display should indicate that the HA pair is enabled, and that aggregates and volumes are online and mirrored.

Reconfiguring an HA pair into two stand-alone systems

To divide an HA pair so that the nodes become stand-alone systems without redundancy, you must disable the HA software features and then remove the hardware connections.

About this task

This procedure applies to all HA pairs regardless of disk shelf type.

Steps

Ensuring uniform disk ownership within disk shelves and loops in the system

If a disk shelf or loop contains a mix of disks owned by Node A *and* Node B, you must use this procedure to move the data and make disk ownership uniform within the disk shelf or loop.

Before you begin

You must ensure the following:

- Disk ownership is uniform within all disk shelves and loops in the system
- All the disks within a disk shelf or loop belong to a single node and pool

About this task

Note: It is a best practice to always assign all disks on the same loop to the same node and pool.

Steps

1. Use the following command to identify any disk shelves or loops that contain both disks belonging to Node A and disks belonging to Node B:

   ```
   disk show -v
   ```

2. Determine which node the disk shelf or loop with mixed ownership will be attached to when the HA feature is unconfigured and record this information.

For example, if the majority of the disks in the loop belong to Node A, you probably want the entire loop to belong to stand-alone Node A.

After you finish

Proceed to disable the HA software.

Disabling controller failover

You need to disable the controller failover functionality in the software before reconfiguring the hardware to create two separate stand-alone storage systems.

Before you begin

Before performing this procedure you must ensure that all loops and disk shelves in the system contain disks belonging to one or the other nodes. The disk shelves and loops cannot contain a mix of disks belonging to Node A and Node B. In any disk shelves or loops containing such a mix of disks, you must move data.

Steps

1. Enter the following command on either node console:

   `cf disable`

2. Disable the cf license by entering the following command:

   `license delete cf`

3. Open the `/etc/rc` file with a text editor and remove references to the partner node in the `ifconfig` entries, as shown in the following example:

   #### Example

   Original entry:

   `ifconfig e0 199.9.204.254 partner 199.9.204.255`

   Edited entry:

   `ifconfig e0 199.9.204.254`

4. Repeat Step 1 through Step 3 on the partner node.

After you finish

Proceed to reconfigure the hardware.

Related concepts

Requirements when changing a node using array LUNs to stand-alone on page 132

Related tasks

Reconfiguring nodes using array LUNs for stand-alone operation on page 133

Reconfiguring nodes using disk shelves for stand-alone operation

You can reconfigure the hardware if you want to return to a single-controller configuration.

Before you begin

The HA pair software must be disabled.

Steps

1. Halt both nodes by entering the following command on each console:

   halt

2. Using the information you recorded earlier in the disk shelves or loops with mixed storage, physically move the disks to a disk shelf in a loop belonging to the node that owns the disks.

   For example, if the disk is owned by Node B, move it to a disk shelf in a loop that is owned by Node B.

   Note: Alternatively, you can move the data on the disks using a product such as Snapshot software, rather than physically moving the disk. See the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.

   After moving the data from the disk you can zero the disk and use the `disk remove_ownership` command to erase the ownership information from the disk. See the *Data ONTAP Storage Management Guide for 7-Mode*.

3. If you are completely removing one node so that all the disk shelves will belong to a single stand-alone node, complete the following substeps:

   a) Boot the node being removed into Maintenance mode, as described in the *Data ONTAP System Administration Guide for 7-Mode*.

   b) Enter the following command to reassign all disk shelves so that they all belong to the remaining node:

      disk reassign [-o *old_name* | -s *old_sysid*] [-n *new_name*] -d *new_sysid*

   c) Halt the node by entering the following command:

      halt

4. Turn off the power to each node, then turn off the power to the disk shelves and unplug them from the power source.

5. Ground yourself, then remove the HA interconnect cables from both nodes.

See the hardware documentation for your system for details.

6. Move or remove the adapter used for the HA interconnect:

If your system uses...	Then...
An HA interconnect adapter or an FC-VI adapter	Remove the adapter from the system.
An NVRAM5 or NVRAM6 adapter	You might need to change the slot position of the adapter. See the *N series Introduction and Planning Guide* for details about expansion slot usage for the adapter.

7. Recable the system, depending on the type of system:

If you are converting a...	Then...
System with nonmirrored disks	**a.** Disconnect all cabling from the Channel B loop on the local node. **b.** Repeat for the partner node.
System with mirrored disks or a redundant Channel B loop	**a.** Connect the local node to the open Channel B loop in its local disk shelves, as described in the appropriate disk shelf guide. **b.** Repeat for the partner node.

8. Power on the disk shelves, then the individual nodes, monitoring the system console for error messages during the boot process.
9. Run all system diagnostics at the boot prompt by entering the following command on the system console:

   ```
   boot_diags
   ```

10. Unset the partner system ID by entering the following command at the prompt:

    ```
    unsetenv partner-sysid
    ```

11. Boot the node by entering the following command:

    ```
    boot_ontap
    ```

12. Check HA pair status by entering the following command:

    ```
    cf status
    ```

    If the HA pair is disabled, you see the following output:
    ```
    Controller Failover disabled.
    ```

13. Repeat Step 1 through Step 12 for the partner node.

Related concepts

Requirements when changing a node using array LUNs to stand-alone on page 132

Related tasks

Reconfiguring nodes using array LUNs for stand-alone operation on page 133

Requirements when changing a node using array LUNs to stand-alone

After uncoupling gateways in an HA pair, you might need to perform additional reconfiguration related to Data ONTAP ownership of array LUNs.

The following table summarizes the requirements when uncoupling a storage system using array LUNs from an HA pair.

If you want to...	Requirements for uncoupling systems are...	Requirements for array LUN assignments to systems are...
Make both systems in the pair stand-alone systems	Remove the HA configuration software and interconnect cabling	No Data ONTAP reconfiguration of array LUNs is necessary. Each system can continue to own the array LUNs assigned to it.
Remove one system in the pair from service	Remove the HA configuration software and interconnect cabling	After uncoupling the pair, you must do one of the following: • If you want to continue to use the array LUNs for Data ONTAP, reassign the array LUNs owned by the system you are removing to another storage system. • Prepare the array LUNs assigned to the system you are removing for use by systems that do not run Data ONTAP.

Related tasks

Disabling controller failover on page 129
Reconfiguring nodes using array LUNs for stand-alone operation on page 133

Reconfiguring nodes using array LUNs for stand-alone operation

After uncoupling the nodes in an HA pair, each node can continue to own its assigned array LUNs, you can reassign its array LUNs to another gateway, or you can release the persistent reservations on the array LUNs so the LUNs can be used by a non Data ONTAP host.

Before you begin

The HA pair software must be disabled.

About this task

If you want both nodes in the HA pair to remain in service and operate as stand-alone systems, each system can continue to own the array LUNs that were assigned to it. Each system, as a stand-alone, continues to see the array LUNs owned by the other system because both systems are still part of the same gateway neighborhood. However, only the system that is the owner of the array LUNs can read from or write to the array LUN, and the systems can no longer fail over to each other.

Steps

1. On each node, halt the node by entering the following command at the console:

   ```
   halt
   ```

2. Turn off the power to each node.
3. Ground yourself, then remove the HA interconnect cables from both nodes. See the hardware documentation for your system for details.
4. Move or remove the adapter used for the HA interconnect.

If your system uses...	Then...
An HA interconnect adapter or an FC-VI adapter	Remove the adapter from the system.
An NVRAM5 or NVRAM6 adapter	You might need to change the slot position of the adapter. See the *N series Introduction and Planning Guide* for details about expansion slot usage for the adapter.

5. On each node, perform the following steps:
   a) Power on the node, monitoring the system console for error messages during the boot process.
   b) Unset the partner system ID by entering the following command at the prompt:

      ```
      unsetenv partner-sysid
      ```

6. Perform the appropriate step in the following table for what you intend to do with your system and its storage:

If you want to...	Then...
Keep both systems in service as stand-alone systems and continue with both systems owning the array LUNs that were already assigned to them	Boot both systems by entering the following command on each system: **`boot_ontap`**
Remove one of the systems from service but still use the storage that was assigned to that system for Data ONTAP	**a.** Boot the node being removed into Maintenance mode, as described in the *Data ONTAP System Administration Guide for 7-Mode*. **b.** Use the `disk reassign` command to reassign all the array LUNs so that they all belong to the node that remains. The `disk reassign` command has the following syntax: **`disk reassign [-o <old_name> | -s <old_sysid>] [-n <new_name>] -d <new_sysid>`** **c.** Remove the node from service. **d.** Boot the node you are keeping in service by entering the following command: **`boot_ontap`**
Remove one of the systems from service and use the array LUNs that are currently assigned to it for a host that does not run Data ONTAP	Release the persistent reservations that Data ONTAP placed on those array LUNs so that the storage administrator can use those LUNs for other hosts. See the *Data ONTAP Storage Management Guide for 7-Mode* for information about what you need to do to prepare for taking a system using array LUNs out of service.

Related tasks

Disabling controller failover on page 129

Configuring an HA pair

Bringing up and configuring a standard or mirrored HA pair for the first time can require enabling licenses, setting options, configuring networking, and testing the configuration.

These tasks apply to all HA pairs regardless of disk shelf type.

Steps

1. *Bringing up the HA pair* on page 135
2. *Enabling licenses* on page 138
3. *Setting options and parameters* on page 139
4. *Configuring network interfaces for HA pairs* on page 146
5. *Testing takeover and giveback* on page 156

Bringing up the HA pair

The first time you bring up the HA pair, you must ensure that the nodes are correctly connected and powered up, and then use the setup program to configure the systems.

Considerations for HA pair setup

When the setup program runs on a storage system in an HA pair, it prompts you to answer some questions specific for HA pairs.

The following list outlines some of the questions about your installation that you should think about before proceeding through the setup program:

- Do you want to configure interface groups for your network interfaces?
  For information about interface groups, see the *Data ONTAP Network Management Guide for 7-Mode*.

  Note: You are advised to use interface groups with HA pairs to reduce SPOFs (single-points-of-failure).
- How do you want to configure your interfaces for takeover?

Note: If you do not want to configure your network for use in an HA pair when you run setup for the first time, you can configure it later. You can do so either by running setup again, or by using the `ifconfig` command and editing the `/etc/rc` file manually. However, you must provide at least one local IP address to exit setup.

Related tasks

Configuring shared interfaces with setup on page 136
Configuring dedicated interfaces with setup on page 137

Configuring standby interfaces with setup on page 137

Configuring shared interfaces with setup

During setup of the storage system, you can assign an IP address to a network interface and assign a partner IP address that the interface takes over if a failover occurs.

Steps

1. Enter the IP address for the interface that you are configuring.

   Example

   ```
   Please enter the IP address for Network Interface e0 []:nnn.nn.nn.nnn
   ```

   nnn.nn.nn.nnn is the local address for the node you are configuring.

2. Enter the netmask for the interface you are configuring, or press Return if the default value is correct.

   Example

   ```
   Please enter the netmask for Network Interface e1 [255.255.0.0]:
   ```

3. Specify that this interface is to take over a partner IP address.

   Example

   ```
   Should interface e1 take over a partner IP address during failover?
   [n]: y
   ```

4. Enter the IP address or interface name of the partner.

   Example

   ```
   Please enter the IP address or interface name to be taken over by e1
   []: :nnn.nn.nn.nnn
   ```

   Note: If the partner is a interface group, you must use the interface name.

Configuring dedicated interfaces with setup

You can assign a dedicated IP address to a network interface, so that the interface does not have a partner IP address.

About this task

This procedure is performed during setup of the storage system.

Steps

1. Enter the IP address for the interface you are configuring.

   Example

   ```
   Please enter the IP address for Network Interface e0 []::nnn.nn.nn.nnn
   ```

   nnn.nn.nn.nnn is the local address for the node you are configuring.

2. Enter the netmask for the interface you are configuring, or press Enter if the default value is correct.

   Example

   ```
   Please enter the netmask for Network Interface e1 [255.255.0.0]:
   ```

3. Specify that this interface does not take over a partner IP address.

   Example

   ```
   Should interface e1 take over a partner IP address during failover?
   [n]: n
   ```

Configuring standby interfaces with setup

You can assign a standby IP address to a network interface, so that the interface does not have a partner IP address.

About this task

This procedure is performed during setup of the storage system.

Steps

1. Do not enter an IP address for a standby interface; press Return.

For example:

```
Please enter the IP address for Network Interface e0 []:
```

2. Enter the netmask for the interface you are configuring, or press Return if the default value is correct.

   For example:

   Please enter the netmask for Network Interface e1 [255.255.0.0]:

3. Specify that this interface is to take over a partner IP address.

   For example:

   Should interface e1 take over a partner IP address during failover? [n]: y

Enabling licenses

You must enable the required licenses for your type of HA pair.

Before you begin

The licenses you need to add depend on the type of HA pair you have. The following table outlines the required licenses for each configuration.

Note: If you have a gateway, you must enable the gateway license on each node in the HA pair.

Configuration type	Required licenses
Standard HA pair	cf
Mirrored HA pair	• cf • syncmirror_local
MetroCluster	• cf • syncmirror_local • cf_remote

Steps

1. Enter the following command on both node consoles for each required license:

   `license add` *`license-code`*

   `license-code` is the license code you received for the feature.

2. Enter the following command to reboot both nodes:

   `reboot`

3. Enter the following command on the local node console:

```
cf enable
```

4. Verify that controller failover is enabled by entering the following command on each node console:

   ```
   cf status
   ```

   The system displays the following output if controller failover is enabled:
   ```
   Controller Failover enabled, filer2 is up.
   ```

Setting options and parameters

Options help you maintain various functions of your node, such as security, file access, and network communication. During takeover, the value of an option might be changed by the node doing the takeover. This can cause unexpected behavior during a takeover. To avoid unexpected behavior, specific option values must be the same on both the local and partner node.

Option types for HA pairs

Some options must be the same on both nodes in the HA pair, while some can be different, and some are affected by failover events.

In an HA pair, options are one of the following types:

- Options that must be the same on both nodes for the HA pair to function correctly
- Options that might be overwritten on the node that is failing over
  These options must be the same on both nodes to avoid losing system state after a failover.
- Options that should be the same on both nodes so that system behavior does not change during failover
- Options that can be different on each node

Note: You can find out whether an option must be the same on both nodes of an HA pair from the comments that accompany the option value when you enter the `option` command. If there are no comments, the option can be different on each node.

Setting matching node options

Because some Data ONTAP options need to be the same on both the local and partner node, you need to check these options with the options command on each node and change them as necessary.

Steps

1. View and note the values of the options on the local and partner nodes by entering the following command on each console:

   ```
   options
   ```

   The current option settings for the node are displayed on the console. Output similar to the following is displayed:
   ```
   autosupport.doit DONT
   ```

```
autosupport.enable on
```

2. Verify that the options with the following comments in parentheses are set to the same value for both nodes:
```
Value might be overwritten in takeover
Same value required in local+partner
Same value in local+partner recommended
```
3. Correct any mismatched options by entering the `options` *`option_name option_value`* command.

   Note: See the na_options man page for more information about the options.

Parameters that must be the same on each node

Lists the parameters that must be the same so that takeover is smooth and data is transferred between the nodes correctly.

The parameters listed in the following table must be the same so that takeover is smooth and data is transferred between the nodes correctly.

Parameter...	Setting for...
date	date, rdate
NDMP (on or off)	ndmp (on or off)
route table published	route
route enabled	routed (on or off)
Time zone	timezone

Best practices for cf options

You should generally use the default options and can.

The following table lists best practices for the `cf` options available in Data ONTAP. These settings are changed with the `options` command and you use the `man options` command to display the man page with detailed descriptions of the options.

Option	Recommended value and notes
`cf.giveback.auto.cancel.on_network_failure`	ON Leave this option on to avoid automatic giveback without ensuring that the partner's network interfaces are fully operational.
`cf.giveback.auto.cifs.terminate.minutes`	5 minutes (default value)

Option	Recommended value and notes
`cf.giveback.auto.delay.seconds`	600 seconds (default value)
`cf.giveback.auto.enable`	OFF (default value) Leave this option off so in cases other than takeover due to reboot or panic, you can verify that the failure that caused the takeover is fixed before doing a giveback. Each node in an HA pair can have a different setting for this option
`cf.giveback.auto.after.panic.takeover`	ON (default value)
`cf.giveback.auto.terminate.bigjobs`	OFF (default value) Each node in an HA pair can have a different setting for this option.
`cf.giveback.check.partner`	ON (default value) Leave on to reduce downtime caused by a unsuccessful giveback. Each node in an HA pair can have a different setting for this option.
`cf.hw_assist.enable`	ON (default value) The node must support remote management via a Remote LAN Management card or Service Processor to enable the hardware-assisted takeover.
`cf.hw_assist.partner.address`	When using hardware-assisted takeover, the value of this option should be equal to partner's node management IP address.
`cf.hw_assist.partner.port`	4444 (default value)
`cf.takeover.change_fsid`	ON (default value) Each node in an HA pair can have a different setting for this option.
`cf.takeover.detection.seconds`	15 seconds (default value) If sk.process.timeout.override has been manually set, it is strongly advised that this option is set to a value larger than or equal to the value of sk.process.timeout.override+5.

Option	Recommended value and notes
`cf.takeover.on_failure`	ON (default value) Changing the value on one node automatically changes the value on the partner.
`cf.takeover.on_network_interface_failure`	OFF (default value) Changing the value on one node automatically changes the value on the partner.
`cf.takeover.on_network_interface_failure.policy`	Use `all_nics` to avoid spurious takeovers due to any one network interface failure.
`cf.takeover.on_panic`	ON (default value) Use caution when manually changing the option value. In normal circumstances, leave this option on to avoid outages due to panics. Changing the value on one node automatically changes the value on the partner.
`cf.takeover.on_reboot`	ON Leave this option on to avoid outages due to long boot times. Changing the value on one node automatically changes the value on the partner.
`cf.takeover.on_short_uptime`	ON (default value) Leave this option on to avoid outages due to failures occurring early in the node's boot process. Changing the value on one node automatically changes the value on the partner.
`cf.takeover.use_mcrc_file`	OFF (default value) Use only in MetroCluster environments.

Disabling the change_fsid option in MetroCluster configurations

In a MetroCluster configuration, you can take advantage of the `change_fsid` option in Data ONTAP to simplify site takeover when the `cf forcetakeover -d` command is used.

About this task

In a MetroCluster configuration, if a site takeover initiated by the `cf forcetakeover -d` command occurs, the following happens:

- Data ONTAP changes the file system IDs (FSIDs) of volumes and aggregates because ownership changes.
- Because of the FSID change, clients must remount their volumes if a takeover occurs.

- If using Logical Units (LUNs), the LUNs must also be brought back online after the takeover.

To avoid the FSID change in the case of a site takeover, you can set the `change_fsid` option to `off` (the default is `on`). Setting this option to `off` has the following results if a site takeover is initiated by the `cf forcetakeover -d` command:

- Data ONTAP refrains from changing the FSIDs of volumes and aggregates.
- Users can continue to access their volumes after site takeover without remounting.
- LUNs remain online.

Attention: If the option is set to `off`, any data written to the failed node that did not get written to the surviving node's NVRAM is lost. Disable the `change_fsid` option with great care.

Step

1. Enter the following command to disable the `change_fsid` option:

   `options cf.takeover.change_fsid off`

   By default, the `change_fsid` option is enabled (set to `on`).

Related concepts

Disaster recovery using MetroCluster configurations on page 196

Clarification of when data loss can occur when the change_fsid option is enabled

Ensure that you have a good understanding of when data loss can occur before you disable the `change_fsid` option. Disabling this option can create a seamless takeover for clients in the event of a disaster, but there is potential for data loss.

If both the ISLs between the two sites in a fabric MetroCluster go down, then both the systems remain operational. However, in that scenario, client data is written only to the local plex and the plexes become unsynchronized.

If, subsequently, a disaster occurs at one site, and the `cf forcetakeover -d` command is issued, the remote plex which survived the disaster is not current. With the `change_fsid` option set to `off`, clients switch to the stale remote plex without interruption.

If the `change_fsid` option is set to `on`, the system changes the fsids when the `cf forcetakeover -d` is issued, so clients are forced to remount their volumes and can then check for the integrity of the data before proceeding.

Verifying and setting the HA state on controller modules and chassis

Some storage system models recognize that they are in an HA pair based on HA state information in the controller module and chassis PROMs. If that state information is incorrect (possibly after a

chassis or controller module replacement), you can verify the state, and, if necessary, update the state.

About this task

- The `ha-config show` and `ha-config modify` commands are Maintenance mode commands.
- The `ha-config` command only applies to the local controller module and, in the case of a dual-chassis HA pair, the local chassis.
  To ensure consistent HA state information throughout the configuration, you must also run these commands on the partner controller module and chassis, if necessary.

Steps

1. Boot into Maintenance mode.
2. Enter the following command to display the HA state of the local controller module and chassis:
   `ha-config show`
3. If necessary, enter the following command to set the HA state of the controller module:
   `ha-config modify controller` ***`ha_state`***
   `ha_state` is `ha` or `non-ha`.
   The HA state of the controller module is changed.
4. If necessary, enter the following command to set the HA state of the chassis:
   `ha-config modify chassis` ***`ha_state`***
   `ha_state` is `ha` or `non-ha`.
   The HA state of the chassis is changed.
5. Repeat the preceding steps on the partner controller module and chassis, if necessary.

Configuring hardware-assisted takeover

You can configure hardware-assisted takeover to speed up takeover times. Hardware-assisted takeover uses the remote management device to quickly communicate local status changes to the partner node.

How hardware-assisted takeover speeds up takeover

Hardware-assisted takeover speeds up the takeover process by using a node's remote management device (SP or RLM) to detect failures and quickly initiate the takeover rather than waiting for Data ONTAP to recognize that the partner's heartbeat has stopped.

Without hardware-assisted takeover, if a failure occurs, the partner waits until it notices that the node is no longer giving a heartbeat, confirms the loss of heartbeat, and then initiates the takeover.

The hardware-assisted takeover feature uses the following process to take advantage of the remote management device and avoid that wait:

1. The remote management device monitors the local system for certain types of failures.
2. If a failure is detected, the remote management device immediately sends an alert to the partner node.
3. Upon receiving the alert, the partner initiates takeover.

The hardware-assisted takeover option (`cf.hw_assist.enable`) is enabled by default.

Disabling and enabling the hardware-assisted takeover option

Hardware-assisted takeover is enabled by default on systems that use remote management. Hardware-assisted takeover speeds the takeover process by using the RLM or SP to quickly detect potential takeover events and alerting the partner node.

Step

1. Enter the following command to disable or enable the `cf.hw_assist` option:

   `options cf.hw_assist.enable off`

   `options cf.hw_assist.enable on`

Setting the partner address for hardware-assisted takeover

By default, on systems with an e0M port, the hardware-assisted takeover feature uses the IP address of the partner's e0M port to communicate with the partner. On systems without an e0M port, the system automatically selects another configured IP address. You can use the `cf.hw_assist.partner.address` option to select a different IP address.

Step

1. Enter the following command to set the IP address or host name to which the hardware failure notification is sent:

   `options cf.hw_assist.partner.address` ***`address`***

Setting the partner port for hardware-assisted takeover

When hardware-assisted takeover is enabled, the remote management (either RLM or SP) sends hardware failure notifications to the partner. The `cf.hw_assist.partner.port` option enables you to change the partner port. The default is 4444.

Step

1. Enter the following command to set the partner port to which the hardware failure notification is sent:

   `options cf.hw_assist.partner.port` ***`port_number`***

Configuring network interfaces for HA pairs

Configuring network interfaces requires that you understand the available configurations for takeover and that you configure different types of interfaces (shared, dedicated, and standby) depending on your needs.

Understanding interfaces in an HA pair

You can configure three types of interfaces on nodes in an HA pair.

What the networking interfaces do

When a node in an HA pair fails, the surviving node must be able to assume the identity of the failed node on the network. Networking interfaces allow individual nodes in the HA pair to maintain communication with the network if the partner fails.

See the *Data ONTAP Network Management Guide for 7-Mode* for a description of available options and the function each performs.

Note: You should always use multiple NICs with interface groups to improve networking availability for both stand-alone storage systems and systems in an HA pair.

Shared, dedicated, and standby interfaces

These different types of interfaces have different roles in normal and takeover mode.

The following table lists the three types of interface configurations that you can enable in an HA pair.

Interface type	Description
Shared	This type of interface supports both the local and partner nodes. It contains both the local node and partner node IP addresses. During takeover, it supports the identity of both nodes.
Dedicated	This type of interface only supports the node in which it is installed. It contains the local node IP address only and does not participate in network communication beyond local node support during takeover. It is paired with a standby interface.
Standby	This type of interface is on the local node, but only contains the IP address of the partner node. It is paired with a dedicated interface.

Note: Most HA pair interfaces are configured as shared interfaces because they do not require an extra NIC.

Interface roles in normal and takeover modes

You can configure shared, dedicated, and standby interfaces in an HA pair. Each type has a different role in normal and takeover mode.

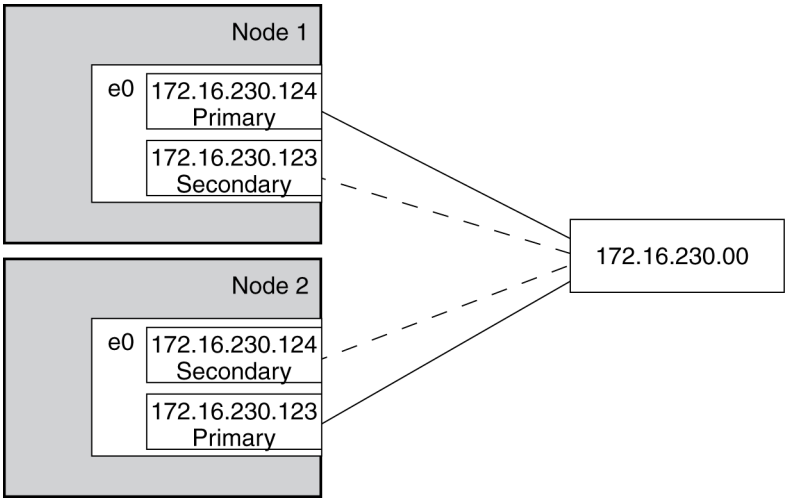
The following table shows the role of each interface type in normal and takeover mode.

Interface type	Normal mode	Takeover mode
Shared	Supports the identity of the local node	Supports the identity of both the local node and the failed node
Dedicated	Supports the identity of the local node	Supports the identity of the local node
Standby	Idle	Supports the identity of the failed node

Takeover configuration with shared interfaces

You can configure two NICs on to provide two shared interfaces to each node.

In the following configuration illustration, you use two NICs to provide the two interfaces.



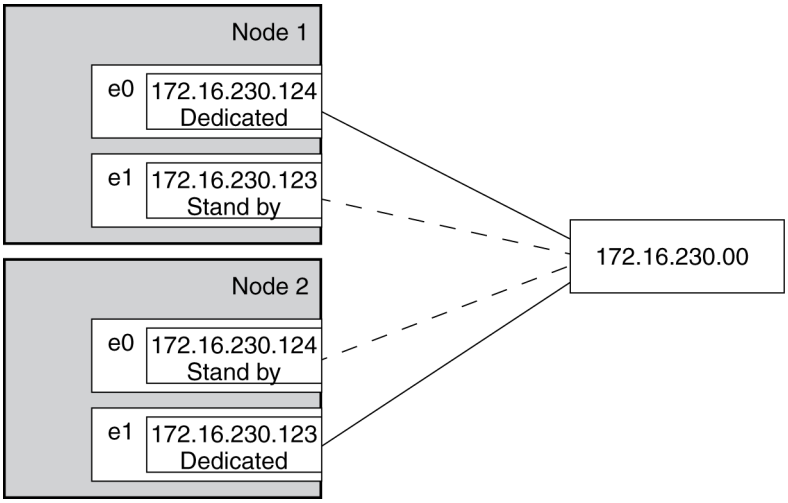
If Node 1 fails, interface e0 on Node 1 stops functioning, but the secondary address on e0 on Node 2 handles the Node 1 network connection with the 230 network.

If Node 2 fails, e0 on Node 2 stops functioning, but e0 on Node 1 substitutes for the failed interface and handles the Node 2 network connection with the 230 network.

Takeover configuration with dedicated and standby interfaces

With two NICs on each node, one can provide a dedicated interface and the other can act as a standby interface.

In the following configuration illustration, you use two NICs for each interface, one on each storage system. One NIC acts as a dedicated interface and the other acts as a standby interface.



If Node 1 fails, interface e0 on Node 1 stops functioning, but e0 on Node 2 substitutes for the failed interface and handles the Node 1 network connection with the 230 network.

If Node 2 fails, e1 on Node 2 stops functioning, but e1 on Node 1 substitutes for the failed interface and handles the Node 2 network connection with the 230 network.

Interface types and configurations

This table lists the configurations supported by each type of interface in an HA pair.

Interface	Shared	Dedicated	Standby	Partner parameter
Ethernet	X	X	X	IP address or interface name
Gigabit Ethernet	X	X	X	IP address or interface name
Virtual interface	X	X	X	Virtual interface name
VLAN interface	X	X	X	IP address or interface name

Note: Some storage systems, such as the N6000 series systems, include an e0M interface that is dedicated to management traffic. This port can be partnered in an HA pair in the same way as a regular Ethernet interface.

IPv6 considerations in an HA pair

When enabled, IPv6 provides features such as address autoconfiguration. Using these IPv6 features requires an understanding of how these features work with the HA pair functionality.

For additional information about IPv6, see the *Data ONTAP Network Management Guide for 7-Mode*.

Configuration requirements for using IPv6

To use IPv6 in an HA pair, IPv6 must be enabled on both nodes. If a node that does not have IPv6 enabled attempts to take over a node using IPv6, the IPv6 addresses configured on the partner's interfaces are lost because the takeover node does not recognize them.

Using the ifconfig command

When using the `ifconfig` command with IPv4, the partner's interface can be mapped to a local interface or the partner's IP address. When using IPv6, you must specify the partner interface, not an IP address.

Generation of addresses during takeover

For manually configured IPv6 addresses, during takeover, the mechanism used to configure partner's IP address remains same as in the case of IPv4.

For link-local auto-configured IPv6 addresses, during takeover, the address is auto-generated based on the partner's MAC address.

Prefix-based auto-configured addresses are also generated during takeover, based on the prefixes in router advertisements (RAs) received on the local link and on the partner's MAC address.

Duplicate Address Detection (DAD) is performed on all IPv6 partner addresses during takeover. This can potentially keep the addresses in *tentative* state for some amount of time.

Making nondisruptive changes to the interface groups

You can use the `cf takeover` and `cf giveback` commands to make changes to interface groups in the HA pair in a nondisruptive manner.

About this task

Changes to the `/etc/rc` file require a reboot to make the changes effective. You can use the `cf takeover` and `cf giveback` commands to take over one node in the HA pair, causing it to reboot while its storage is taken over by the partner.

Steps

1. Edit the `/etc/rc` file on the desired node to modify the interface groups.

   See the *Data ONTAP Network Management Guide for 7-Mode* for more information about configuring interface groups.

2. From the partner node (the partner of the node on which you performed step 1), enter the following command:

   `cf takeover`

3. Enter the following command:

   `cf giveback`

   The node on which the changes were made reboots and its `/etc/rc` file is reread. The `rc` file is responsible for creating the interface groups.

4. Repeat these steps, making any required changes to the `/etc/rc` file on the partner node.

Configuring network interfaces for the HA pair

You must configure network interfaces so that if takeover occurs, interfaces on the operating node takes over interfaces on the failed-over node and hosts can still reach data over the network.

Before you begin

Both nodes in the HA pair must have interfaces that access the same collection of networks and subnetworks.

You must gather the following information before configuring the interfaces:

- The IP address for both the local node and partner node.
- The netmask for both the local node and partner node.
- The MTU size for both the local node and partner node.
  The MTU size must be the same on both the local and partner interface.

Note: You should always use multiple NICs with interface groups to improve networking availability for both stand-alone storage systems and systems in an HA pair.

About this task

If you configured your interfaces using setup when you first applied power to your storage systems, you do not need to configure them again.

Note: For information about configuring an HA pair to use FC, see the *Data ONTAP SAN Administration Guide for 7-Mode*.

Steps

1. Determine whether to use shared interfaces or dedicated and standby interfaces.

2. Configure your interfaces on one node.
3. Repeat Step 2 on the other node.
4. If desired, configure automatic takeover for the interfaces or interface groups.

Configuring a partner interface in an HA pair

To prepare for a successful takeover in an HA configuration, you can map a network interface to an IP address or to another network interface on the partner node. During a takeover, the network interface on the surviving node assumes the identity of the partner interface.

Before you begin

When specifying the partner IP address, both the local network interface and the partner's network interface must be attached to the same network segment or network switch.

About this task

- If the network interface is an interface group, the partner interface must be denoted by an interface name and not an IP address.

  The partner interface can be an interface group or a physical network interface.
- You cannot specify the underlying physical ports of an interface group in a partner configuration.
- If IPv6 addresses are to be taken over, you must specify the partner interface, and not an IP address.

  Address to address mapping is not supported for IPv6 addresses.
- For the partner configuration to be persistent across reboots, you must include the `ifconfig` command in the `/etc/rc` file.

  For a successful takeover in both directions, you must repeat the partner configuration in the `/etc/rc` files of each node.
- When specifying the partner interface name, you can configure the interfaces symmetrically, for example map interface e1 on one node to interface e1 on the partner node.

  Though symmetrical configuration is not mandatory, it simplifies administration and troubleshooting tasks.

Step

1. Depending on the partner configuration that you want to specify, enter the following command:

If you want specify a...	Enter the following command...
Partner IP address	**`ifconfig`** ***`interface_name`*** **`partner`** ***`address`*** ***`interface_name`*** is the name of the network interface. ***`address`*** is the partner IP address.

If you want specify a...	Enter the following command...
Partner interface name	**ifconfig** ***interface_name*** **partner** ***partner_interface*** *partner_interface* is the name of the partner network interface.

Example: Specifying a partner IP address and partner interface name

Consider node1 and node2 are two storage systems in an HA configuration.

If the IP address of the interface e8 on node2 is 198.9.200.38, the following command allows the interface e1 of node1 to take over the IP address of node2 for the duration of the takeover:

```
node1> ifconfig e1 partner 198.9.200.38
```

Instead of specifying the IP address, you can also specify the partner interface name. The following command allows the interface e1 of node1 to assume the identity of e8 of node2 for the duration of the takeover:

```
node1> ifconfig e1 partner e8
```

Configuring partner addresses on different subnets (MetroCluster configurations only)

On MetroCluster configurations, you can configure partner addresses on different subnets. To do this, you must create a separate `/etc/mcrc` file and enable the `cf.takeover.use_mcrc_file` option. When taking over its partner, the node uses the partner's `/etc/mcrc` file to configure partner addresses locally. These addresses will reside on the local subnetwork.

The /etc/mcrc file

The `/etc/mcrc` file, in conjunction with the `cf.takeover.use_mcrc_file` option, should be used on MetroCluster configurations in which the partner nodes reside on separate subnetworks.

Normally, when a node (for example, nodeA) takes over its partner (nodeB), nodeA runs nodeB's `/etc/rc` file to configure interfaces on nodeA to handle incoming traffic for the taken-over partner, nodeB. This requires that the local and partner addresses are on the same subnetwork.

When the `cf.takeover.use_mcrc_file` option is enabled on nodeA, nodeA will use nodeB's `/etc/mcrc` file upon takeover, instead of nodeB's `/etc/rc` file. The ifconfig commands in the `/etc/mcrc` file can configure IP addresses on nodeA's subnetwork. With the correct ifconfig, virtual IP (VIP), and routing commands in the `/etc/mcrc` file, the resulting configuration allows hosts connecting to nodeB to connect to node A.

Note: The `/etc/mcrc` file must be maintained manually and kept in sync with the `/etc/rc` file.

Example /etc/rc and /etc/mcrc files

NodeA's `/etc/rc` file, which configures its local addresses and a partner address (which matches the address configured in NodeB's `/etc/mcrc` file):

```
hostname nodeA
ifconfig e0a 10.1.1.1 netmask 255.255.255.0
ifconfig e0a partner 10.1.1.100
ifconfig vip add 5.5.5.5
route add default 10.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

NodeA's `/etc/mcrc` file, which configures a partner address on NodeB's subnetwork:

```
hostname nodeA
ifconfig e0a 20.1.1.200 netmask 255.255.255.0
ifconfig vip add 5.5.5.5
route add default 20.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

NodeB's `/etc/rc` file, which configures its local addresses and a partner address (which matches the address configured in NodeA's `/etc/mcrc` file):

```
hostname nodeB
ifconfig e0a 20.1.1.1 netmask 255.255.255.0
ifconfig e0a partner 20.1.1.200
ifconfig vip add 7.7.7.7
route add default 20.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

NodeB's `/etc/mcrc` file, which configures a partner address on NodeA's subnetwork:

```
hostname nodeB
ifconfig e0a 10.1.1.100 netmask 255.255.255.0
ifconfig vip add 7.7.7.7
```

```
route add default 10.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

Creating an /etc/mcrc file

You should create an `/etc/mcrc` file on each node of your MetroCluster configuration if the nodes are on separate subnetworks.

Steps

1. Create an `/etc/mcrc` file on one node (nodeA) and place it in the /etc directory.

   You might want to create the `/etc/mcrc` file by copying the `/etc/rc` file.

   Note: The `/etc/mcrc` file must be configured manually. It is not updated automatically. It must include all commands necessary to implement the network configuration on the partner node in the event the node is taken over by the partner.

2. Enter the following commands in nodeA's `/etc/mcrc` file:

   ```
   hostname nodeA
   ifconfig interface MetroCluster-partner-address netmask netmask
   ifconfig vip add virtual-IP-address
   route add default route-for-MetroCluster-partner-address 1
   routed on
   other-required-options
   ```

   `interface` is the interface on which the corresponding *`MetroCluster-partner-address`* will reside.

   `MetroCluster-partner-address` is the partner address of nodeB. It corresponds to the partner address configured by an `ifconfig` command in nodeB's `/etc/rc` file.

   `virtual-IP-address` is the virtual address of the partner (nodeB).

   `other-required-options` denotes whatever other options are needed to correctly configure the interface in your network environment.

Example

Example of nodeA's `/etc/mcrc` file:

```
hostname nodeA
ifconfig e0a 20.1.1.200 netmask 255.255.255.0
```

```
ifconfig vip add 5.5.5.5
route add default 20.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

3. Create an `/etc/mcrc` file on the other node (nodeB) and place it in the `/etc` directory.

   The `/etc/mcrc` file must include an `ifconfig` command that configures the address that corresponds to the address specified in the `partner` parameter in the partner node's `/etc/rc`.

   You might want to create the `/etc/mcrc` file by copying the `/etc/rc` file.

   Note: The `/etc/mcrc` file must be configured manually. It is not updated automatically. It must include all commands necessary to configure the interfaces.

4. Enter the following commands in nodeB's `/etc/mcrc` file:

```
hostname nodeB
ifconfig interface MetroCluster-partner-address netmask netmask
ifconfig vip add virtual-IP-address
route add default route-for-MetroCluster-partner-address 1
routed on
other-required-options
```

   `interface` is the interface on which the corresponding *`MetroCluster-partner-address`* will reside.

   `MetroCluster-partner-address` is the partner address of nodeA. It corresponds to the partner address configured by an `ifconfig` command in nodeA's `/etc/rc` file.

   `virtual-IP-address` is the virtual address of the partner (nodeA).

   `other-required-options` denotes whatever other options are needed to correctly configure the interface in your network environment.

   Example

   Example of nodeB's `/etc/mcrc` file:

```
hostname nodeB
ifconfig e0a 10.1.1.100 netmask 255.255.255.0
ifconfig vip add 7.7.7.7
route add default 10.1.1.50 1
routed on
options dns.domainname mycompany.com
options dns.enable on
options nis.enable off
savecore
```

Setting the system to use the partner's /etc/mcrc file at takeover

You must enable the `cf.takeover.use_mcrc_file` option to cause the system to use the partner's `/etc/mcrc` in the event that the local system takes over the partner. This allows the partner IP addresses to reside on separate subnetworks. This option should be set on both nodes in the MetroCluster.

Step

1. Enter the following command on both nodes:

   `options cf.takeover.use_mcrc_file on`
   The default is `off`.

Testing takeover and giveback

After you configure all aspects of your HA pair, you need to verify that it operates as expected.

Steps

1. Check the cabling on the HA interconnect cables to make sure that they are secure.
2. Verify that you can create and retrieve files on both nodes for each licensed protocol.
3. Enter the following command from the local node console:

   `cf takeover`

   See the man page for command details.

   The local node takes over the partner node and gives the following output:
   `Failover monitor: takeover completed`
4. Use the `fcstat device_map` command to ensure that one node can access the other node's disks.
5. Give back the partner node after it displays the `Waiting for giveback` message by entering the following command:

   `cf giveback`

   The local node releases the partner node, which reboots and resumes normal operation. The following message is displayed on the console when the process is complete:
   `giveback completed`
6. Proceed depending on whether you saw the message that giveback was completed successfully:

If takeover and giveback...	Then...
Is completed successfully	Repeat Step 2 through Step 7 on the partner node.

If takeover and giveback...	Then...
Fails	Correct the takeover or giveback failure and then repeat this procedure.

Managing takeover and giveback

An HA pair allows one partner to take over the storage of the other, and return the storage using the giveback operation. Management of the nodes in the HA pair differs depending on whether one partner has taken over the other, and the takeover and giveback operations themselves have different options.

This information applies to all HA pairs regardless of disk shelf type.

Monitoring an HA pair in normal mode

You can display information about the status and configuration of HA pair in normal mode (when neither node has taken over the other).

Monitoring HA pair status

You can use commands on the local node to determine whether the controller failover feature is enabled and whether the other node in the HA pair is up.

Step

1. Enter the following command:

   `cf status`

   The following example shows that the HA pair is enabled and the interconnect is up and working correctly.

   ```
   node1>cf statusController Failover enabled, node2 is up.
   RDMA Interconnect is up (Link 0 up).
   ```

   If the output shows that one link is down, the HA pair is degraded and you must configure the link so that it is up while the other link is still active.

   Note: Depending on the storage system model, the output might display either `RDMA interconnect` or `VIA interconnect` in the last line.

   Note: Data ONTAP can disable controller failover if a software or hardware problem exists that prevents a successful takeover. In this case, the message returned from the `cf status` command describes the reason failover is disabled.

Description of HA pair status messages

The `cf status` command displays information about the status of the HA pair.

The following table shows messages that the `cf status` command can display.

Message	**Meaning**
HA enabled, partner_name is up.	The HA pair is operating normally.
local_node has taken over partner_node.	One node took over the other node.
Interconnect not present.	The system does not recognize the existence of a HA interconnect adapter.
Interconnect is down.	The HA interconnect adapter cannot access the partner. This might be due to cabling problems or the partner might be down.
Interconnect is up.	The HA interconnect adapter is active and can transmit data to the partner.
partner_name_1 has detected a mailbox disk error, takeover of partner_name_2 disabled.	One node cannot access multiple mailbox disks. Check access to both the local and partner root volumes and mirrors, if they exist. Also check for disk or FC-AL problems or offline storage adapters.
partner_name_2 may be down, takeover disabled because of reason (partner halted in notakeover mode) partner_name_1 has disabled takeover by partner_name_2 (interconnect error) interconnect_type Interconnect is down (Link 0 down).	One node might be down.
Version mismatch.	The partner node has an incompatible version of Data ONTAP.
partner_name_1 is attempting takeover of partner_name_2, takeover is in module n of N modules.	A takeover is being attempted (includes information about how far the takeover has progressed).
partner_name_1 has taken over partner_name_2, giveback in progress, giveback is in module n of N modules.	A giveback is being attempted (includes information about how far the giveback has progressed).
partner_name_1 has taken over partner_name_2, partner_name_2 is ready for giveback.	The takeover node received information that the failed node is ready for giveback.
partner_name_1 has taken over partner_name_2, partner_name_2 is ready for giveback. Automatic giveback is disabled due to exceeding retry count.	The takeover node received information that the failed node is ready for giveback, but giveback cannot take place because the number of retries exceeded the limit.

Monitoring the hardware-assisted takeover feature

You can check and test the hardware-assisted takeover configuration using the `hw_assist` command. You can also use the command to review statistics relating to hardware-assisted takeover.

Checking the hardware-assisted takeover status of the local and partner node

You can check the status of the hardware-assisted takeover configuration with the `cf hw_assist status` command. It shows the current status for the local and partner nodes.

Step

1. Enter the following command to display the hardware-assisted takeover status:

   `cf hw_assist status`

Example of hardware-assisted takeover status

The following example shows output from the `cf hw_assist status` command:

```
Local Node Status - ha1

        Active: Monitoring alerts from partner(ha2)

        port 4004 IP address 172.27.1.14


Partner Node Status - ha2

        Active: Monitoring alerts from partner(ha1)

        port 4005 IP address 172.27.1.15
```

Testing the hardware-assisted takeover configuration

You can test the hardware-assisted takeover configuration with the `cf hw_assist test` command.

About this task

The `cf hw_assist test` command sends a test alert to the partner. If the alert is received the partner sends back an acknowledgment, and a message indicating the successful receipt of the test alert is displayed on the console.

Step

1. Enter the following command to test the hardware-assisted takeover configuration:

cf hw_assist test

After you finish

Depending on the message received from the `cf hw_assist test` command, you might need to reconfigure options so that the HA pair and the remote management card are operating.

Checking hardware-assisted takeover statistics

You can display statistics about hardware-assisted takeovers to determine how many alert events of each type have been received from the partner.

Step

1. Enter the following command to display or clear the hardware-assisted takeover statistics, respectively:

   cf hw_assist stats

   cf hw_assist stats clear

Example of hardware-assisted takeover statistics

The following example shows output from the `cf hw_assist stats` command on a system that has received a variety of alerts from the partner:

```
# cf hw_assist: stats

Known hw_assist alerts received from partner

   alert type       alert event                          num of
alerts

   ----------       -----------
-------------

   system_down       post_error                           0

   system_down       power_loss                           0

   system_down       abnormal_reboot                      0

   system_down       l2_watchdog_reset                    0

   system_down       power_off_via_rlm                    0

   system_down       power_cycle_via_rlm                  0

   system_down       reset_via_rlm                        0

   keep_alive        loss_of_heartbeat                    0

   keep_alive        periodic_message                     18
```

```
   test              test                                  6
Unknown hw_assist alerts received from partner
   Partner nvramid mismatch alerts 5
   Shared secret mismatch alerts 10
   Unknown alerts 23
Number of times hw_assist alerts throttled: 3
```

Displaying the partner's name

You can display the name of the other node with the `cf partner` command.

Step

1. Enter the following command:

   `cf partner`

   Note: If the node does not yet know the name of its partner because the HA pair is new, this command returns `partner`.

Displaying disk and array LUN information on an HA pair

To find out about the disks, array LUNs, or both on both the local and partner node, you can use the `sysconfig` and `aggr status` commands, which display information about both nodes.

About this task

For each node, the `sysconfig` command output displays disks on both channel A and channel B:

- The information about disks on channel A is the same as for storage systems not in an HA pair.
- The information about disks on channel B is for hardware only; the `sysconfig` command displays information about the adapters supporting the disks.
  The command does not show whether a disk on channel B is a file system disk, spare disk, or parity disk.

Step

1. Enter one of the following commands:

   `sysconfig -r`
   or

   `aggr status -r`

What takeover and giveback are

Takeover is the process in which a node takes over the storage of its partner. Giveback is the process in which that storage is returned to the partner. Both processes can be initiated manually or configured for automatic initiation.

When takeovers occur

Takeovers occur when you manually enter a command to do a takeover, and automatic takeovers can occur when a failover event happens, depending on how you configure the HA pair. In some cases, takeovers occur automatically regardless of configuration.

Takeovers can be initiated when one of the following conditions occur:

- You initiate a takeover manually.
- A node is in an HA pair with the default configuration for immediate takeover on panic, and it undergoes a software or system failure that leads to a panic.
  By default, after the partner recovers from the panic and boots up, the node automatically does a giveback to return the partner to normal operation.
- A node that is in an HA pair undergoes a system failure (for example, a loss of power) and cannot reboot.

  Note: If the storage for a node also loses power at the same time, a standard takeover is not possible. For MetroCluster configurations, you can initiate a forced takeover in this situation.
- One or more network interfaces that are configured to support failover become unavailable.
- A node cannot send heartbeat messages to its partner.
  This could happen if the node experienced a hardware or software failure that did not result in a panic but still prevented it from functioning correctly.
- You halt one of the nodes without using the `-f`or `-inhibit-takeover true` parameter.
- Hardware-assisted takeover is enabled and triggers a takeover when the remote management device (RLM or Service Processor) detects an issue on the partner node.

What happens during takeover

When a takeover occurs, the unimpaired partner node takes over the functions and disk drives of the failed node by creating an emulated storage system.

The emulated system performs the following tasks:

- Assumes the identity of the failed node
- Accesses the failed node's disks, array LUNs, or both and serves its data to clients

The partner node maintains its own identity and its own primary functions, but also handles the added functionality of the failed node through the emulated node.

Note: When a takeover occurs, existing CIFS sessions are terminated. A graceful shutdown of the CIFS sessions is not possible, and some data loss could occur for CIFS users.

What happens after takeover

After a takeover occurs, you view the surviving partner as having two identities, its own and its partner's, that exist simultaneously on the same storage system. Each identity can access only the appropriate volumes and networks. You can send commands or log in to either storage system by using the `rsh` command, allowing remote scripts that invoke storage system commands through a Remote Shell connection to continue to operate normally.

Access with rsh

Commands sent to the failed node through a Remote Shell connection are serviced by the partner node, as are `rsh` command login requests.

Access with telnet

If you log in to a failed node through a Telnet session, you see a message alerting you that your storage system failed and to log in to the partner node instead. If you are logged in to the partner node, you can access the failed node or its resources from the partner node by using the partner command.

What happens during giveback

After any issues on the partner node are resolved or maintenance completed, the partner node has booted up, and giveback is initiated (either manually or automatically), the local node returns the ownership of the partner's aggregates and volumes to the partner.

When the failed node is functioning again, the following events can occur:

- You issue a `giveback` command that terminates the emulated node on the partner.
- The failed node resumes normal operation, serving its own data.
- The HA pair resumes normal operation, with each node ready to take over for its partner if the partner fails.

Configuring automatic takeover

You can control when automatic takeovers happen by setting the appropriate options.

Reasons for automatic takeover

You can set options to control whether automatic takeovers occur due to different system errors. In some cases, automatic takeover occurs by default unless you disable the option, and in some cases automatic takeover cannot be prevented.

Takeovers can happen for several reasons. Some system errors must cause a takeover; for example, when a system in an HA pair loses power, it automatically fails over to the other node.

However, for some system errors, a takeover is optional, depending on how you set up your HA pair. The following table outlines which system errors can cause a takeover to occur, and whether you can configure the HA pair for that error.

System error	Option used to configure	Default value	Notes
A node undergoes a system failure and cannot reboot.	`cf.takeover.on_failure` set to `on`	On	You should leave this option enabled unless instructed otherwise by technical support.
A node undergoes a software or system failure leading to a panic.	`cf.takeover.on_panic` set to `on`	On	
A node reboots.	`cf.takeover.on_reboot` set to `on`	On, unless FC or iSCSI is licensed. **Note:** In releases prior to Data ONTAP 8.0, the system would only take over the partner after reboot if the partner took longer than 90 seconds to boot.	

System error	Option used to configure	Default value	Notes
All the network interface cards (NICs) or interface groups enabled for negotiated failover on a node failed.	`cf.takeover.on_network_interface_failure` set to `on`, `cf.takeover.on_network_interface_failure.policy` set to `all_nics`	By default, takeover on network failure is disabled.	To enable a network interface for negotiated failover, you use the `ifconfig if_name nfo` command. For more information, see the *Data ONTAP MultiStore Management Guide for 7-Mode*.
One or more of the NICs or interface groups enabled for negotiated failover failed. **Note:** If interfaces fail on both nodes in the HA pair, takeover won't occur.	`cf.takeover.on_network_interface_failure` set to `on` `cf.takeover.on_network_interface_failure.policy` set to `any_nic`	By default, takeover on network failure is disabled.	To enable a network interface or interface group for negotiated failover, you use the `ifconfig if_name nfo` command. For more information, see the *Data ONTAP MultiStore Management Guide for 7-Mode*.
A node fails within 60 seconds of booting up.	`cf.takeover.on_short_uptime` set to `on`	On	Changing the value of this option on one node automatically updates the option on the partner node.
A node cannot send heartbeat messages to its partner.	n/a		You cannot prevent this condition from causing a takeover.
You halt one of the nodes *without* using the `-f` flag.	n/a		You cannot prevent this condition from causing a takeover. If you include the `-f` flag, the takeover is prevented.
You initiate a takeover manually using the `cf takeover` command.	n/a		You cannot prevent this condition from causing a takeover.

Commands for performing a manual takeover

You need to know the commands you can use when initiating a takeover. You can initiate a takeover on a node in an HA pair to perform maintenance on that node while still serving the data on its disks, array LUNs, or both to users.

Command	Description
`cf takeover`	Initiates a takeover of the partner of the local node. Takeover is aborted if a core dump is in progress on the partner (if the `cf.takeover.on_panic` option is set to `off`). The takeover starts either after the partner halts successfully or after a timeout.
`cf takeover -f`	Initiates an immediate takeover of the partner of the local node regardless of whether the other node is dumping its core. The partner node is not allowed to halt gracefully.
`cf forcetakeover`	Tells the HA monitor to ignore some configuration problems that would otherwise prevent a takeover, such as unsynchronized NVRAM due to a faulty HA interconnect connection. It then initiates a takeover of the partner of the local node.
`cf forcetakeover -d`	Initiates a takeover of the local partner even in the absence of a quorum of partner mailbox disks or partner mailbox LUNs. The `cf forcetakeover -d` command is valid only if the cf_remote license is enabled. **Attention:** You should only use the `-d` option after you verify that the partner is down. The `-d` option is used in conjunction with RAID mirroring to recover from disasters in which one partner is not available. For more information, see the *Data ONTAP Data Protection Online Backup and Recovery Guide for 7-Mode*.
`cf takeover -n`	Initiates a takeover for a nondisruptive upgrade. For more information, see the *Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode*.

Halting a node without takeover

You can halt the node and prevent its partner from taking over.

About this task

You can halt the node and prevent its partner from taking over. For example, you might need to perform maintenance on both the storage system and its disks and want to avoid an attempt by the partner node to write to those disks.

Step

1. Enter the following command:

   `halt -f`

Rebooting a node without takeover

You can reboot the node and prevent its partner from taking over, overriding the `cf.takeover.on_reboot` option.

Step

1. Enter the following command:

   `reboot -f`

Enabling and disabling takeover

You might want to use the `cf disable` command to disable takeover if you are doing maintenance that typically causes a takeover. You can reenable takeover with the `cf enable` command after you finish maintenance.

Step

1. Enter the following command:

   `cf enable|disable`

   Use `cf enable` to enable takeover or `cf disable` to disable takeover.

   Note: You can enable or disable takeover from either node.

Enabling and disabling takeover on reboot

The takeover on reboot option enables you to control whether an automatic takeover occurs when a node reboots. This automatic takeover, and the automatic giveback that follows after the reboot is

complete, can reduce the outage during which the storage belonging to the rebooting system is unavailable.

About this task

If this option is enabled and a takeover occurs because of a reboot, then an automatic giveback is performed after the partner has booted. This giveback occurs even if the `cf.giveback.auto.enable` option is set to `off`. However, if a node takes over its partner due to a reboot and that node itself reboots before it can execute a giveback, it performs automatic giveback only if `cf.giveback.auto.enable` is set to `on`.

If the `cf.takeover.on_reboot` is off and a node is rebooted then the partner will not take over immediately. But the partner could take over later if the node takes more than 180 seconds to boot.

Note: If the `reboot -f` command is used, then the partner does not take over under any circumstances, even if the reboot timer expires.

Step

1. Enter the following command:

   `options cf.takeover.on_reboot on`

   The default is `on`, unless FC or iSCSI is licensed, in which case the default is `off`.

   Note: If you enter this command on one node, the value applies to both nodes.

   This option is persistent across reboots.

Enabling and disabling automatic takeover of a panicked partner

Data ONTAP is configured by default to initiate a takeover immediately if the partner node panics. This shortens the time between the initial failure and the time that service is fully restored because the takeover can be quicker than the recovery from the panic, although the subsequent giveback causes another brief outage.

About this task

- If you enter this command on one node, the value applies to both nodes.
- The setting of this option is persistent across reboots.
- By default, Data ONTAP will initiate an automatic giveback after a takeover on panic. The `cf.giveback.auto.after.panic.takeover` option can be used to disable this automatic giveback.

Steps

1. Verify that controller takeover is enabled by entering the following command:

   `cf enable`

2. Enable or disable automatic takeover on panic by entering the following command:

   options cf.takeover.on_panic {on|off}

   `on` enables immediate takeover of a panicked node. This is the default value.

   `off` disables immediate takeover of a panicked node. If you disable this option, normal takeover procedures apply: if a node panics and stays down without sending messages to its partner for 15 seconds, the partner then automatically takes over the failed node.

Specifying the time period before takeover

You can specify how long (in seconds) a partner in an HA pair can be unresponsive before the other partner takes over.

About this task

Both partners do not need to have the same value for this option. Thus, you can have one partner that takes over more quickly than the other.

Note: If your HA pair is failing over because one of the nodes is too busy to respond to its partner, increase the value of the `cf.takeover.detection.seconds` option on the partner.

Step

1. Enter the following command:

   options cf.takeover.detection.seconds *number_of_seconds*

   The valid values for *number_of_seconds* are 10 through 180; the default is 15.

   Note: If the specified time is less than 15 seconds, unnecessary takeovers can occur, and a core might not be generated for some system panics. Use caution when assigning a takeover time of less than 15 seconds.

Enabling or disabling negotiated failover for a network interface

You can enable or disable negotiated failover for a network interface to trigger automatic takeover if the interface experiences a persistent failure. You can use the `nfo` option of the `ifconfig` command to enable or disable negotiated failover.

About this task

You can specify the `nfo` option for an interface group. However, you cannot specify the `nfo` option for any underlying physical interface of the interface group.

Steps

1. To enable takeover during interface failure, enter the following command:

   options cf.takeover.on_network_interface_failure on

2. To enable or disable negotiated failover, enter the following command:

ifconfig *interface_name* {nfo|-nfo}

interface_name is the name of the network interface.

`nfo` enables negotiated failover.

`-nfo` disables negotiated failover.

Example

To enable negotiated failover on the interface e8 of an HA configuration, enter the following command:

ifconfig e8 nfo

Takeover of vFiler units and the vFiler limit

The vFiler limit, set with the `vfiler limit` command, determines how many vFiler units can exist on a system. In an HA pair, if the two systems have different vFiler limits, some vFiler units might not be taken over if a takeover occurs.

When performing a takeover, a system can take over only the number of vFiler units that were specified by that system's vFiler limit. For example, if the limit is set to 5, the system can only take over five vFiler units from the partner. If the partner that is being taken over has a higher vFiler limit, some vFiler units cannot be taken over successfully.

For more information about setting the vFiler limit, see the *Data ONTAP MultiStore Management Guide for 7-Mode*.

Managing an HA pair in takeover mode

You manage an HA pair in takeover mode by performing a number of management actions.

Determining why takeover occurred

You can use the `cf status` command to determine why a takeover occurred.

Step

1. At the takeover prompt, enter the following command:

   cf status

Result

This command can display the following information:

- Whether controller failover is enabled or disabled
- Whether a takeover is imminent due to a negotiated failover
- Whether a takeover occurred, and the reason for the takeover

Statistics in takeover mode

Explains differences in system statistics when in takeover mode.

In takeover mode, statistics for some commands differ from the statistics in normal mode in the following ways:

- Each display reflects the sum of operations that take place on the takeover node plus the operations on the failed node.
  The display does not differentiate between the operations on the takeover node and the operations on the failed node.
- The statistics displayed by each of these commands are cumulative.
- After giving back the failed partner's resources, the takeover node does not subtract the statistics it performed for the failed node in takeover mode.
- The giveback does not reset (zero out) the statistics.
  To get accurate statistics from a command after a giveback, you can reset the statistics as described in the man page for the command you are using.

Note: You can have different settings on each node for SNMP options, but any statistics gathered while a node was taken over do not distinguish between nodes.

Managing emulated nodes

An emulated node is a software copy of the failed node that is hosted by the takeover node. You access the emulated node in partner mode by using the `partner` command.

Management exceptions for emulated nodes

The management of disks and array LUNs and some other tasks are different when you are managing an emulated node.

You manage an emulated node as you do any other storage system, including managing disks or LUNs, with the following exceptions, which are described in greater detail later in this section:

- An emulated node can access only its own disks or LUNs.
- Some commands are unavailable.
- Some displays differ from normal displays.

Accessing the emulated node from the takeover node

You access the emulated node from the takeover node in takeover mode with the `partner` command.

About this task

You can issue the `partner` command in two forms:

- Using the `partner` command without an argument

This toggles between *partner mode*, in which you manage the emulated node, and *takeover mode*, in which you manage the takeover node.

- Using the `partner` command with a Data ONTAP command as an argument

  This executes the command on the emulated node in partner mode and then returns to takeover mode.

Accessing the remote node using the partner command without arguments

You can use the `partner` command to toggle between the partner mode, in which commands are executed on the partner node, and takeover mode.

Step

1. From the takeover prompt, enter the following command:

   `partner`

Result

The prompt changes to the partner-mode prompt, which has the following form:
`emulated_node/takeover_node>`

Example of the change to partner mode

The following example shows the change from takeover mode to partner mode and back:

```
filer1(takeover)> partner
Login from console: filer2
Thu Aug 20 16:44:39 GMT [filer1: rc]: Login from console: filer2
filer2/filer1> partner
Logoff from console: filer2
filer1(takeover)> Thu Aug 20 16:44:54 GMT [filer1: rc]: Logoff from
console: filer2
filer1(takeover)>
```

Accessing the takeover node with the partner command containing arguments

You use the `partner` command with a Data ONTAP command as an argument, so you can execute single commands on the takeover node without entering partner mode.

Step

1. From the takeover prompt, enter the following command:

   `partner` ***`command`***

   `command` is the command you want to initiate on the emulated node.

Example of issuing the partner command with an argument

```
filer1(takeover)>partner cf status
filer2 has been taken over by filer1.
filer1(takeover)>
```

Accessing the emulated node remotely using Remote Shell

You can access the emulated node remotely using a Remote Shell (rsh) connection. You cannot access the emulated node using Secure Shell (ssh) or Telnet.

Step

1. Enter the following command:

   rsh failed_node command

   failed_node is the name of the failed node.

   command is the Data ONTAP command you want to run.

Example of an rsh command

In the following example, filer2 is the failed node:

```
rsh filer2 df
```

Emulated node command exceptions

Almost all the commands that are available to a takeover node are available on the emulated node. Some commands, however, are either unavailable or behave differently in emulated mode.

Unavailable commands

The following commands are not available on an emulated node:

- `cf disable`
- `cf enable`
- `cf forcegiveback`
- `cf forcetakeover`
- `cf giveback`
- `cf takeover`
- `date`
- `halt`
- `ifconfig partner`
- `ifconfig -partner`

- `ifconfig mtusize`
- `license cf`
- `rdate`
- `reboot`
- `timezone`

Commands with different behaviors

Command	Difference
`ifconfig` *`interface`*	• Displays emulated interface mappings based on the failed node's `/etc/rc` file rather than the takeover node interface mappings. • Displays emulated interface names rather than the interface names of the takeover node. • Displays only interfaces that have been configured, rather than all interfaces, configured or not, as displayed on the takeover node.
`mt`	Uses the tape devices on the takeover node because the failed node has no access to its tape devices.
`netstat -i`	Appends a plus sign (+) to shared interfaces. A shared interface is one that has two IP addresses assigned to it: an IP address for the node in which it physically resides and an IP address for its partner node in the HA pair.
`sysconfig`	When it displays hardware information, it displays information only about the hardware that is attached to the takeover node. It does not display information about the hardware that is attached only to the failed node. For example, the disk adapter information that the partner `sysconfig -r` command displays is about the disk adapters on the takeover node.
`uptime`	Displays how long the failed node has been down and the host name of the takeover node.

Command	Difference
`aggr status`	When it displays hardware information, it displays information only about the hardware that is attached to the takeover node. It does not display information about the hardware that is attached only to the failed node. For example, the disk adapter information that the partner `aggr status -r` command displays is about the disk adapters on the takeover node.

Performing dumps and restores for a failed node

You can use the emulated node and peripheral devices attached to the takeover node to perform dumps and restores for the failed node.

Before you begin

Any `dump` commands directed to the failed node's tape drives are executed on the takeover node's tape drives. Therefore, any `dump` commands that you execute using a scheduler, such as the `cron` command, succeed only under the following conditions:

- The device names are the same on both nodes in the HA pair.
- The `dump` commands for the takeover node and the emulated node are not scheduled to occur during the same time period; the takeover node and the emulated node cannot access the tape drives simultaneously.

About this task

Because the peripheral devices for a failed node are inaccessible, you perform dumps and restores for a failed node by using the emulated node (available using the `partner` command on the takeover node), making sure that you use a peripheral device attached to the takeover node.

For more information about performing dumps and restores, see the *Data Protection Tape Backup and Recovery Guide for 7-Mode*.

Step

1. Issue the `backup` or `restore` command, either in partner mode or as an argument in the `partner` command.

Example

Issuing a `restore` command in partner mode:

```
node1 (takeover)> partner
node1/node2> restore [options [arguments]]
node1 (takeover)> partner
```

Example

Issuing a `restore` command as an argument in the `partner` command:

```
node1 (takeover)> partner restore [options [arguments]]
```

Giveback operations

Giveback can be implemented and configured in a number of different ways. It can also be configured to occur automatically.

Performing a manual giveback

You can perform a normal giveback, a giveback in which you terminate processes on the partner node, or a forced giveback.

Note: Prior to performing a giveback, you must remove failed drives in the taken-over system, as described in the *Data ONTAP Storage Management Guide for 7-Mode*.

Option for shortening giveback time

You can shorten the client service outage during giveback by using the `cf.giveback.check.partner` option. You should always set this option to `on`.

Removing failed disks prior to attempting giveback

For taken-over systems that use disks, you must remove the failed disk or disks prior to attempting to implement giveback.

Step

1. Remove the failed disks, as described in the *Data ONTAP Storage Management Guide for 7-Mode*.

After you finish

When all failed disks are removed or replaced, proceed with the giveback operation.

Initiating normal giveback

You can return control to a taken-over partner with the `cf giveback` command.

Before you begin

For fabric-attached MetroCluster configurations, the aggregates on the surviving node and the partner node must already be rejoined to reestablish the MetroCluster configuration.

Step

1. Enter the following command on the command line of the takeover node:

   `cf giveback`

   Note: If the giveback fails, there might be a process running that prevents giveback. You can wait and repeat the command, or you can initiate giveback using the `-f` option to terminate the processes that are preventing giveback.

   After a giveback, the takeover node's ability to take over its partner automatically is not reenabled until the partner reboots successfully. If the partner fails to reboot, you can enter the `cf takeover` command to initiate a takeover of the partner manually.

Troubleshooting if giveback fails

If the `cf giveback` command fails, you should check for system processes that are currently running and might prevent giveback, check that the HA interconnect is operational, and check for any failed disks on systems using disks.

Steps

1. For systems using disks, check for and remove any failed disks, as described in the *Data ONTAP Storage Management Guide for 7-Mode*.
2. Check for the following message on the console:
   `cf.giveback.disk.check.fail`

   Both nodes should be able to detect the same disks. This message indicates that there is a disk mismatch: for some reason, one node is not seeing all the disks attached to the HA pair.
3. Check the HA interconnect and verify that it is correctly connected and operating.
4. Check whether any of the following processes were taking place on the takeover node at the same time you attempted the giveback:
   - Advanced mode repair operations, such as wafliron
   - Aggregate creation
   - Backup dump and restore operations
   - Disks being added to a volume (vol add)
   - Disk ownership assignment

- Disk sanitization operations
- Outstanding CIFS sessions
- Quota initialization
- RAID disk additions
- Snapshot copy creation, deletion, or renaming
- SnapMirror transfers (if the partner is a SnapMirror destination)
- SnapVault restorations
- Storage system panics
- Volume creation (traditional volume or FlexVol volume)

If any of these processes are taking place, either cancel the process or wait until it is complete, and then try the giveback operation.

5. If the `cf giveback` operation still does not succeed, use the `cf giveback -f` command to force giveback.

Related tasks

Forcing giveback on page 179

Forcing giveback

Because the takeover node might detect an error condition on the failed node that typically prevents a complete giveback such as data not being flushed from NVRAM to the failed node's disks, you can force a giveback, if necessary.

About this task

You can use this procedure to force the takeover node to give back the resources of the failed node even if the takeover node detects an error that typically prevents a complete giveback.

Note: The `cf forcegiveback` command should be used with caution because it can cause a loss of data. If you cannot risk loss of data and are unable to complete the giveback, contact technical support.

Steps

1. On the takeover node, enter the following command:

   `cf giveback -f`

   The `-f` parameter allows giveback to proceed as long as it would not result in data corruption or an error on the storage system.

2. If giveback is still not successful, and if you can risk possible loss of data, enter the following command on the takeover node:

   `cf forcegiveback`

Attention: Use `cf forcegiveback` only when you cannot get `cf giveback -f` to succeed. When you use this command, you risk losing any data committed to NVRAM but not to disk.

If a `cifs terminate` command is running, allow it to finish before forcing a giveback.

If giveback is interrupted

If the takeover node experiences a failure or a power outage during the giveback process, the giveback process stops and the takeover node returns to takeover mode until the failure is repaired or the power is restored.

Configuring giveback

You can configure how giveback occurs, setting different Data ONTAP options to improve the speed and timing of giveback.

Configuring automatic giveback

You can enable automatic giveback by using the `cf.giveback.auto.enable` command.

About this task

You should use the automatic giveback feature with care:

- Do not enable automatic giveback in MetroCluster configurations.
  Before the giveback operation is undertaken, you must rejoin the aggregates on the surviving node and the partner node to reestablish the MetroCluster configuration. If automatic giveback is enabled, this crucial step cannot be performed before the giveback.
- You should leave this option disabled unless your clients are unaffected by failover, or you have processes in place to handle repetitive failovers and givebacks.
- If an automatic takeover occurred because the partner node panicked, the default behavior is that an automatic giveback occurs even if this option is set to `off`.

Step

1. Enter the following command to enable automatic giveback:

   `option cf.giveback.auto.enable on`

   The `on` value enables automatic giveback. The `off` value disables automatic giveback. This option is `off` by default.

Related tasks

Enabling and disabling automatic giveback after takeover due to partner panicking on page 182

Adjusting the giveback delay time for automatic giveback

By default, there is a 600-second minimum time that a node stays in the takeover state before performing an automatic giveback. This delay reduces the overall outage that can occur while the

taken-over partner reboots. Instead of a single longer outage, there are two brief outages (first when the partner is taken over, the second when giveback occurs). This option affects all types of automatic giveback but does not affect manual giveback.

Step

1. Enter the following command:

   options cf.giveback.auto.delay.seconds *number of seconds*

   The valid values for *number_of_seconds* are 0 to 600. The default is 600.

   Attention: If `cf.giveback.auto.delay.seconds` is set to 0, the combined outage during takeover and giveback results in a long total client outage.

Setting giveback delay time for CIFS clients

You can specify the number of minutes to delay an automatic giveback before terminating CIFS clients that have open files.

About this task

This option specifies the number of minutes to delay an automatic giveback before terminating CIFS clients that have open files. During the delay, the system periodically sends notices to the affected clients. If you specify 0, CIFS clients are terminated immediately.

This option is used only if automatic giveback is On.

Step

1. Enter the following command:

   options cf.giveback.auto.cifs.terminate.minutes *minutes*

   Valid values for *minutes* are 0 through 999. The default is 5 minutes.

Terminating long-running processes to speed automatic giveback

You can use the `cf.giveback.auto.terminate.bigjobs` option to speed implementation of automatic giveback.

The `cf.giveback.auto.terminate.bigjobs` option, when on, specifies how automatic giveback handles long-running operations, such as dump or restore. When the option is set to `on`, long-running operations are terminated immediately when automatic giveback is initiated. When the option is set to `off`, automatic giveback is deferred until the long-running operations are complete. This option is used only if automatic giveback is enabled.

Setting giveback to terminate long-running processes

You can set the automatic giveback process to terminate long-running processes that might prevent the giveback.

Step

1. Enter the following command:

   options cf.giveback.auto.terminate.bigjobs {on|off}

   The `on` argument enables this option. The `off` argument disables this option. This option is `off` by default.

Enabling and disabling automatic giveback after takeover due to partner panicking

Data ONTAP is configured by default to initiate an automatic giveback after an automatic takeover that occurred due to the partner node panicking. This shortens the time between the initial failure and the full restoration of service.

About this task

- If you enter this command on one node, the value applies to both nodes.
- The setting of this option is persistent across reboots.
- This option is not affected by the setting of the `cf.giveback.auto.enable` option.
  If `cf.giveback.auto.enable` is set to OFF, automatic giveback after takeover due to panic still occurs if `cf.giveback.auto.after.panic.takeover` is set to ON.

Steps

1. Ensure that you enabled controller takeover by entering the following command:

   cf enable

2. Enable or disable automatic giveback after takeover on panic by entering the following command:

   options cf.giveback.auto.after.panic.takeover {on|off}

   `on` enables automatic giveback to the partner. This is the default value.

   `off` disables automatic giveback after takeover on panic.

Troubleshooting HA issues

If takeover or giveback fails for an HA pair, or you cannot enable HA, you need to check the HA status and proceed based on messages you receive.

Steps

1. Check communication between the local and partner nodes by entering the following command and observing the messages:

   ```
   cf status
   ```

2. Review the messages and take the appropriate action:

If the error message indicates...	Then...
A HA adapter error	Check the HA adapter cabling. Make sure that the cabling is correct and properly seated at both ends of the cable.
That the NVRAM adapter is in the wrong slot number	Check the NVRAM slot number. Move it to the correct slot if needed.
A Channel B cabling error	Check the cabling of the Channel B disk shelf loops and reseat and tighten any loose cables.
A networking error	Check for network connectivity. See the *Data ONTAP Network Management Guide for 7-Mode* for more information.

3. Reboot the HA pair and rerun the takeover and giveback tests.
4. If you still do not have takeover enabled, contact technical support.

Managing EXN1000, EXN2000, or EXN4000 unit disk shelves in an HA pair

You must follow specific procedures to add disk shelves to an HA pair or a MetroCluster configuration, or to upgrade or replace disk shelf hardware in an HA pair.

If your configuration includes SAS disk shelves, see the following documents on the N series support website (accessed and navigated as described in *Websites* on page 11):

- For SAS disk shelf management, see the *Hardware and Service Guide* for your disk shelf model.
- For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.
- For cabling SAS disk shelves in a MetroCluster configuration, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges*.

Adding EXN1000, EXN2000, or EXN4000 unit disk shelves to a multipath HA loop

To add supported EXN1000, EXN2000, or EXN4000 unit disk shelves to an HA pair configured for multipath HA, you need to add the new disk shelf to the end of a loop, ensuring that it is connected to the previous disk shelf and to the controller.

About this task

This procedure does not apply to SAS disk shelves.

Steps

1. Confirm that there are two paths to every disk by entering the following command:

   ```
   storage show disk -p
   ```

   Note: If two paths are not listed for every disk, this procedure could result in a data service outage. Before proceeding, address any issues so that all paths are redundant. If you do not have redundant paths to every disk, you can use the nondisruptive upgrade method (failover) to add your storage.

2. Install the new disk shelf in your cabinet or equipment rack, as described in the *EXN2000 and EXN4000 Hardware and Service Guide*.
3. Find the last disk shelf in the loop for which you want to add the new disk shelf.

   Note: The Channel A Output port of the last disk shelf in the loop is connected back to one of the controllers.

   Note: In Step 4 you disconnect the cable from the disk shelf. When you do this, the system displays messages about adapter resets and eventually indicates that the loop is down. These

messages are normal within the context of this procedure. However, to avoid them, you can optionally disable the adapter prior to disconnecting the disk shelf.

If you choose to, disable the adapter attached to the Channel A Output port of the last disk shelf by entering the following command:

```
fcadmin config -d <adapter>
```

<adapter> identifies the adapter by name. For example: 0a.

4. Disconnect the SFP and cable coming from the Channel A Output port of the last disk shelf.

   Note: Leave the other ends of the cable connected to the controller.

5. Using the correct cable for a shelf-to-shelf connection, connect the Channel A Output port of the last disk shelf to the Channel A Input port of the new disk shelf.
6. Connect the cable and SFP you removed in Step 4 to the Channel A Output port of the new disk shelf.
7. If you disabled the adapter in Step 3, reenable the adapter by entering the following command:

   ```
   fcadmin config -e <adapter>
   ```

8. Repeat Step 4 through Step 7 for Channel B.

   Note: The Channel B Output port is connected to the other controller.

9. Confirm that there are two paths to every disk by entering the following command:

   ```
   storage show disk -p
   ```

   Two paths should be listed for every disk.

Related tasks

Determining path status for your HA pair on page 187

Upgrading or replacing modules in an HA pair

In an HA pair with redundant pathing, you can upgrade or replace disk shelf modules without interrupting access to storage.

About this task

These procedures are for EXN1000, EXN2000, or EXN4000 unit disk shelves.

Note: If your configuration includes SAS disk shelves, refer to the following documents on the N series support website (accessed and navigated as described in *Websites* on page 11):

- For SAS disk shelf management, see the *Hardware and Service Guide* for your disk shelf model.
- For cabling SAS disk shelves in an HA pair, see the *Universal SAS and ACP Cabling Guide*.

- For cabling SAS disk shelves in a MetroCluster configuration, see *Configuring a MetroCluster system with SAS disk shelves and FibreBridge 6500N bridges*.

About the disk shelf modules

A disk shelf module (ESH4 or AT-FCX) in an EXN2000 or EXN1000 unit includes a SCSI-3 Enclosure Services Processor that maintains the integrity of the loop when disks are swapped and provides signal retiming for enhanced loop stability. When upgrading or replacing a module, you must be sure to cable the modules correctly.

The EXN2000 or EXN1000 unit disk shelves support the ESH4 or AT-FCX modules.

There are two modules in the middle of the rear of the disk shelf, one for Channel A and one for Channel B.

Note: The Input and Output ports on module B on the EXN2000 unit are the reverse of module A.

Restrictions for changing module types

If you plan to change the type of any module in your HA pair, make sure that you understand the restrictions.

You cannot mix ESH4 modules in the same loop with AT-FCX modules.

Best practices for changing module types

If you plan to change the type of any module in your HA pair, make sure that you review the best practice guidelines.

- Whenever you remove a module from an HA pair, you need to know whether the path you will disrupt is redundant.

  If it is, you can remove the module without interfering with the storage system's ability to serve data. However, if that module provides the only path to any disk in your HA pair, you must take action to ensure that you do not incur system downtime.
- When you replace a module, make sure that the replacement module's termination switch is in the same position as the module it is replacing.

  Note: ESH4 modules are self-terminating; this guideline does not apply to ESH4 modules.
- If you replace a module with a different type of module, make sure that you also change the cables, if necessary.

  For more information about supported cable types, see the hardware documentation for your disk shelf.
- Always wait 30 seconds after inserting any module before reattaching any cables in that loop.

Testing the modules

You should test your disk shelf modules after replacing or upgrading them to ensure that they are configured correctly and operating.

Steps

1. Verify that all disk shelves are functioning properly by entering the following command:

   `environ shelf`

2. Verify that there are no missing disks by entering the following command:

   `aggr status -r`

   Local disks displayed on the local node should be displayed as partner disks on the partner node, and vice-versa.

3. Verify that you can create and retrieve files on both nodes for each licensed protocol.

Determining path status for your HA pair

If you want to remove a module from your HA pair, you need to know whether the path you will disrupt is redundant. You can use the `storage show disk -p` command to indicate whether the disks have redundant paths.

About this task

If the disks have redundant paths, you can remove the module without interfering with the storage system's ability to serve data. On the other hand, if that module provides the only path to any of the disks in your HA pair, you must take action to ensure that you do not incur system downtime.

Step

1. Use the `storage show disk -p` command at your system console.

   This command displays the following information for every disk in the HA pair:

   - Primary port
   - Secondary port
   - Disk shelf
   - Bay

Examples for configurations with and without redundant paths

The following example shows what the `storage show disk -p` command output might look like for a redundant-path HA pair consisting of filers:

```
PRIMARY  PORT SECONDARY  PORT SHELF BAY
-------  ---- ---------  ---- --------
0c.112    A    0b.112     B     7    0
0b.113    B    0c.113     A     7    1
0b.114    B    0c.114     A     7    2
0c.115    A    0b.115     B     7    3
0c.116    A    0b.116     B     7    4
0c.117    A    0b.117     B     7    5
0b.118    B    0c.118     A     7    6
0b.119    B    0c.119     A     7    7
0b.120    B    0c.120     A     7    8
0c.121    A    0b.121     B     7    9
0c.122    A    0b.122     B     7   10
0b.123    B    0c.123     A     7   11
```

Notice that every disk (for example, 0c.112/0b.112) has two ports active: one for A and one for B. The presence of the redundant path means that you do not need to fail over one system before removing modules from the system.

Attention: Make sure that every disk has two paths. Even in an HA pair configured for redundant paths, a hardware or configuration problem can cause one or more disks to have only one path. If any disk in your HA pair has only one path, you must treat that loop as if it were in a single-path HA pair when removing modules.

The following example shows what the `storage show disk -p` command output might look like for an HA pair consisting of filers that do not use redundant paths:

```
filer1> storage show disk -p
PRIMARY  PORT SECONDARY  PORT SHELF BAY
-------  ---- ---------  ---- --------
5b.16     B                    1    0
5b.17     B                    1    1
5b.18     B                    1    2
5b.19     B                    1    3
5b.20     B                    1    4
5b.21     B                    1    5
5b.22     B                    1    6
5b.23     B                    1    7
5b.24     B                    1    8
5b.25     B                    1    9
5b.26     B                    1   10
5b.27     B                    1   11
5b.28     B                    1   12
5b.29     B                    1   13
5b.32     B                    2    0
5b.33     B                    2    1
5b.34     B                    2    2
```

```
5b.35     B                   2    3
5b.36     B                   2    4
5b.37     B                   2    5
5b.38     B                   2    6
5b.39     B                   2    7
5b.40     B                   2    8
5b.41     B                   2    9
5b.42     B                   2   10
5b.43     B                   2   11
5b.44     B                   2   12
5b.45     B                   2   13
```

For this HA pair, there is only one path to each disk. This means that you cannot remove a module from the configuration, thereby disabling that path, without first performing a takeover.

Hot-swapping a module

You can hot-swap a faulty disk shelf module, removing the faulty module and replacing it without disrupting data availability.

About this task

When you hot-swap a disk shelf module, you must ensure that you never disable the only path to a disk, which results in a system outage.

Attention: If there is newer firmware in the `/etc/shelf_fw` directory than that on the replacement module, the system automatically runs a firmware update. On non-multipath HA AT-FCX installations, multipath HA configurations running versions of Data ONTAP prior to 7.3.1, and non-RoHS modules, this firmware update causes a service interruption.

Steps

1. Verify that your storage system meets the minimum software requirements to support the disk shelf modules that you are hot-swapping.

   See the appropriate *Storage Expansion Unit Hardware and Service Guide* for more information.

2. Determine which loop contains the module you are removing, and determine whether any disks are single-pathed through that loop.

3. If any disks use this loop for their only path to a controller, complete the following steps:

   a) Follow the cables from the module you want to replace back to one of the nodes, called NodeA.

   b) At the NodeB console, enter the following command:

      `cf takeover`

   c) Wait for takeover to be complete and make sure that the partner node, or NodeA, reboots and is waiting for giveback.

Any module in the loop that is attached to NodeA can now be replaced.

4. Put on the antistatic wrist strap and grounding leash.
5. Disconnect the module that you are removing from the Fibre Channel cabling.
6. Using the thumb and index finger of both hands, press the levers on the CAM mechanism on the module to release it and pull it out of the disk shelf.
7. Slide the replacement module into the slot at the rear of the disk shelf and push the levers of the cam mechanism into place.

   Attention: Do not use excessive force when sliding the module into the disk shelf; you might damage the connector.

   Wait 30 seconds after inserting the module before proceeding to the next step.

8. Recable the disk shelf to its original location.
9. Check the operation of the new module by entering the following command from the console of the node that is still running:

   ```
   environ shelf
   ```

   The node reports the status of the modified disk shelves.

10. If you performed a takeover previously, complete the following steps:
    a) At the console of the takeover node, return control of NodeA's disk shelves by entering the following command:

       ```
       cf giveback
       ```

    b) Wait for the giveback to be completed before proceeding to the next step.
11. Test the replacement module.
12. Test the configuration.

Related concepts

Best practices for changing module types on page 186

Related tasks

Determining path status for your HA pair on page 187

Performing nondisruptive shelf replacement in a MetroCluster configuration

In a MetroCluster configuration, you can replace the EXN2000 unit and EXN4000 unit disk shelves nondisruptively. Performing nondisruptive shelf replacement (NDSR) involves preparing for the procedure, replacing disk shelves, and verifying the disk shelves after the shelf replacement.

About this task

For performing the nondisruptive shelf replacement procedure, the two MetroCluster nodes referred in the steps are mc-nodeA and mc-nodeB and the affected node is mc-nodeB. The disk shelf that requires replacement is part of the loop connected to the port 9 of the FC switches on mc-nodeB.

Steps

1. *Preparing for nondisruptive shelf replacement* on page 191
2. *Replacing the disk shelf nondisruptively* on page 192
3. *Verifying the disks after the shelf replacement* on page 194

Preparing for nondisruptive shelf replacement

You need to prepare the storage system before performing a nondisruptive shelf replacement procedure.

Before you begin

All disks on loops affected by a disk shelf must be mirrored.

Steps

1. Verify that all aggregates and volumes contained in the disks on the affected loop are mirrored and the mirroring is operational by using the `aggr status` and `sysconfig -r` commands at both the nodes.

   Example

   ```
   mc-nodeB> aggr status
   ```

   ```
   mc-nodeB> sysconfig -r
   ```

   ```
   mc-nodeA> aggr status
   ```

   ```
   mc-nodeA> sysconfig -r
   ```

2. Trigger an AutoSupport to indicate the start of the disk shelf replacement process by using the `options autosupport.doit` command at both the nodes.

   Example

   ```
   mc-nodeB> options autosupport.doit "SHELF REPLACE: START"
   ```

   ```
   mc-nodeA> options autosupport.doit "SHELF REPLACE: START"
   ```

3. Get the total number of the disks on both the nodes by using the `sysconfig` command.

   You should save this output for comparing with the output received after the disk shelf replacement.

4. Disable the automatic deletion of Snapshot copies on the aggregate of the affected node by using the `snapshot_autodelete` command.

   Example

   ```
   mc-nodeB> aggr options aggr1 snapshot_autodelete off
   ```

Replacing the disk shelf nondisruptively

After preparing your storage system to perform nondisruptive shelf replacement, you can replace the disk shelf.

Steps

1. Take the aggregate and plex on the mc-nodeB offline by using the `aggr offline` command.

   Example

   ```
   mc-nodeB> aggr offline aggr1/plex0
   ```

2. Disable switch port 9 on both the switches in the mc-nodeB by logging in as admin and using the `portdisable` command.

   Example

   ```
   mc-nodeB_sw0> portdisable 9
   ```

   ```
   mc-nodeB_sw2> portdisable 9
   ```

3. Wait until all disks missing notifications on both the nodes are complete and verify that the disks and shelves are no longer visible by using the `sysconfig` and `sysconfig -a` commands.

Example

```
mc-nodeB> sysconfig
```

```
mc-nodeB> sysconfig -a
```

```
mc-nodeA> sysconfig
```

```
mc-nodeA> sysconfig –a
```

4. Power off the disk shelf connected to port 9 of the switches in the mc-nodeB.
5. Remove disks from the shelf.

   You must ensure that you place the disks at a safe place.
6. Disconnect all FC and SFP cables from the disk shelf that are part of the loop connected to port 9.
7. Remove the disk shelf that are part of the loop connected to port 9 from the rack cabinet.
8. Remove module A (top slot), including SFP, from the disk shelf and insert into slot A (top slot) on replacement shelf.

   If required, replace the module and the SFP.
9. Remove module B (bottom slot), including SFP, from the disk shelf and insert into slot B (bottom slot) on replacement shelf.

   If required, replace the module and the SFP.
10. Insert the replacement shelf into rack cabinet.

    You must set the module speed and ensure that shelf ID is the same as the one that was replaced.
11. Cable all the connections again.

    If required, replace the cables.
12. Reconnect all SFP cables between the new disk shelf and the disk shelf of the same loop connected to port 9.

    Note: You must not perform this step if the switch port 9 has a single disk shelf loop.
13. Insert the disks removed in step 5 into replacement shelf only after shelf is replaced (completely installed with requisite install kit) into rack cabinet.
14. Power on disk shelf on the loop connected to port 9.

    Verify that power is up and no alarms are reported.

Verifying the disks after the shelf replacement

After replacing the disk shelf, you must perform certain steps to ensure that the disks are operational.

Steps

1. Enable the switch port on both the switches at mc-nodeB by using the `portenable` command.

   Example

   ```
   mc-nodeB_sw0> portenable 9
   ```

   ```
   mc-nodeB_sw2> portenable 9
   ```

2. Verify that all the disks and aggregates appear correctly in each offline plex by using the `aggr status`, `sysconfig -a`, and `sysconfig -r` commands at mc-nodeB.

   Example

   ```
   mc-nodeB> aggr status -r aggr1
   ```

   ```
   mc-nodeB> sysconfig –a
   ```

   ```
   mc-nodeB> sysconfig –r
   ```

3. Verify that the disks are active and online by using the `sysconfig` command at both the nodes.

   You should compare this output with the output generated before replacing the disk shelf to confirm that the same number of disks appears under each FC host adapter listed as active and online.

4. Take the aggregate and plex online on the mc-nodeB by using the `aggr online` command.

   Example

   ```
   mc-nodeB> aggr online aggr1/plex0
   ```

   Note: You should wait for all *`aggr1`* volumes on mc-nodeB to finish resyncing and return to a mirrored state. This step might take minutes or hours.

5. Enable the automatic deletion of Snapshot copies on all the aggregates in the mc-nodeB by using the `snapshot_autodelete` command.

Example

```
mc-nodeB> aggr options aggr1 snapshot_autodelete on
```

6. Trigger an AutoSupport from both the nodes to indicate the completion of the disk shelf replacement process.

Example

```
mc-nodeB> options autosupport.doit “SHELF REPLACE:FINISH”.
```

```
mc-nodeA> options autosupport.doit “SHELF REPLACE:FINISH”
```

Disaster recovery using MetroCluster configurations

In situations such as prolonged power outages or natural disasters, you can use the optional MetroCluster feature of Data ONTAP to provide a quick failover to another site that contains a nearly real time copy of the data at the disaster site.

Conditions that constitute a disaster

The disaster recovery procedure is an extreme measure that you should use only if the failure disrupts all communication from one MetroCluster site to the other for a prolonged period of time.

The following are examples of disasters that could cause such a failure:

- Fire
- Earthquake
- Prolonged power outages at a site
- Prolonged loss of connectivity from clients to the storage systems at a site

Ways to determine whether a disaster occurred

You should declare a disaster only after using predefined procedures to verify that service cannot be restored.

It is critical that you follow a predefined procedure to confirm that a disaster occurred. The procedure should include determining the status of the disaster site by:

- Using external interfaces to the storage system, such as the following:
  - `ping` command to verify network connectivity
  - Remote shell
  - FilerView administration tool
- Using network management tools to verify connectivity to the disaster site
- Physically inspecting the disaster site, if possible

You should declare a disaster only after verifying that service cannot be restored.

Failures that do not require disaster recovery

There are some failures that do not require any disaster recovery such as, a failure of the interconnect, failure of cables and so on. If you can reestablish the MetroCluster connection after fixing the problem, you should not perform the disaster recovery procedure.

You should not perform the disaster recovery procedure for the following failures:

- A failure of the HA interconnect between the two sites, which can be caused by the following:

  - Failure of the interconnect cable
  - Failure of one of the FC-VI adapters
  - If using switches, a failure of the SFP connecting a node to the switch

  With this type of failure, both nodes remain running. Automatic takeover is disabled because Data ONTAP cannot synchronize the NVRAM logs. After you fix the problem and reestablish the connection, the nodes resynchronize their NVRAM logs and the MetroCluster configuration returns to normal operation.
- The storage from one site (site A) is not accessible to the node at the other site (site B), which can be caused by the following:
  - Failure of any of the cables connecting the storage at one site to the node at the other site or switch
  - If using switches, failure of any of the SFPs connecting the storage to the switch or the node to the switch
  - Failure of the Fibre Channel adapter on the node
  - Failure of a storage disk shelf (disk shelf module; power; access to disk shelves; and so on)

  With this type of failure, you see a `mailbox disk invalid` message on the console of the storage system that cannot see the storage. After you fix the problem and reestablish the connection, the MetroCluster configuration returns to normal operation.
- If you are using switches, the inter-switch link between each pair of switches fails.
  With this type of failure, both nodes remain running. You see a `mailbox disk invalid` message because a storage system at one site cannot see the storage system at the other site. You also see a message because the two nodes cannot communicate with each other. After you fix the problem and reestablish the connection, the nodes resynchronize their NVRAM logs and the MetroCluster configuration returns to normal operation.
- If you are using FibreBridge 6500N bridge, a failure of the FibreBridge 6500N bridge.

Recovering from a disaster

After determining that there is a disaster, you should take steps to recover access to data, fix problems at the disaster site, and re-create the MetroCluster configuration.

About this task

Complete the following tasks in the order shown.

Attention: If for any reason the primary node has data that was not mirrored to the secondary node prior to the execution of the `cf forcetakeover -d` command, data could be lost. Do not resynchronize the original disks of the primary site for a SnapLock volume until an additional backup has been made of those disks to ensure availability of all data. This situation could arise, for example, if the link between the sites was down and the primary node had data written to it in the interim before the `cf forcetakeover -d` command was issued.

For more information about backing up data in SnapLock volumes using SnapMirror, see the *Data ONTAP Archive and Compliance Management Guide for 7-Mode*.

Steps

1. *Restricting access to the disaster site node* on page 198
2. *Forcing a node into takeover mode* on page 199
3. *Remounting volumes of the failed node* on page 199
4. *Recovering LUNs of the failed node* on page 200
5. *Fixing failures caused by the disaster* on page 201
6. *Reestablishing the MetroCluster configuration* on page 202

Restricting access to the disaster site node

You must restrict access to the disaster site node to prevent the node from resuming service. If you do not restrict access, you risk the possibility of data corruption.

About this task

You can restrict access to the disaster site node in the following ways:

- Turning off power to the disaster site node.
- Manually fencing off the node.

Steps

1. *Restricting access to the node by turning off power* on page 198
2. *Restricting access to the node by fencing off* on page 198

Restricting access to the node by turning off power

This is the preferred method for restricting access to the disaster site node. You can perform this task at the disaster site or remotely, if you have that capability.

Step

1. Switch off the power at the back of the storage system.

Restricting access to the node by fencing off

You can use manual fencing as an alternative to turning off power to the disaster site node. The manual fencing method restricts access using software and physical means.

Steps

1. Disconnect the HA interconnect and Fibre Channel adapter cables of the node at the surviving site.
2. Use the appropriate fencing method depending on the type of failover you are using:

If you are using...	Then fencing is achieved by...
Application failover	Using any application-specified method that either prevents the application from restarting at the disaster site or prevents the application clients from accessing the application servers at the disaster site. Methods can include turning off the application server, removing an application server from the network, or any other method that prevents the application server from running applications.
IP failover	Using network management procedures to ensure that the storage systems at the disaster site are isolated from the external public network.

Forcing a node into takeover mode

If a disaster has occurred, you can force the surviving node into takeover mode, so that the surviving node serves the data of the failed node.

Step

1. Enter the following command on the surviving node:

   ```
   cf forcetakeover -d
   ```

Result

Data ONTAP causes the following to occur:

- The surviving node takes over the functions of the failed node.
- The mirrored relationships between the two plexes of mirrored aggregates are broken, thereby creating two unmirrored aggregates.
  This is called splitting the mirrored aggregates.

The overall result of using the `cf forcetakeover -d` command is that a node at the surviving site is running in takeover mode with all the data in unmirrored aggregates.

Remounting volumes of the failed node

If the `cf.takeover.change_fsid` option is set to `on`, you must remount the volumes of the failed node because the volumes are accessed through the surviving node.

About this task

For more information about mounting volumes, see the *Data ONTAP File Access and Protocols Management Guide for Cluster-Mode*.

Note: You can disable the `change_fsid` option to avoid the necessity of remounting the volumes.

Steps

1. On an NFS client at the surviving site, create a directory to act as a mount point by entering the following command.

mkdir *directory_path*

Example

mkdir /n/toaster/home

2. Mount the volume by entering the following command.

mount *volume_name*

Example

mount toaster:/vol/vol0/home /n/toaster/home

Related tasks

Disabling the change_fsid option in MetroCluster configurations on page 142

Recovering LUNs of the failed node

You must actively track whether LUNs are online or offline in a MetroCluster configuration. If the `cf.takeover.change_fsid` option is set to `on`, and there is a disaster, all LUNs in the aggregates that were mirrored at the surviving site are offline. You can't determine if they were online prior to the disaster unless you track their state.

About this task

If you have a MetroCluster configuration, you must actively track the state of LUNs (track whether they are online or offline) on the node at each site. If there is a failure to a MetroCluster configuration that qualifies as a disaster and the node at one site is inaccessible, all LUNs in the aggregates that were mirrored at the surviving site are offline. There is no way to distinguish the LUNs that were offline before the disaster from the LUNs that were online before the disaster unless you have been tracking their status.

When you recover access to the failed node's LUNs, it is important to bring back online only the LUNs that were online before the disaster. To avoid igroup mapping conflicts, do not bring a LUN online if it was offline before the disaster. For example, suppose you have two LUNs with IDs of 5 mapped to the same igroup, but one of these LUNs was offline before the disaster. If you bring the previously offline LUN online first, you cannot bring the second LUN online because you cannot have two LUNs with the same ID mapped to the same host.

Note: You can disable the `change_fsid` option to avoid the necessity of remounting the volumes.

Steps

1. Identify the LUNs that were online before the disaster occurred.
2. Make sure that the LUNs are mapped to an igroup that contains the hosts attached to the surviving node.

For more information about mapping LUNs to igroups, see your *Data ONTAP SAN Administration Guide for 7-Mode*.

3. On the surviving node, enter the following command:

   lun online *lun-path* ...

   lun-path is the path to the LUN you want to bring online. You can specify more than one path to bring multiple LUNs online.

   Example

   lun online /vol/vol1/lun5

   Example

   lun online /vol/vol1/lun3 /vol/vol1/lun4

   Note: After you bring LUNs back online, you might have to perform some application or host-side recovery procedures. For example, the File System Identifiers (FSIDs) are rewritten, which can cause the LUN disk signatures to change. For more information, see the documentation for your application and for your host operating system.

Fixing failures caused by the disaster

You need to fix the failures caused by the disaster, if possible. For example, if a prolonged power outage to one of the MetroCluster sites caused the failure, restoring the power fixes the failure.

About this task

You cannot fix failures if the disaster causes a site to be destroyed. For example, a fire or an earthquake could destroy one of the MetroCluster sites. In this case, you fix the failure by creating a new partner for a MetroCluster configuration at a different site.

Step

1. Fix the failures at the disaster site.

After you finish

After the node at the surviving site can see the disk shelves at the disaster site, Data ONTAP renames the mirrored aggregates that were split, and the volumes they contain, by adding a number in parenthesis to the name. For example, if a volume name was vol1 before the disaster and the split, the renamed volume name could be vol1(1).

Reestablishing the MetroCluster configuration

You can reestablish a MetroCluster configuration after a disaster, depending on the state of the mirrored aggregate at the time of the takeover.

About this task

Depending on the state of a mirrored aggregate before you forced the surviving node to take over its partner, you use one of two procedures to reestablish the MetroCluster configuration:

- If the mirrored aggregate was in a normal state before the forced takeover, you can rejoin the two aggregates to reestablish the MetroCluster configuration.
  This is the most typical case.
- If the mirrored aggregate was in an initial resynchronization state (level-0) before the forced takeover, you cannot rejoin the two aggregates.
  You must re-create the synchronous mirror to reestablish the MetroCluster configuration.

Rejoining the mirrored aggregates to reestablish a MetroCluster configuration

Describes how to rejoin the mirrored aggregates if the mirrored aggregate was in a normal state before the forced takeover.

About this task

Attention: If you attempt a giveback operation prior to rejoining the aggregates, you might cause the node to boot with a previously failed plex, resulting in a data service outage.

Steps

1. Validate that you can access the remote storage by entering the following command:

   `aggr status -r`

2. Turn on power to the node at the disaster site.

   After the node at the disaster site boots, it displays the following message:
   `Waiting for Giveback...`

3. Determine which aggregates are at the surviving site and which aggregates are at the disaster site by entering the following command:

   `aggr status`

   Aggregates at the disaster site show plexes that are in a failed state with an out-of-date status. Aggregates at the surviving site show plexes as online.

4. If aggregates at the disaster site are online, take them offline by entering the following command for each online aggregate:

   `aggr offline` ***`disaster_aggr`***

   `disaster_aggr` is the name of the aggregate at the disaster site.

Note: An error message appears if the aggregate is already offline.

5. Re-create the mirrored aggregates by entering the following command for each aggregate that was split:

   aggr mirror *aggr_name* -v *disaster_aggr*

   aggr_name is the aggregate on the surviving site's node.

   disaster_aggr is the aggregate on the disaster site's node.

   The *aggr_name* aggregate rejoins the *disaster_aggr* aggregate to reestablish the MetroCluster configuration.

6. Verify that the mirrored aggregates have been re-created by entering the following command:

   aggr status -r mir

   The giveback operation only succeeds if the aggregates have been rejoined.

7. Enter the following command at the partner node:

   cf giveback

   The node at the disaster site reboots.

Example of rejoining aggregates

The following example shows the commands and status output when you rejoin aggregates to reestablish the MetroCluster configuration.

First, the aggregate status of the disaster site's storage after reestablishing access to the partner node at the surviving site is shown.

```
filer1> aggr status -r
Aggregate mir (online, normal) (zoned checksums)
  Plex /mir/plex5 (online, normal, active)
    RAID group /filer1/plex5/rg0 (normal)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity   8a.2   8a   0     2   FC:B  34500/70656000 35003/71687368
data     8a.8   8a   1     0   FC:B  34500/70656000 35003/71687368

Aggregate mir(1) (failed, out-of-date) (zoned checksums)
  Plex /mir(1)/plex1 (offline, normal, out-of-date)
    RAID group /mir(1)/plex1/rg0 (normal)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity   6a.0   6a   0     0   FC:B  34500/70656000 35003/71687368
data     6a.1   6a   0     1   FC:B  34500/70656000 35003/71687368

  Plex /mir(1)/plex5 (offline, failed, out-of-date)
```

Next, the mirror is reestablished using the `aggr mirror -v` command.

Note: The node at the surviving site is called filer1; the node at the disaster site is called filer2.

```
filer1> aggr mirror mir -v mir(1)
This will destroy the contents of mir(1).  Are you sure? y
Mon Nov 18 15:36:59 GMT [filer1:
raid.mirror.resync.snapcrtok:info]: mir: created mirror
resynchronization snapshot mirror_resync.1118153658(filer2)
Mon Nov 18 15:36:59 GMT [filer1: raid.rg.resync.start:notice]: /mir/
plex6/rg0: start resynchronization (level 1)
Mon Nov 18 15:36:59 GMT [filer1: raid.mirror.resync.start:notice]: /
mir: start resynchronize to target /mir/plex6
```

After the aggregates rejoin, the synchronous mirrors of the MetroCluster configuration are reestablished.

```
filer1> aggr status -r mir
Aggregate mir (online, mirrored) (zoned checksums)
  Plex /mir/plex5 (online, normal, active)
    RAID group /mir/plex5/rg0 (normal)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity   8a.2   8a   0     2   FC:B  34500/70656000 35003/71687368
data     8a.8   8a   1     0   FC:B  34500/70656000 35003/71687368

  Plex /mir/plex6 (online, normal, active)
    RAID group /mir/plex6/rg0 (normal)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity   6a.0   6a   0     0   FC:B  34500/70656000 35003/71687368
data     6a.1   6a   0     1   FC:B  34500/70656000 35003/71687368
```

Re-creating mirrored aggregates to return a MetroCluster configuration to normal operation

To return the MetroCluster configuration to normal operation, you must re-create the mirrored aggregates.

Steps

1. Validate that you can access the remote storage by entering the following command:

   `aggr status -r`

   Note: A (`level-0 resync in progress`) message indicates that a plex cannot be rejoined.

2. Turn on the power to the node at the disaster site.

   After the node at the disaster site boots up, it displays the following:

```
Waiting for giveback...(Press Ctrl-C to abort wait)
```

3. If the aggregates at the disaster site are online, take them offline by entering the following command for each aggregate that was split:

   aggr offline *disaster_aggr*

   disaster_aggr is the name of the aggregate at the disaster site.

   Note: An error message appears if the aggregate is already offline.

4. Destroy every target plex that is in a level-0 resync state by entering the following command:

   aggr destroy *plex_name*

5. Re-create the mirrored aggregates by entering the following command for each aggregate that was split:

   aggr mirror *aggr_name* -v *disaster_aggr*

   aggr_name is the aggregate on the surviving site's node.

   disaster_aggr is the aggregate on the disaster site's node.

   The *aggr_name* aggregate rejoins the *disaster_aggr* aggregate to reestablish the MetroCluster configuration.

6. Enter the following command at the partner node:

   cf giveback

   The node at the disaster site reboots.

Example of re-creating a mirrored aggregate

The following example shows the commands and status output when re-creating aggregates to reestablish the MetroCluster configuration.

The following output shows the aggregate status of the disaster site's storage after reestablishing access to the partner at the surviving site:

```
filer1>aggr status -r
Aggregate mir1 (online, normal) (zoned checksums)
  Plex /mir1/plex0 (online, normal, active)
  RAID group /mir1/plex0/rg0 (normal)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity   8a.3   8a   0     3   FC:B  34500/70656000 35003/71687368
data     8a.4   8a   0     4   FC:B  34500/70656000 35003/71687368
data     8a.6   8a   0     6   FC:B  34500/70656000 35003/71687368
data     8a.5   8a   0     5   FC:B  34500/70656000 35003/71687368

Aggregate mir1(1) (failed, partial) (zoned checksums)
  Plex /mir1(1)/plex0 (offline, failed, inactive)
```

```
  Plex /mir1(1)/plex6 (online, normal, resyncing)
    RAID group /mir1(1)/plex6/rg0 (level-0 resync in progress)

RAID Disk Device HA  SHELF BAY CHAN  Used (MB/blks) Phys (MB/blks)
--------- ------ ------------- ----- -------------- -------------
parity    6a.6   6a  0     6   FC:B  34500/70656000 35003/71687368
data      6a.2   6a  0     2   FC:B  34500/70656000 35003/71687368
data      6a.3   6a  0     3   FC:B  34500/70656000 35003/71687368
data      6a.5   6a  0     5   FC:B  34500/70656000 35003/71687368
```

The mir1(1)/plex6 plex shows that a level-0 resynchronization was in progress; therefore, an attempt to rejoin the plexes fails, as shown in the following output:

```
filer1> aggr mirror mir1 -v mir1(1)
aggr mirror: Illegal mirror state for aggregate 'mir1(1)'
```

Because the mir1(1)/plex6 plex had a level-0 resynchronization in progress, the mir1(1) aggregate must be destroyed and the mir aggregate remirrored to reestablish a synchronous mirror, as shown in the following output:

```
filer1> aggr mirror mir1 -v mir1(1)
aggr mirror: Illegal mirror state for aggregate 'mir1(1)'
filer1> aggr destroy mir1(1)
Are you sure you want to destroy this aggregate? y
Aggregate 'mir1(1)' destroyed.
filer1> aggr mirror mir1
Creation of a mirror plex with 4 disks has been initiated. The
disks need to be zeroed before addition to the aggregate. The
process has been initiated and you will be notified via the system
log as disks are added.
```

Where to find procedures for nondisruptive operations with HA pairs

By taking advantage of an HA pair's takeover and giveback operations, you can change hardware components and perform software upgrades in your configuration without disrupting access to system storage. You can refer to the specific documents for the required procedures.

You can perform nondisruptive operations on a system by having its partner take over the system's storage, performing maintenance, and then giving back the storage. Use the specific procedures as shown in the following table:

If you want to perform this task nondisruptively...	**See the...**
Upgrade Data ONTAP	*Data ONTAP Upgrade and Revert/Downgrade Guide for 7-Mode*
Replace a hardware FRU component	FRU procedures for your platform

Controller failover and single-points-of-failure

A storage system has a variety of single-points-of-failure that you can reduce by using the controller failover operations available in an HA pair. In an HA pair, there are a number of failures that can cause a controller to fail over.

Single-point-of-failure definition

A single-point-of failure (SPOF) represents the failure of a single hardware component that can lead to loss of data access or potential loss of data.

SPOF does not include multiple/rolling hardware errors, such as triple disk failure, dual disk shelf module failure, and so on.

All hardware components included with your storage system have demonstrated very good reliability with low failure rates. If a hardware component fails, such as a controller or adapter, you can use controller failover to provide continuous data availability and preserve data integrity for client applications and users.

SPOF analysis for HA pairs

Different individual hardware components and cables in the storage system are SPOFs, but an HA configuration can eliminate these SPOFs to improve data availability.

Hardware components	SPOF		How controller failover eliminates SPOF
	Stand-alone	HA pair	
Controller	Yes	No	If a controller fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
NVRAM	Yes	No	If an NVRAM adapter fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
CPU fan	Yes	No	If the CPU fan fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.

Hardware components	**SPOF**		**How controller failover eliminates SPOF**
	Stand-alone	**HA pair**	
Multiple NICs with interface groups (virtual interfaces)	Maybe, if all NICs fail	No	If one of the networking links within an interface group fails, the networking traffic is automatically sent over the remaining networking links on the same node. No failover is needed in this situation. If all the NICs in an interface group fail, the node automatically fails over to its partner node if failover is enabled for the interface group.
Single NIC	Yes	No	If a NIC fails, the node automatically fails over to its partner node if failover is enabled for the NIC.
FC-AL adapter or SAS HBA	Yes	No	If an FC-AL adapter for the primary loop fails for a configuration without multipath HA, the partner node attempts a takeover at the time of failure. With multipath HA, no takeover is required. If the FC-AL adapter for the secondary loop fails for a configuration without multipath HA, the failover capability is disabled, but both nodes continue to serve data to their respective applications and users, with no impact or delay. With multipath HA, failover capability is not affected.
FC-AL or SAS cable (controller-to-shelf, shelf-to-shelf)	No, if dual-path cabling is used	No	If an FC-AL loop or SAS stack breaks in a configuration that does not have multipath HA, the break could lead to a failover, depending on the shelf type. The partnered nodes invoke the negotiated failover feature to determine which node is best for serving data, based on the disk shelf count. When multipath HA is used, no failover is required.
Disk shelf module	No, if dual-path cabling is used	No	If a disk shelf module fails in a configuration that does not have multipath HA, the failure could lead to a failover. The partnered nodes invoke the negotiated failover feature to determine which node is best for serving data, based on the disk shelf count. When multipath HA is used, there is no impact.

Hardware components	SPOF		How controller failover eliminates SPOF
	Stand-alone	HA pair	
Disk drive	No	No	If a disk fails, the node can reconstruct data from the RAID4 parity disk. No failover is needed in this situation.
Power supply	Maybe, if both power supplies fail	No	Both the controller and disk shelf have dual power supplies. If one power supply fails, the second power supply automatically kicks in. No failover is needed in this situation. If both power supplies fail, the node automatically fails over to its partner node, which serves data for both nodes.
Fan (controller or disk shelf)	Maybe, if both fans fail	No	Both the controller and disk shelf have multiple fans. If one fan fails, the second fan automatically provides cooling. No failover is needed in this situation. If both fans fail, the node automatically fails over to its partner node, which serves data for both nodes.
HA adapter	N/A	No	If a HA adapter fails, the failover capability is disabled but both nodes continue to serve data to their respective applications and users.
HA interconnect cable	N/A	No	The HA interconnect adapter supports dual HA interconnect cables. If one cable fails, the heartbeat and NVRAM data are automatically sent over the second cable with no delay or interruption. If both cables fail, the failover capability is disabled but both nodes continue to serve data to their respective applications and users.

Failover event cause-and-effect table

Failover events cause a controller failover in HA pairs. The configuration responds differently depending on the event and the type of HA pair.

Cause-and-effect table for standard or mirrored HA pairs

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	Standard or mirrored HA pair
Single disk failure	No	No	Yes	Yes
Double disk failure (2 disks fail in same RAID group)	Yes, unless you are using SyncMirror or RAID-DP, then no.	Maybe. If root volume has double disk failure, or if the mailbox disks are affected, no failover is possible.	No, unless you are using RAID-DP or SyncMirror, then yes.	No, unless you are using RAID-DP or SyncMirror, then yes.
Triple disk failure (3 disks fail in same RAID group)	Maybe. If SyncMirror is being used, no takeover occurs; otherwise, yes.	Maybe. If root volume has triple disk failure, no failover is possible.	No	No
Single HBA (initiator) failure, Loop A	Maybe. If SyncMirror or multipath HA is in use, then no; otherwise, yes.	Maybe. If root volume has double disk failure, no failover is possible.	Yes, if multipath HA or SyncMirror is being used.	Yes, if multipath HA or SyncMirror is being used, or if failover succeeds.

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	Standard or mirrored HA pair
Single HBA (initiator) failure, Loop B	No	Yes, unless you are using SyncMirror or multipath HA and the mailbox disks are not affected, then no.	Yes, if multipath HA or SyncMirror is being used.	Yes, if multipath HA or SyncMirror is being used, or if failover succeeds.
Single HBA initiator failure (both loops at the same time)	Yes, unless the data is mirrored on a different (up) loop or multipath HA is in use, then no takeover needed.	Maybe. If the data is mirrored or multipath HA is being used and the mailbox disks are not affected, then no; otherwise, yes.	No, unless the data is mirrored or multipath HA is in use, then yes.	No failover needed if data is mirrored or multipath HA is in use.
AT-FCX failure (Loop A)	Only if multidisk volume failure or open loop condition occurs, and neither SyncMirror nor multipath HA is in use.	Maybe. If root volume has double disk failure, no failover is possible.	No	Yes, if failover succeeds.
AT-FCX failure (Loop B)	No	Maybe. If SyncMirror or multipath HA is in use, then no; otherwise, yes.	Yes, if multipath HA or SyncMirror is in use.	Yes

Event	**Does the event trigger failover?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	**Is data still available on the affected volume after the event?**	
			Single storage system	**Standard or mirrored HA pair**
IOM failure (Loop A)	Only if multidisk volume failure or open loop condition occurs, and neither SyncMirror nor multipath HA is in use.	Maybe. If root volume has double disk failure, no failover is possible.	No	Yes, if failover succeeds.
IOM failure (Loop B)	No	Maybe. If SyncMirror or multipath HA is in use, then no; otherwise, yes.	Yes, if multipath HA or SyncMirror is in use.	Yes
Shelf (backplane) failure	Only if multidisk volume failure or open loop condition occurs, and data isn't mirrored.	Maybe. If root volume has double disk failure or if the mailboxes are affected, no failover is possible.	Maybe. If data is mirrored, then yes; otherwise, no.	Maybe. If data is mirrored, then yes; otherwise, no.
Shelf, single power failure	No	No	Yes	Yes
Shelf, dual power failure	Only if multidisk volume failure or open loop condition occurs and data is not mirrored.	Maybe. If root volume has double disk failure, or if the mailbox disks are affected, no failover is possible.	Maybe. If data is mirrored, then yes; otherwise, no.	Maybe. If data is mirrored, then yes; otherwise, no.
Controller, single power failure	No	No	Yes	Yes

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	Standard or mirrored HA pair
Controller, dual power failure	Yes	Yes, until power is restored.	No	Yes, if failover succeeds.
HA interconnect failure (1 port)	No	No	n/a	Yes
HA interconnect failure (both ports)	No	Yes	n/a	Yes
Ethernet interface failure (primary, no interface group)	Yes, if set up to do so.	No	Yes	Yes
Ethernet interface failure (primary, interface group)	Yes, if set up to do so.	No	Yes	Yes
Ethernet interface failure (secondary, interface group)	Yes, if set up to do so.	No	Yes	Yes

Event	**Does the event trigger failover?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	**Is data still available on the affected volume after the event?**	
			Single storage system	**Standard or mirrored HA pair**
Ethernet interface failure (interface group, all ports)	Yes, if set up to do so.	No	Yes	Yes
Tape interface failure	No	No	Yes	Yes
Heat exceeds permissible amount	Yes	No	No	No
Fan failures (disk shelves or controller)	No	No	Yes	Yes
Reboot	Yes	No	Maybe. Depends on cause of reboot.	Maybe. Depends on cause of reboot.
Panic	Yes	No	Maybe. Depends on cause of panic.	Maybe. Depends on cause of panic.

Cause-and-effect table for stretch and fabric-attached MetroClusters

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Stretch MetroCluster	Fabric Attached MetroCluster
Single disk failure	No	No	Yes	Yes
Double disk failure (2 disks fail in same RAID group)	No	No	Yes	Yes, with no failover necessary.
Triple disk failure (3 disks fail in same RAID group)	No	No	No	Yes, with no failover necessary.
Single HBA (initiator) failure, Loop A	No	No	Yes	Yes, with no failover necessary.
Single HBA (initiator) failure, Loop B	No	No	Yes	Yes, with no failover necessary.
Single HBA initiator failure, (both loops at the same time)	No	No	Yes, with no failover necessary.	Yes, with no failover necessary.
AT-FCX failure (Loop A)	No	No	Yes	Yes, with no failover necessary.
AT-FCX failure (Loop B)	No	No	Yes	Yes

Event	**Does the event trigger failover?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	**Is data still available on the affected volume after the event?**	
			Stretch MetroCluster	**Fabric Attached MetroCluster**
Shelf (backplane) failure	No	No	Yes	Yes, with no failover necessary.
Shelf, single power failure	No	No	Yes	Yes
Shelf, dual power failure	No	No	Yes	Yes, with no failover necessary.
Controller, single power failure	No	No	Yes	Yes
Controller, dual power failure	Yes	Yes, until power is restored.	Yes, if failover succeeds.	Yes, if failover succeeds.
HA interconnect failure (1 port)	No	No	Yes	Yes
HA interconnect failure (both ports)	No	No; failover is possible.	Yes	Yes
Ethernet interface failure (primary, no interface group)	Yes, if set up to do so.	No	Yes	Yes

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Stretch MetroCluster	Fabric Attached MetroCluster
Ethernet interface failure (primary, interface group)	Yes, if set up to do so.	No	Yes	Yes
Ethernet interface failure (secondary, interface group)	Yes, if set up to do so.	No	Yes	Yes
Ethernet interface failure (interface group, all ports)	Yes, if set up to do so.	No	Yes	Yes
Tape interface failure	No	No	Yes	Yes
Heat exceeds permissible amount	Yes	No	No	No
Fan failures (disk shelves or controller)	No	No	Yes	Yes
Reboot	Yes	No	Maybe. Depends on cause of reboot.	Maybe. Depends on cause of reboot.
Panic	Yes	No	Maybe. Depends on cause of panic.	Maybe. Depends on cause of panic.

Feature update record

Provides a record of the history of changes made to this guide. When a change is implemented, it applies to the release in which it was implemented and all subsequent releases, unless otherwise specified.

Feature updates	Feature first implemented in	Feature release date
• Incorporation of the Cluster Administration chapter from the *Data ONTAP System Administration Guide for 7-Mode* and the *Disaster Protection Using MetroCluster* appendix from the *Data ONTAP Data Protection Online Backup and Recovery Guide*.	Data ONTAP 7.1	June 2005
• Updated MetroCluster information for N5000 series	Data ONTAP 7.1	October 2005
• Updated module replacement information • Fixed problem in Brocade switch configuration information	Data ONTAP 7.1	December 2005
• Updated and extended HA pair information • Moved Brocade switch configuration to Brocade Switch Description Page. • Moved from *cluster* to *active/active configuration* • Added information about Multipath Storage for HA pairs	Data ONTAP 7.1.1	June 2006

Feature updates	Feature first implemented in	Feature release date
• Generalized standard and mirrored HA pair cabling instructions • Updated standard and mirrored HA pair cabling instructions to include N7000 series • Changed name of document from *Cluster Installation and Administration Guide* to *Active/Active Configuration Guide*. • Added N7000 series information • Updated and extended HA pair configuration information • Moved Brocade switch configuration to Brocade Switch Description Page. • Moved from *cluster* to *active/active configuration*	Data ONTAP 7.2	May 2006
• Added information about Multipath Storage for HA pairs.	Data ONTAP 7.2.1	November 2006

Feature updates	Feature first implemented in	Feature release date
• Added quad-port, 4-Gb Fibre Channel HBA, ESH4 module, EXN4000 disk shelf information • Added information to explain that automatic giveback should not be used in MetroClusters • Updated Multipath Storage information • Updated MetroCluster disaster recovery information • Corrected failover and single-point-of-failure table	Data ONTAP 7.2.2	March 2007
• Added procedures for configuring fabric-attached MetroClusters on systems using software-based disk management • Added procedure for unconfiguring an active/active pair and returning to stand-alone operation	Data ONTAP 7.2.3	June 2007
• Added support for 504 disks in MetroClusters • Added support for the N7700 and N7900 systems • Added support for the `change_fsid` option • Added procedure for removing an HA pair	Data ONTAP 7.2.4	November 2007

Feature updates	Feature first implemented in	Feature release date
• Changed name of document from *Active/Active Configuration Guide* to *High-Availability Configuration Guide* • Moved from *active/active configuration* to *High Availability pair* • Added information about configuration on gateway systems • Added support for 672 disks in MetroClusters • Added MetroCluster support for the Brocade 300 and 5100 switches	Data ONTAP 8.0 RC1	
• Added references to the EXN3000 disk shelf documentation	Data ONTAP 8.0 GA	
• Added support for N6210 and N6240 systems • Added support for the 8-Gbps FC-VI adapter on MetroCluster configurations	Data ONTAP 8.0.1	November 2010
• Added support for N6210 and N6240 systems • Added support for the 8-Gbps FC-VI adapter on MetroCluster configurations	Data ONTAP 7.3.5	October 2010
• Added support for N6270 systems	Data ONTAP 8.0.1	March 2011
• Added support for N6270 systems	Data ONTAP 7.3.5	March 2011

Feature updates	Feature first implemented in	Feature release date
• Added more SAS disk shelf information and references to SAS disk shelf documentation. • Updated multipath HA as a requirement. • Removed procedures for non-multipath HA cabling. • Replaced the term *Multipath Storage* with *multipath HA* for consistency with other documentation.	Data ONTAP 8.0.2 Data ONTAP 7.3.5.1	May 2011
• Added support for MetroCluster with SAS disk shelves. • Added support for MetroCluster shared-switches configurations. • Added procedures for MetroCluster installation with third-party storage. • Added support for N3220 and N3240 systems.	Data ONTAP 8.1 RC1	September 2011
• Removed references to ESH2 and LRC modules	Data ONTAP 8.1 GA	December 2011
• Changed title to *Data ONTAP High Availability and MetroCluster Configuration Guide for 7-Mode*. • Added MetroCluster support for Cisco switches.	Data ONTAP 8.1.1 RC1	May 2012
Added support for automatic giveback after takeover-on-panic.	Data ONTAP 8.1.2 RC1	October 2012

Copyright and trademark information

References in this documentation to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of IBM's or NetApp's intellectual property rights may be used instead of the IBM or NetApp product, program, or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM and NetApp, are the user's responsibility.

No part of this document covered by copyright may be reproduced in any form or by any means— graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT

Trademark information

Builder, Shadow Tape, Simplicity, Simulate ONTAP, SnapCopy, SnapDirector, SnapDrive, SnapFilter, SnapLock, SnapManager, SnapMigrator, SnapMirror, SnapMover, SnapProtect, SnapRestore, Snapshot, SnapSuite, SnapValidator, SnapVault, StorageGRID, StoreVault, the StoreVault logo, SyncMirror, Tech OnTap, The evolution of storage, Topio, vFiler, VFM, Virtual File Manager, VPolicy, WAFL, Web Filer, and XBB are trademarks or registered trademarks of NetApp, Inc. in the United States, other countries, or both.

All other brands or products are trademarks or registered trademarks of their respective holders and should be treated as such.

NetApp, Inc. is a licensee of the CompactFlash and CF Logo trademarks.

NetApp, Inc. NetCache is certified RealSystem compatible.

Notices

This information was developed for products and services offered in the U.S.A.

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 on 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 give you any license to these patents. You can send license inquiries, in writing to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, N.Y. 10504-1785
U.S.A.

For additional information, visit the web at:
http://www.ibm.com/ibm/licensing/contact/

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

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 states 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 web sites are provided for convenience only and do not in any manner serve as an endorsement of those web sites. The materials at those web sites are not part of the materials for this IBM product and use of those web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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.

If you are viewing this information in softcopy, the photographs and color illustrations may not appear.

Index

A

B

C

D

E

F

G

H

I

L

M

N

O

P

R

S

T

U

V

Z

NA 210-05801_A0, Printed in USA

GA32-1036-03