IBM System Storage N series

Data ONTAP 7.3 Active/Active Configuration Guide

GC27-2208-08

Contents

About this guide

Note: 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, which is accessed and navigated as described in *Websites* on page 13.

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 IBM N series support website 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 IBM N series support website, which is accessed and navigated as described in *Websites* on page 13.

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 IBM 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, it is recommended that you run the latest level of firmware. Any firmware updates are posted to the IBM N series support website, which is accessed and navigated as described in *Websites* on page 13.

Note: If you do not see new firmware updates on the IBM 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 e-mail 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

Active/active configuration types and requirements

There are four types of active/active configurations, each having different advantages and requirements.

Overview of active/active configurations

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

What an active/active configuration is

An active/active configuration 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.

You can configure the active/active configuration 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 and subnets.

Depending on the model, nodes are connected to each other through an NVRAM adapter, gigabit Ethernet connections, or, in the case of systems with two controllers in a single chassis, through an internal 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).

Nondisruptive operations and fault tolerance with active/active configurations

Active/active configurations provide fault tolerance and the ability to perform nondisruptive upgrades and maintenance.

Configuring storage systems in an active/active configuration provides the following benefits:

- Fault tolerance
  When one node fails or becomes impaired a takeover occurs, and the partner node continues to serve the failed node's data.
- Nondisruptive software upgrades
  When you halt one node and allow takeover, the partner node continues to serve data for the halted node while you upgrade the node you halted.
- Nondisruptive hardware maintenance
  When you halt one node and allow takeover, the partner node continues to serve data for the halted node while you replace or repair hardware in the node you halted.

Related concepts

Nondisruptive operations with active/active configurations on page 175
Managing EXN1000, EXN2000, or EXN400 units in an active/active configuration on page 155

Characteristics of nodes in an active/active configuration

To configure and manage nodes in an active/active configuration, you should be familiar with the characteristics of active/active configurations.

- The controllers in the active/active configuration are connected to each other either through a cluster 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 do 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.
  - Each node in takeover mode manages its partner's disks or array LUNs. For third-party storage, the partner node takes over read/write access to the array LUNs owned by the failed node until the failed node becomes available again.

  Note: For systems using software-based disk ownership, 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:
  - Two if it is an N-series system.
  - One if it is a gateway system.

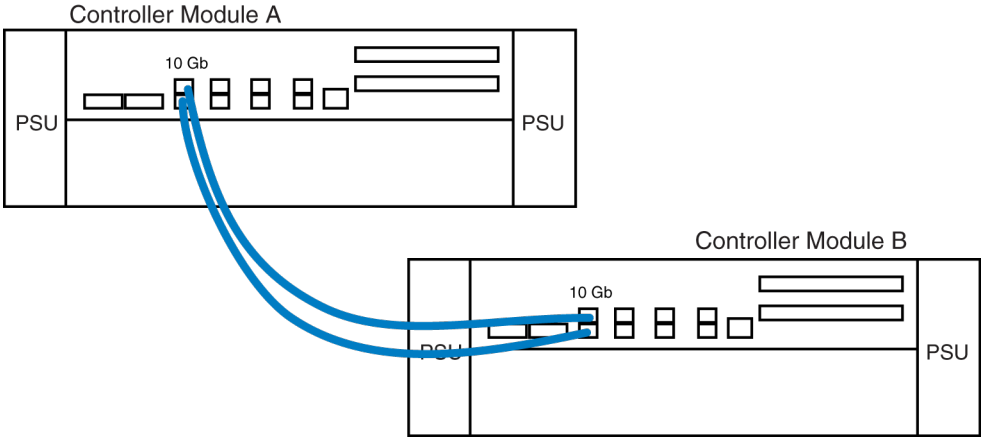
  The mailbox disks or LUNs are used to do the following tasks:
  - Maintain consistency between the pair
  - Continually check whether the other node is running or whether it has performed a takeover
  - Store configuration information that is not specific to any particular node
- They can reside on the same Windows domain or on different domains.

Systems with variable HA configurations

N6200 series systems can be physically configured as single-chassis active/active configurations in which both controller modules are in the same chassis, or as dual-chassis active/active configurations in which the controller modules are in separate chassis.

The following example shows a single-chassis active/active configuration:

The following example shows a dual-chassis active/active configuration and the HA interconnect cables:



Interconnect cabling for systems with variable HA configurations

Single-chassis and dual-chassis active/active configurations have different interconnect cabling requirements.

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

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

The HA state on systems with variable HA configurations

N6200 series controller modules and chassis automatically record whether they are in an active/active configuration or stand-alone. This record is the *HA state* and must be the same on all components within the stand-alone system or active/active configuration. The HA state can be manually configured if necessary.

Other systems recognize that they are in an active/active configuration by the physical presence of an NVRAM adapter in a specific slot or, for systems with an internal interconnect, the presence of a

controller module in the second bay of the chassis. Because N6200 series systems can have different physical HA configurations, they cannot rely on that method.

The HA state is recorded in the hardware PROM in the chassis and in the controller module.

The HA state can be one of the following:

ha	Indicates that the controller module or chassis is in an active/active configuration.
non-ha	Indicates that the controller module or chassis is in a stand-alone system.
default	Indicates that the controller module or chassis has not yet been inserted in a system. After being installed using the proper procedures, the controller module or chassis is automatically assigned the correct state depending on the state of the rest of the system.

The HA state must be consistent across all components of the system, as shown in the following table:

If the system or systems are...	The HA state is recorded on these components:	The HA state on the components must be:
Stand-alone	The chassis Controller module A	non-ha
In a single-chassis active/active configuration	The chassis Controller module A Controller module B	ha
In a dual-chassis active/active configuration	Chassis A Controller module A Chassis B Controller module B	ha

Related concepts

Systems with two controller modules in the same chassis on page 45

Related tasks

Verifying and setting the HA state of N6200 series controller modules and chassis on page 113

If the HA state becomes inconsistent

The HA state is maintained automatically, but if for some reason it becomes inconsistent between the chassis and controller modules, an EMS message is displayed.

In that case, you can use the `ha-config show` and `ha-config modify` commands to verify and set the HA state for the controller module or modules and chassis.

Best practices for active/active configurations

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

- Make sure that the controllers and disk shelves are on different power supplies or grids, so that a single power outage does not affect both components.
- Use VIFs (virtual interfaces) to provide redundancy and improve availability of network communication.
- Follow the documented procedures in the *Data ONTAP Upgrade Guide* when upgrading your active/active configuration.
- 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 (via an RLM or Service Processor), make sure you configure it properly, as described in the *Data ONTAP System Administration Guide* .
- Higher numbers of traditional and FlexVol volumes on your system can affect takeover and giveback times.

  When adding traditional or FlexVol volumes to an active/active configuration, 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*.
- Multipath HA is required on all active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections.

Comparison of active/active configuration types

The different types of active/active configurations support different capabilities for data duplication, distance between nodes, and failover.

Active/active configuration type	Data duplication?	Distance between nodes	Failover possible after loss of entire node (including storage)?	Notes
Standard active/active configuration	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 active/active configuration	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 active/active configuration.
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). MetroCluster configurations do not support SAS disk shelves.

Active/active configuration type	Data duplication?	Distance between nodes	Failover possible after loss of entire node (including storage)?	Notes
Fabric-attached MetroCluster	Yes	Up to 100 kilometers, depending on switch configuration. For filers, see the *Brocade Switch Configuration Guide for Fabric-attached MetroClusters*. For gateway systems, up to 30 km.	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. Fabric-attached MetroCluster configurations do not support SAS disk shelves.

Standard active/active configurations

Standard active/active configurations 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 active/active configurations on page 178

How Data ONTAP works with standard active/active configurations

In a standard active/active configuration, 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), 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 may indicate that the partner could not be reached. Wait until the partner has fully rebooted and reissue the command.

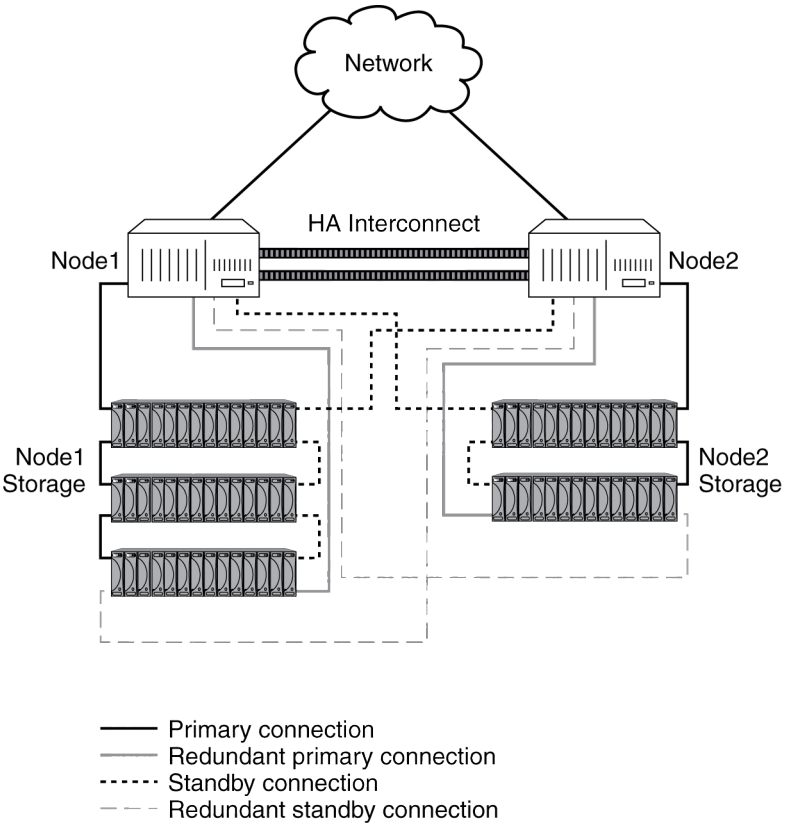
In some cases (such as the `lun config_check` command) these commands are issued automatically when the interconnect comes up. The resulting error can generate an AutoSupport

indicating a configuration problem when in fact the underlying problem is that Data ONTAP has not fully booted.

Standard active/active configuration diagram

A standard active/active configuration using native disk storage with multipath HA connects to the data network, has an HA interconnect between the controllers, and has primary, standby, and redundant connections to both node's disk shelves.

Standard active/active configuration



This diagram shows a standard active/active configuration with native disk shelves and multipath HA. Multipath HA is required on all active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections.

This diagram shows Fibre Channel disk shelves from the EXN1000 or EXN2000 unit/ EXN4000 unit family. For information on cabling SAS disk shelves, see the *Universal SAS*

and ACP Cabling Guide on the N series support website at *www.ibm.com/storage/support/nseries/*.

Setup requirements and restrictions for standard active/active configurations

You must follow certain requirements and restrictions when setting up a new standard active/active configuration.

The following list specifies the requirements and restrictions to be aware of when setting up a new standard active/active configuration:

- Architecture compatibility

  Both nodes must have the same system model and be running the same firmware version. See the *Data ONTAP Release Notes* for the list of supported systems.

  Note: In the case of systems with two controller modules in a single chassis (except the N6040, N6060, or N6070 systems), both nodes of the active/active configuration are located in the same chassis and have an internal interconnect.

- 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 appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html*. For a system using array LUNs, disks, or both, see the *Gateway Interoperability Matrix*.

  Note: After a failover, the takeover node temporarily serves data from all the storage in the active/active configuration. When the single-node capacity limit is less than the total active/active configuration capacity limit, the total disk space in an active/active configuration can be greater than the single-node capacity limit. It is acceptable for the takeover node to temporarily serve more than the single-node capacity would normally allow, as long as it does not own more than the single-node capacity.

- Disks and disk shelf compatibility

  - Fibre Channel, SATA, and SAS storage are supported in standard active/active configurations, as long as the storage types are not mixed on the same loop.
  - 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 active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections.

- Cluster 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 Common Internet File System (CIFS), Network File System (NFS), or SyncMirror, 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 active/active configuration loses functionality after a takeover.

Configuration variations for standard active/active configurations

Active/active configurations can be configured asymmetrically, as an active/passive pair, or with shared disk shelf stacks.

- Asymmetrical configurations

  In an asymmetrical standard active/active configuration, 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

  If your standard active/active configuration is using software-based disk ownership, 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.

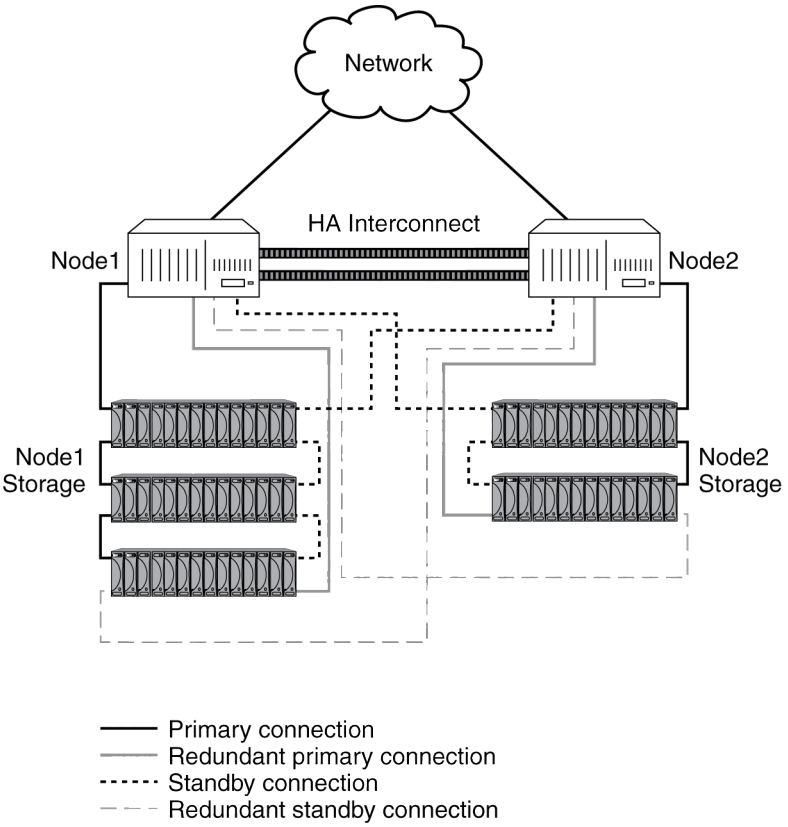
Multipath HA requirements and recommendations

Multipath HA is required on all active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections. Multipath HA was previously referred to as *Multipath Storage*.

What multipath HA for active/active configurations 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 active/active configuration without multipath HA has only one path from each controller to every disk, but an active/active configuration 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 active/active configuration using multipath HA. The redundant primary connections and the redundant standby connections are the additional connections required for multipath HA for active/active configurations.



How the connection types are used

Outlines the connection types used for multipath HA in active/active configurations.

The following table outlines the connection types used for multipath HA for active/active configurations, 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 active/active configuration 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.

Related concepts

Understanding redundant pathing in active/active configurations on page 158

Requirements for multipath HA

Multipath HA for active/active configurations has certain requirements.

Storage systems and disk shelves requiring multipath HA

Multipath HA is required on all active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections.

Disk shelf requirements

Multipath HA for active/active configurations is available on all supported disk shelves and disk modules.

Note: Multipath HA is not supported with third-party storage connected to gateway systems but is supported with native disks connected to gateway systems.

Software-based disk ownership requirements

Both nodes must be using software-based disk ownership.

To convert an active/active configuration to use software-based disk ownership, you must boot both nodes into Maintenance mode at the same time (during scheduled downtime).

Note: Plan to convert to software-based disk ownership before adding any cabling for multipath HA. After you add the cabling for multipath HA, you must manually assign all disks.

For more information about software-based disk ownership, see the chapter about disks in the *Data ONTAP Storage Management Guide*.

Fibre Channel port requirements for Fibre Channel disk shelves

Each node must have enough onboard Fibre Channel ports or HBAs to accommodate the cables required for multipath HA. You need two Fibre Channel ports for each loop.

If you are scheduling downtime to convert to software-based disk ownership, you should add the HBAs then. Otherwise, you can use the nondisruptive upgrade method to add the HBA; this method does not require downtime.

Note: See the appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* for information about which slots to use for the HBAs and in what order to use them.

SAS port requirements for SAS disk shelves

Each node must have enough onboard SAS ports and/or SAS HBAs to accommodate the cables required for multipath HA. You need two SAS ports on each controller for each stack.

Note: See appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* for information about which slots to use for the HBAs and in what order to use them.

Boot environment variable requirement for gateway systems

To use multipath HA on a gateway system, you must configure the `fc-nonarray-adapter-list` environment variable for each new loop before you connect and configure the disk shelf for multipath HA. See the *Gateway Implementation Guide for Native Disk Shelves*.

Understanding mirrored active/active configurations

Mirrored active/active configurations provide high availability through failover, just as standard active/active configurations do. Additionally, mirrored active/active configurations 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 active/active configurations 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 active/active configurations use SyncMirror. For more information about SyncMirror, see the *Data ONTAP Data Protection Online Backup and Recovery Guide*.

Advantages of mirrored active/active configurations

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 active/active configuration.

Similar to standard active/active configurations, if either node in a mirrored active/active configuration 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 active/active configurations

The restrictions and requirements for mirrored active/active configurations include those for a standard active/active configuration 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.
  - If hardware-based ownership is used on your systems, the disk shelves must be connected to the controllers so that the disks do not have to change pools when a takeover occurs.

    For example, on an N5200 system, if you connect an HBA in slot 2 to Channel A (the A Input port on the disk shelf), you should connect Channel B to an HBA that is also in pool 0 on the partner node. If you connect Channel B to an HBA in, for example, slot 4, the disks would have to change from pool 0 to pool 1 when a takeover occurs.

    For more information about how Data ONTAP assigns pools ownership, see the section about hardware-based disk ownership in the *Data ONTAP Storage Management Guide*.

- There must be sufficient spares in each pool to account for a disk or array LUN failure.

  Note: If your systems are using hardware-based disk ownership, pool membership is determined by the physical connections between the disk shelves and the controllers. If your systems are using software-based disk ownership, pool membership is determined explicitly using the Data ONTAP command-line interface. For more information, see the section on disk ownership in the *Data ONTAP Storage Management Guide*.

- On systems using software ownership, 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 more information about requirements for setting up SyncMirror with third-party storage.

- You must enable the following licenses on both nodes:
  - cluster
  - 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 active/active configurations on page 27

Configuration variations for mirrored active/active configurations

A number of configuration variations are supported for mirrored active/active configurations.

The following list describes some configuration variations that are supported for mirrored active/active configurations:

- Asymmetrical mirroring

  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 active/active configuration.

Understanding stretch MetroClusters

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

Like mirrored active/active configurations, stretch MetroClusters contain two complete copies of the specified data volumes or file systems that you indicated as being mirrored volumes or file systems in your active/active configuration. 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.

Unlike mirrored active/active configurations, MetroClusters provide the capability to force a failover when an entire node (including the controllers and storage) is destroyed or unavailable.

Note: In previous versions of this document, stretch MetroClusters were called nonswitched MetroClusters.

Note: If you are a gateway system customer, see the *Gateway MetroCluster Guide* for information about configuring and operating a gateway system in a MetroCluster configuration

Continued data service after loss of one node with MetroCluster

The 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.

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

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

Advantages of stretch MetroCluster configurations

MetroClusters provide the same advantages of mirroring as mirrored Active/active configurations, with the additional ability to initiate failover if an entire site becomes lost or unavailable.

For a MetroCluster, the advantages of a stretch MetroCluster are:

- 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.
-

In addition, a MetroCluster 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

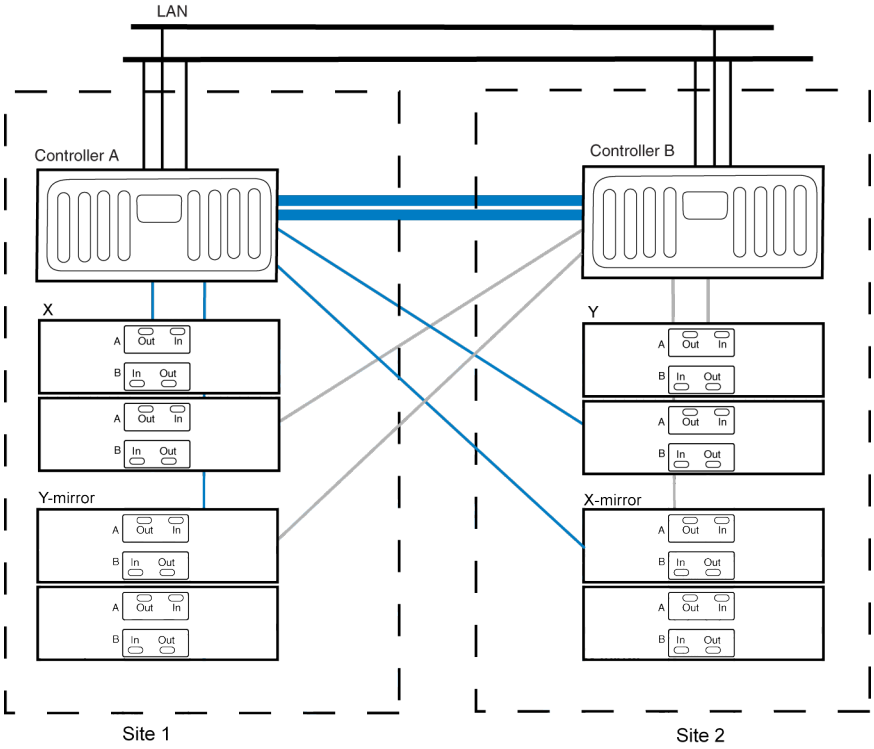
Disaster recovery using MetroCluster on page 163

Stretch MetroCluster configuration

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

- 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.

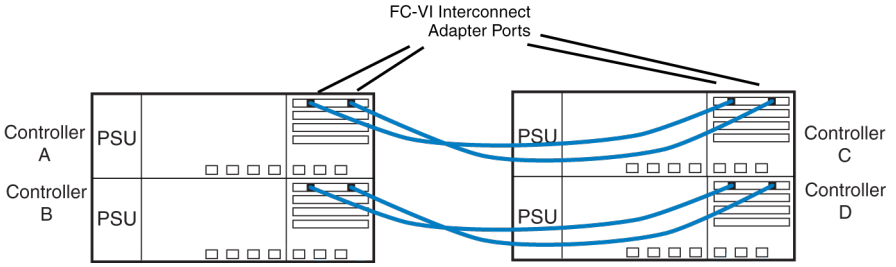
Stretch MetroCluster configuration on single- enclosure active/active configurations

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

To implement the stretch MetroCluster, you must install an FC-VI adapter in each controller to provide the cluster 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 MetroClusters, which use NVRAM adapters to provide the interconnect.

The following figure shows a stretch MetroCluster on single-enclosure active/active configuration systems.



Note: The dual-controller N3700, N3300, and N3600 systems do not support MetroCluster.

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.

Setup requirements and restrictions for stretch MetroCluster configurations

You must follow certain requirements and restrictions when setting up a new Stretch MetroCluster configuration.

The restrictions and requirements for stretch MetroClusters include those for a standard active/active configuration and those for a mirrored active/active configuration. In addition, the following requirements apply:

- SAS, SATA and Fibre Channel storage is supported on stretch MetroClusters, but both plexes of the same aggregate must use the same type of storage.
- Stretch MetroCluster configurations support SAS disk shelves with a distance limit of 5 meters between the nodes.
- MetroCluster is not supported on the N3300, N3400, or N3600 platforms.

- 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.
- If you want to convert the stretch MetroCluster configuration to a fabric-attached MetroCluster configuration, you must unset the speed of the nodes before conversion by using the `unsetenv` command.
- The following licenses must be enabled on both nodes:
  - cluster
  - syncmirror_local
  - cluster_remote

Note: See the MetroCluster Compatibility Matrix on the N series support site for more information about hardware and firmware requirements for this configuration.

Related concepts

Setup requirements and restrictions for standard active/active configurations on page 27
Setup requirements and restrictions for mirrored active/active configurations on page 32

Configuration variations for stretch MetroCluster configurations

Stretch MetroClusters have asymmetrical and active/passive variations.

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

- 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) 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 active/active configuration.

- Active/passive MetroClusters
  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 MetroCluster licenses installed so that remote takeover is possible.

Understanding fabric-attached MetroClusters

Like mirrored active/active configurations, fabric-attached MetroClusters contain two complete, separate copies of the data volumes or file systems that you configured as mirrored volumes or file systems in your active/active configuration. The fabric-attached MetroCluster nodes can be physically distant from each other, beyond the 500 meter limit of a stretch MetroCluster.

Note: If you are a gateway system customer, see the *Gateway MetroCluster Guide* for information about configuring and operating a gateway system in a MetroCluster configuration.

Fabric-attached MetroClusters use Brocade Fibre Channel switches

A MetroCluster configuration for distances greater than 500 meters connects the two nodes using four Brocade 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 active/active configurations.

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 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

Brocade Switch Configuration Guide for Fabric MetroClusters - www.ibm.com/storage/support/nseries

Advantages of fabric-attached MetroCluster configurations

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

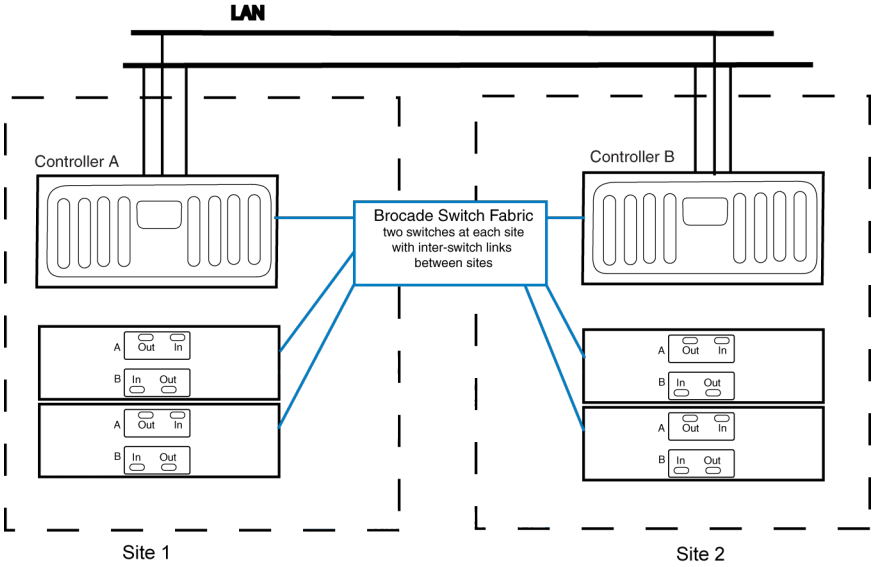
The advantages of a fabric-attached MetroCluster over a stretch MetroCluster include the following:

- The two halves of the configuration can be more than 500 meters apart, which provides increased disaster protection.
- 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.

Fabric-attached MetroCluster configuration

A fabric-attached MetroCluster includes two Brocade Fibre Channel switch fabrics that provide long distance connectivity between the nodes. Through the Brocade 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.

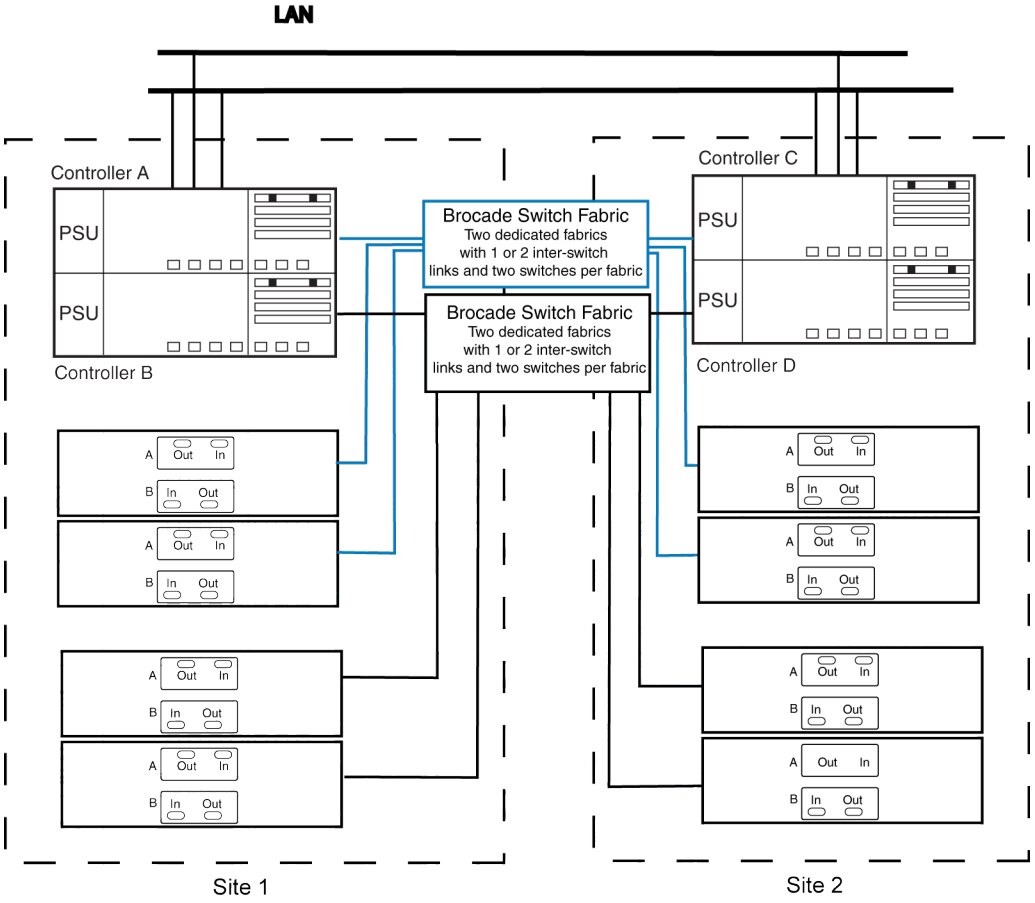


Note: This is a simplified figure that does not show disk shelf-to-disk shelf connections.

Fabric-attached MetroCluster configuration on single enclosure active/active configuration systems

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

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



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 as on a stretch MetroCluster.

Related concepts

How Data ONTAP works with stretch MetroCluster configurations on page 36

Setup requirements and restrictions for fabric-attached MetroClusters

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

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

Note: See the MetroCluster Compatibility Matrix on the N series support website 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
  - N6040, N6060, or N6070 systems
  - The N6040, N6060, or N6070 systems can have two controllers in the same chassis.
    When in a MetroCluster configuration, only a single controller in each system is supported (rather than two).
  - N7600, N7700, N7800, or N7900
  - N6200 series systems
- Each node requires a FC-VI (Fibre Channel-Virtual Interface) adapter; the slot position is dependent on the controller model.
  The 4-Gbps FC-VI adapter is supported on the following systems using software-based disk ownership:
  - N7600, N7700, N7800, or N7900
  - N5300
  - N5600
  - N6040, N6060, or N6070

    Note: For information about supported cards and slot placement, see the appropriate hardware and service guide on the N series support site.

    The FC-VI adapter is also called a VI-MC or VI-MetroCluster adapter.
- The 8-Gbps FC-VI (Fibre Channel-Virtual Interface) adapter is supported only on the N6200 series systems.

Disk and disk shelf requirements

- Only EXN2000 and EXN4000 units are supported.
- Only Fibre Channel disks are supported; you cannot use SATA drives or AT-FCX modules for fabric-attached MetroCluster configurations.
- Only homogeneous stacks of SAS disk shelves are supported. For example, a stack of disk shelves can only contain either EXN3000 or EXN3500 disk shelves.
- You can connect a maximum of two EXN2000 unit disk shelves to each loop.

Capacity limits

The maximum capacity for a system configured in a fabric-attached MetroCluster 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 site.
- 840 Fibre Channel disks (60 disk shelves).

Fibre Channel switch requirements

Note: For the most up-to-date switch information, including supported switches and firmware downloads, see the Fabric-Attached MetroCluster Switch Description page on the N series support website.

- Each site of the MetroCluster requires two switches.
- Switches must be a supported Brocade model supplied by IBM. Customer supplied switches are not supported.
- The two switches at a site must be the same model and must be licensed for the same number of ports.
- All the four switches for a particular Fabric-attached MetroCluster must support the same maximum speed.
- Switches must be running the correct firmware version.

License requirements

- cluster
- syncmirror_local
- cluster_remote

Related concepts

Setup requirements and restrictions for standard active/active configurations on page 27
Setup requirements and restrictions for mirrored active/active configurations on page 32

Configuration limitations for fabric-attached MetroClusters

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:

- SATA and AT-FCX storage is not supported.
- You cannot use the MetroCluster switches to connect Fibre Channel tape devices, or for Fibre Channel Protocol (FCP) traffic of any kind.

  You can connect only system controllers and disk shelves to the MetroCluster switches.
- You can connect a tape storage area network (SAN) to either of the nodes, but the tape SAN must not use the MetroCluster switches.

Configuration variations for fabric-attached MetroClusters

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

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

- 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 active/active configuration.

- Active/passive MetroClusters

  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 MetroCluster licenses installed so that remote takeover is possible.

Active/active configuration installation

To install and cable a new standard or mirrored active/active configuration 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). You must also cable the cluster interconnect between the nodes. Active/active configurations 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 active/active configuration.
- Whether you are using Fibre Channel 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, which is accessed and navigated as described in *Websites* on page 13 for information on disk shelf cabling. For cabling the HA interconnect between the nodes, use the procedures in this guide.

Multipath HA is required on all active/active configurations except for some N3300, N3400, or N3600 system configurations which use single-path HA. Single-path HA configurations lack the redundant standby connections.

Systems with two controller modules in the same chassis

In an active/active configuration, some storage systems (such as the N3300, N3400, or N3600 systems) support two controller modules in the same chassis.

This simplifies system cabling, because an internal InfiniBand connector between the two controller modules replaces the interconnect adapters and cabling used in other configurations. The following illustration shows a system with two controller modules in the chassis forming a single chassis active/active configuration.

This configuration is different from other examples in this section, which show systems that must be cabled together with an interconnect to enable the active/active configuration.

Related concepts

Systems with variable HA configurations on page 20

System cabinet or equipment rack installation

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

Active/active configurations 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.

Active/active configurations in a system cabinet

If you ordered an active/active configuration in a system cabinet, it comes in one or more IBM system cabinets, depending on the amount of storage.

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 the active/active configuration you ordered has six or fewer disk shelves, it arrives in a single system cabinet. This system cabinet has both the Channel A and Channel B disk shelves cabled, and also has the cluster adapters cabled.

If the active/active configuration you ordered has more than six disk shelves, the active/active configuration arrives in two or more system cabinets. 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 cluster interconnects. If the active/active configuration 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, tools, and equipment

Installation of an active/active configuration requires the correct documentation, tools, and equipment.

Required documentation

Installation of an active/active configuration requires the correct documentation.

IBM hardware and service documentation comes with your hardware and is also available at the N series support site.

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

Manual name	**Description**
The appropriate system cabinet guide	This guide describes how to install IBM N series equipment into a system cabinet.
Site Requirements Guide	This guide describes the physical requirements your site must meet to install IBM N series equipment.
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 Guide	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	This guide describes, among other topics, SyncMirror technology, which is used for mirrored active/active configurations.
Data ONTAP System Administration Guide	This guide describes general storage system administration.
Data ONTAP Software Setup Guide	This guide describes how to configure the software of a new storage system for the first time.

Note: If you are installing a gateway active/active configuration, refer also to the *Gateway Installation Requirements and Reference Guide* for information about cabling gateway systems to storage arrays and to the gateway Implementation Guides for information about configuring storage arrays to work with gateway systems.

Required tools

Installation of an active/active configuration requires the correct tools.

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

- #1 and #2 Phillips screwdrivers

- Hand level
- Marker

Required equipment

When you receive your active/active configuration, you should receive the equipment listed in the following table. See the appropriate hardware and service guide at the N series support site to confirm your storage system type, storage capacity, and so on.

Required equipment	Standard or mirrored active/active configuration
Storage system	Two of the same type of storage systems.
Storage	See the appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html*.
Cluster interconnect adapter (for controller modules that do not share a chassis) **Note:** When N6200 series systems are in a dual-chassis active/active configuration, they use the c0a and c0b 10-GbE ports for the HA interconnect. They do not require an HA interconnect adapter.	InfiniBand (IB) cluster adapter (The NVRAM adapter functions as the cluster interconnect adapter on N5000 series and later storage systems, except the N6200 series systems.)
For EXN1000 or EXN2000 unit/EXN4000 unit family 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
SFP (Small Form Pluggable) modules	N/A
NVRAM cluster adapter media converter	Only if using fiber cabling.

Required equipment	Standard or mirrored active/active configuration
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 cluster adapter • The N6200 series systems, when in a dual-chassis active/active configuration, 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 active/active configuration.

Installing the nodes in a system cabinet

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

Steps

1. Install the system cabinets, 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.
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 active/active configuration.

Cabling a standard active/active configuration

To cable a standard active/active configuration, you identify the ports you need to use on each node, then you cable the ports, and then you cable the cluster interconnect.

About this task

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

For cabling SAS disk shelves in an active/active configuration, see the *Universal SAS and ACP Cabling Guide*.

The sections for cabling the cluster 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 51
2. *Cabling Node A to EXN1000 or EXN2000 unit or EXN4000 unit disk shelves* on page 52
3. *Cabling Node B to EXN1000 or EXN2000 unit or EXN4000 unit disk shelves* on page 54
4. *Cabling the cluster interconnect (all systems except N6200 series)* on page 56

5. *Cabling the cluster interconnect (N6200 series systems in separate chassis)* on page 57

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

Before cabling your active/active configuration, 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 active/active configuration requires two ports on the node, one for the primary connection and one for the redundant multipath HA connection.

  A standard active/active configuration 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 appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* for your platform for an active/active configuration.
- If using Fibre Channel HBAs, insert the adapters in the same slots on both systems.

See the appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* to obtain all slot assignment information for the various adapters you use to cable your active/active configuration.

When complete, 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 or ESH2 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. Use port A for the first loop, then B, 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 appropriate hardware and service guide must be observed.

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

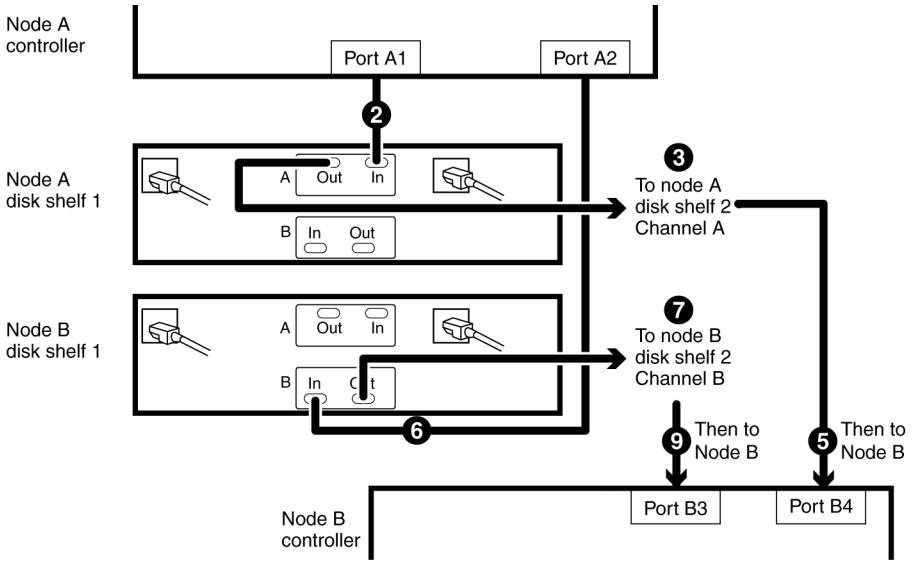
To cable Node A, you must use the Fibre Channel ports you previously identified and cable the disk shelf loops owned by the node to these ports.

About this task

- This procedure uses multipath HA, required on all systems.
- This procedure does not apply to SAS disk shelves.
  For cabling SAS disk shelves in an active/active configuration, see the *Universal SAS and ACP Cabling Guide*.
- For additional cabling diagrams, you can refer to your system's *Installation and Setup Instructions* on the N series support website.

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 active/active configuration, 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 or EXN2000 unit or EXN4000 unit disk shelves

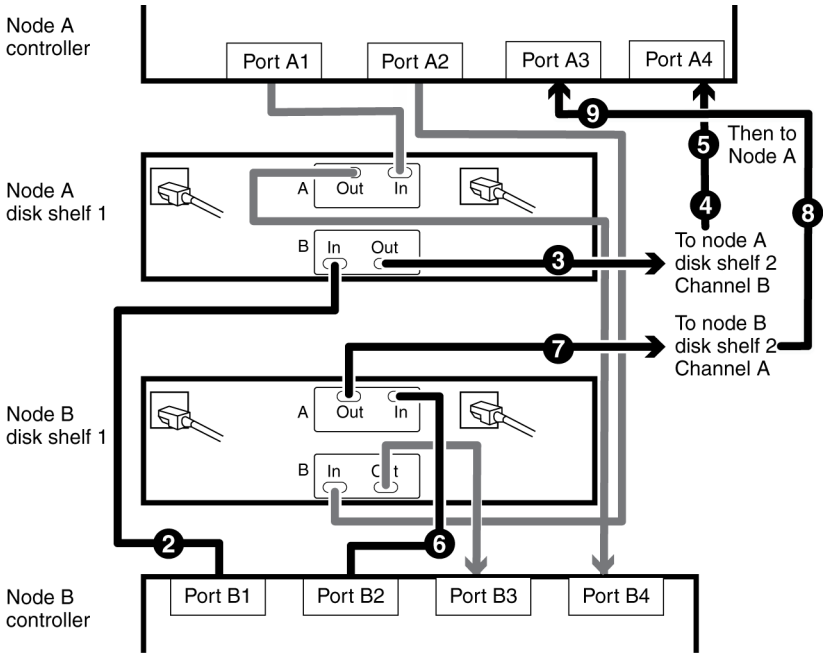
To cable Node B, you must use the Fibre Channel ports you previously identified and cable the disk shelf loops owned by the node to these ports.

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 active/active configuration, see the *Universal SAS and ACP Cabling Guide*.
- For additional cabling diagrams, you can refer to your system's *Installation and Setup Instructions* on the N series support website.

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 active/active configuration 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 active/active configuration, 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 cluster interconnect.

Cabling the cluster interconnect (all systems except N6200 series)

To cable the cluster interconnect between the active/active configuration 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 active/active configurations (active/active configurations in which the two controller modules reside in separate chassis) except N6200 series systems, regardless of disk shelf type.

Steps

1. See the appropriate hardware and service guide on the N series support site to ensure that your interconnect adapter is in the correct slot for your system in an active/active configuration.

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

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

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

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

```
Cluster 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 cluster adapters.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

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

To enable the cluster 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 active/active configuration

To cable a mirrored active/active configuration, you identify the ports you need to use on each node, and then you cable the ports, and then you cable the cluster interconnect.

About this task

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

For cabling SAS disk shelves in an active/active configuration, see the *Universal SAS and ACP Cabling Guide*.

The sections for cabling the cluster 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 51
2. *Creating your port list for mirrored active/active configurations* on page 59
3. *Cabling the Channel A EXN1000 or EXN2000 unit or EXN4000 unit disk shelf loops* on page 60
4. *Cabling the Channel B EXN1000 or EXN2000 unit or EXN4000 unit disk shelf loops* on page 62
5. *Cabling the redundant multipath HA connection for each loop* on page 64
6. *Cabling the cluster interconnect (all systems except N6200 series)* on page 66
7. *Cabling the cluster interconnect (N6200 series systems in separate chassis)* on page 67

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

Before cabling your active/active configuration, 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 active/active configuration requires two ports on the node, one for the primary connection and one for the redundant multipath HA connection.

  A standard active/active configuration 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 appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* for your platform for an active/active configuration.
- If using Fibre Channel HBAs, insert the adapters in the same slots on both systems.

See the appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html* to obtain all slot assignment information for the various adapters you use to cable your active/active configuration.

When complete, 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 or ESH2 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. Use port A for the first loop, then B, 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 appropriate hardware and service guide must be observed.

Creating your port list for mirrored active/active configurations

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 active/active configurations, 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. To ensure proper disk pool access, your cabling depends on whether you have hardware-based or software-based disk ownership.

If your system uses hardware-based disk ownership, you must cable your mirrored active/active configuration according to the pool rules for your platform. For more information about pool rules, see the section on hardware-based disk ownership in the *Data ONTAP Storage Management Guide*.

If your system uses software-based disk ownership, follow the guidelines for software-based disk ownership in the *Data ONTAP Storage Management Guide*.

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

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).

Example

For an N5000 series active/active configuration that has two mirrored loops using hardware-based disk ownership, the port list would 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

Pool 0	Pool 1
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 or EXN4000 unit disk shelf loops

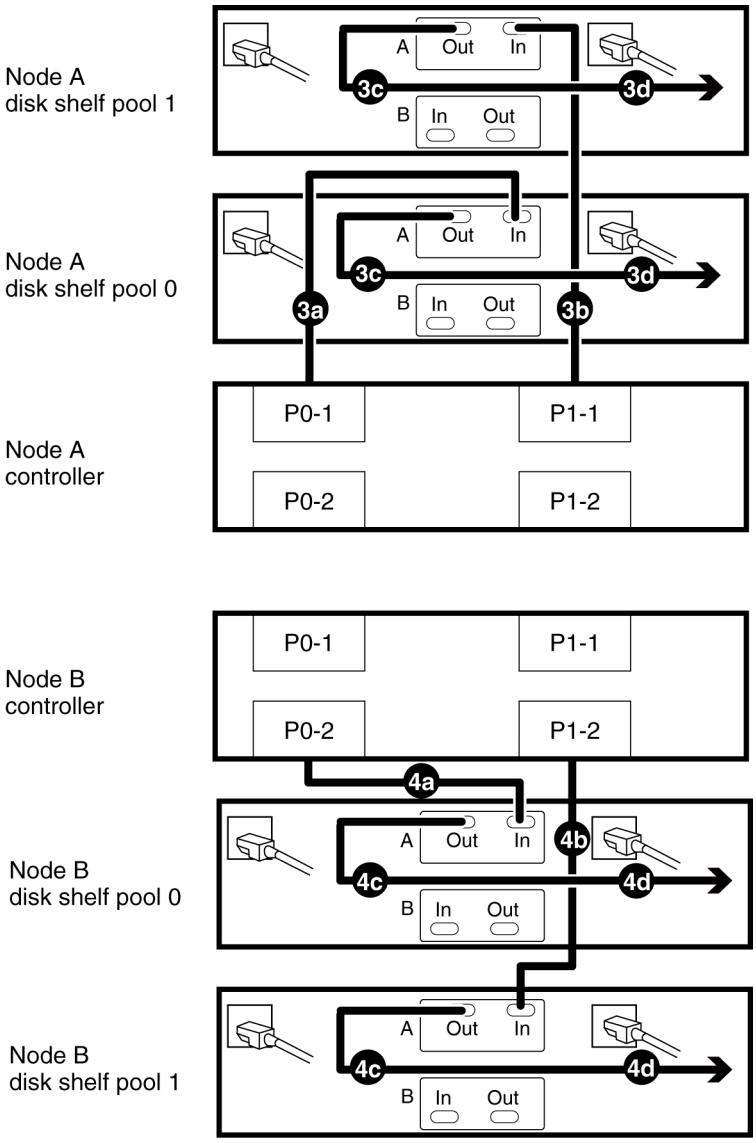
To begin cabling of the disk shelves, you cable the appropriate pool ports on the node to the Channel A modules of the disk shelf stack for the pool.

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 active/active configuration, 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 2c, connecting Channel A output to input, for any remaining disk shelves in this loop for each disk pool.

e. Repeat Substep a through Substep e 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 3.c, connecting Channel A output to input, for any remaining disk shelves in each disk pool.

e. Repeat substep 3.a through substep 3.e 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 or EXN2000 unit 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.

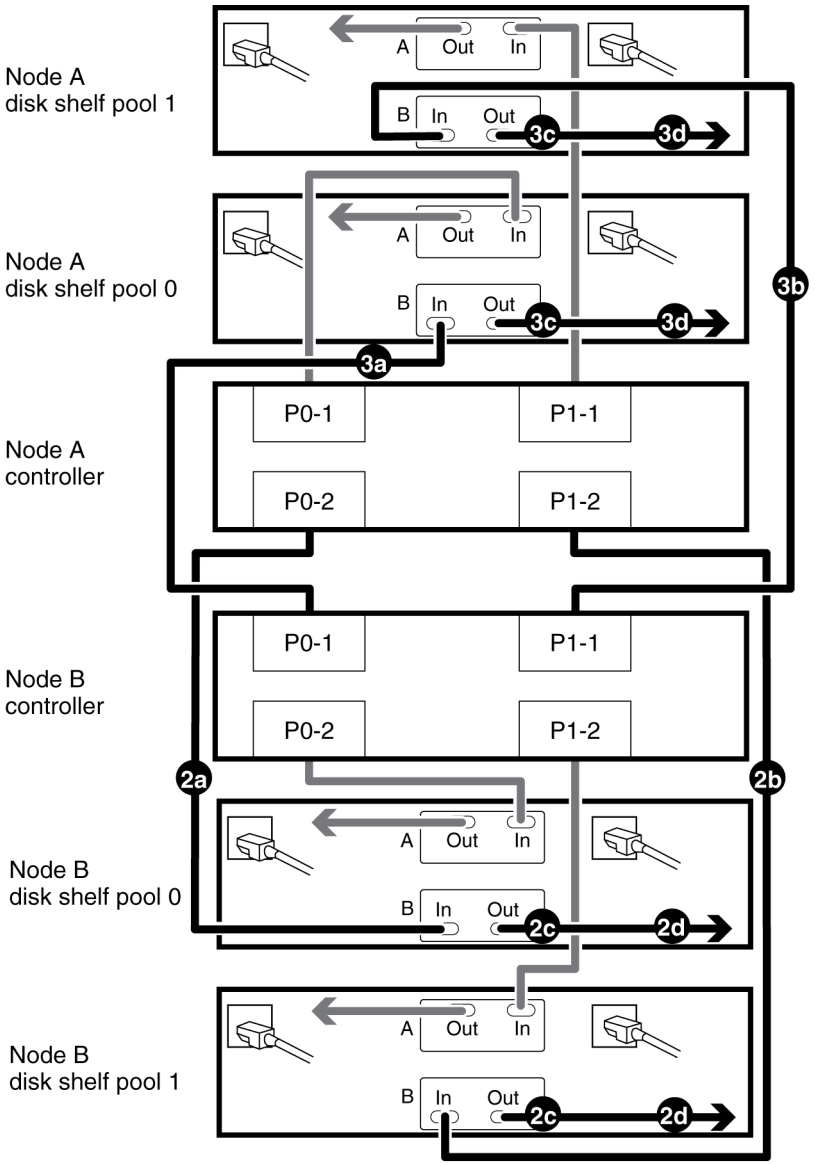
Steps

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 active/active configuration 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 c, connecting Channel B output to input, for any remaining disk shelves in each disk pool.

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

e. Repeat Substep a through Substep d 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 cluster 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

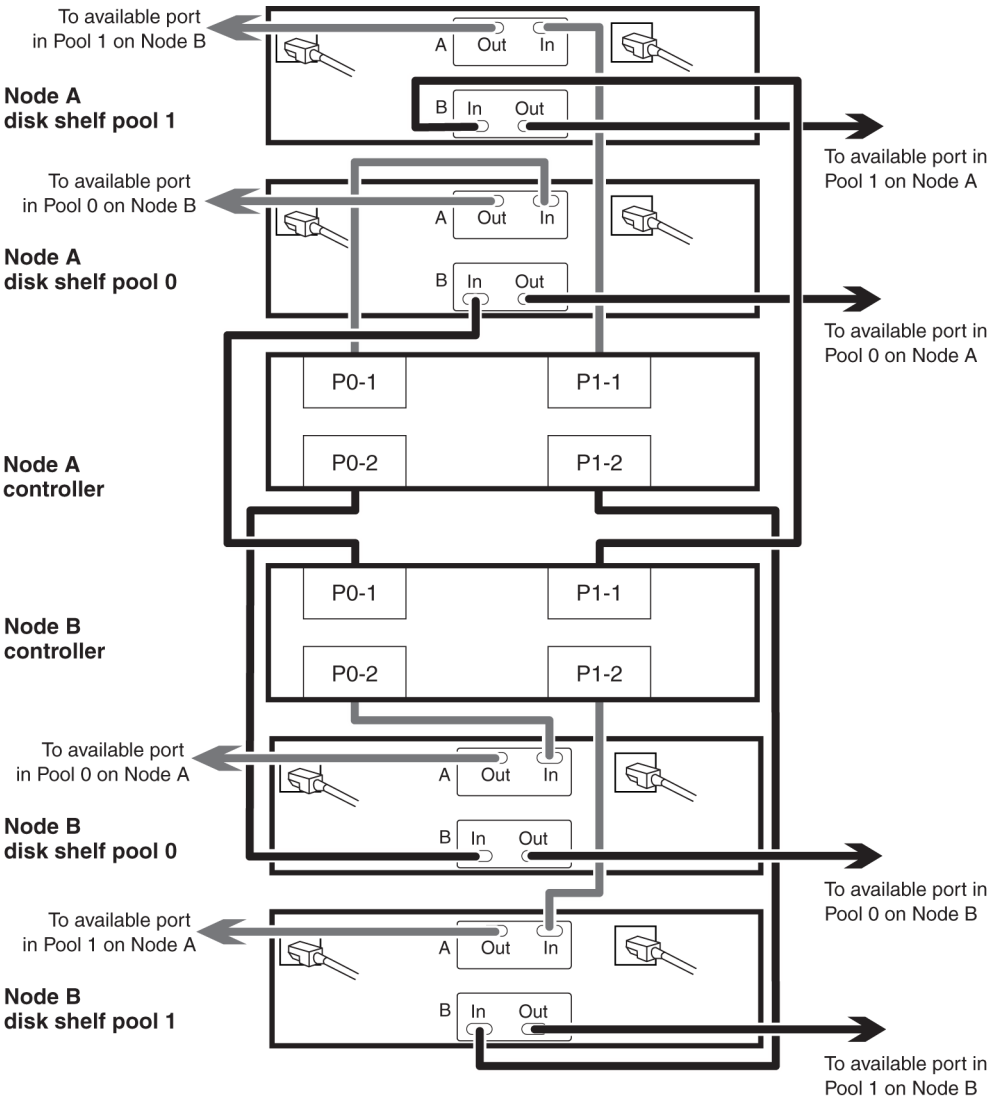
1. Review the cabling diagram before proceeding to the cabling steps.

- 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 cluster interconnect (all systems except N6200 series)

To cable the cluster interconnect between the active/active configuration 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 active/active configurations (active/active configurations in which the two controller modules reside in separate chassis) except N6200 series systems, regardless of disk shelf type.

Steps

1. See the appropriate hardware and service guide on the N series support site to ensure that your interconnect adapter is in the correct slot for your system in an active/active configuration.

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

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

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

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

   ```
   Cluster 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 cluster adapters.

Result

The nodes are connected to each other.

After you finish

Proceed to configure the system.

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

To enable the cluster 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 active/active configurations

You can use a UPS (uninterruptible power supply) with your active/active configuration. 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 active/active configuration.

For a standard active/active configuration, you must connect the controller, disks, and any Fibre Channel switches in use.

For a mirrored active/active configuration, you must connect the controller and any Fibre Channel switches to the UPS, as for a standard active/active configuration. 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 active/active configuration cannot shut down gracefully.

MetroCluster installation

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

Note: If you are a gateway system customer, see the *Gateway MetroCluster Guide* for information about configuring and operating a gateway system in a MetroCluster configuration.

Related concepts

Disaster recovery using MetroCluster on page 163
Setup requirements and restrictions for stretch MetroCluster configurations on page 36
Setup requirements and restrictions for fabric-attached MetroClusters on page 40

Required documentation, tools, and equipment

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

Required documentation

Describes the flyers and guides required to install a new MetroCluster, or convert two stand-alone systems into a MetroCluster.

IBM hardware and service documentation comes with your hardware and is also available at the N series support site.

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	This guide describes how to install IBM N series equipment into a system cabinet.
Site Requirements Guide	This guide describes the physical requirements your site must meet to install IBM N series equipment.
The appropriate disk shelf guide	These guides describe how to cable a disk shelf to a storage system.

Manual name	Description
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.
Upgrade Guide	This guide describes how to upgrade storage system and disk firmware, and how to upgrade storage system software.
Data Protection Online Backup and Recovery Guide	This guide describes, among other topics, SyncMirror technology, which is used for mirrored Active/active configurations.
Data ONTAP System Administration Guide	This guide describes general storage system administration.
Software Setup Guide	This guide describes how to configure the software of a new storage system for the first time.
Brocade Switch Configuration Guide for Fabric-attached MetroClusters	This document describes how to configure Brocade switches for a fabric-attached MetroCluster. You can find this document on the N series support site.
The appropriate Brocade manuals	These guides describe how to configure and maintain Brocade switches. These guides are available from the N series support site.

Required tools

Lists the tools you need to install the active/active configuration.

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

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

Required equipment

When you receive your MetroCluster, you should receive the equipment listed in the following table. See the appropriate hardware and service guide at the N series support site to confirm your storage system type, storage capacity, and so on.

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

Required equipment	**Stretch MetroCluster**	**Fabric-attached MetroCluster**
Storage system	Two of the same type of storage systems.	
Storage	See the appropriate hardware and service guide and the N series Interoperability Matrix at *www.ibm.com/systems/storage/network/interophome.html*.	
cluster interconnect adapter	InfiniBand adapter (Required only for systems that do not use an NVRAM5 or NVRAM6 adapter, which functions as the cluster interconnect adapter.) FC-VI adapter (Required only for the N6040, N6060, or N6070 dual-controller systems.) **Note:** When the FC-VI adapter is installed in n N6040, N6060, or N6070 system, the internal InfiniBand interconnect is automatically deactivated.	FC-VI adapter
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.	

Required equipment	Stretch MetroCluster	Fabric-attached MetroCluster
Fibre Channel switches	N/A	Two pairs of Brocade 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
Cables (provided with shipment unless otherwise noted)	• Four SC/LC (standard connector to low-profile connector) controller-to-disk shelf cables • Two SC/LC IB cluster 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 site.	• LC/LC controller-to-switch cables • LC/LC (for EXN2000) disk shelf-to-switch cables • Two LC/LC inter-switch link cables, not provided in the shipment • Multiple disk shelf-to-disk shelf cables

MetroCluster and software-based disk ownership

Systems using software-based disk ownership in a MetroCluster require different configuration than systems using hardware-based disk ownership.

Some systems use software-based disk ownership to control which disks in a disk shelf loop belong to which controller and pool.

- Software commands in Data ONTAP are used to assign disks, or they are auto-assigned by the software.
  This is because disk ownership is determined by the software, rather than by the physical cabling of the shelves.
- Systems that use software disk ownership require different cabling of their disk shelves when you configure your MetroCluster.
  This is because different Brocade port usage rules are used with software-based disk ownership.

For details about software-based disk ownership, see the *Data ONTAP Storage Management Guide*.

The 4-Gbps FC-VI adapter requires software disk ownership

If you want to take advantage of the performance provided by the 4-Gbps adapter, you must upgrade to a system that uses software-based disk ownership.

Converting an active/active configuration to a fabric-attached MetroCluster

With the correct hardware, you can reconfigure an active/active configuration to a fabric-attached MetroCluster.

Before you begin

- If you are upgrading an existing active/active configuration 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. You can download the latest disk firmware from *www.ibm.com/storage/support/nseries/*.
- If you are upgrading from an existing active/active configuration on a system that supports both software-based and hardware-based disk ownership and is currently using software-based disk ownership, you must convert disk assignment to hardware ownership before the upgrade. However, converting an active/active configuration from software-based disk ownership to hardware-based disk ownership is a complicated process. If done incorrectly, your system might not boot. You are advised to contact technical support for assistance with this conversion.
- If you are upgrading an N6040, N6060, or N6070 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. You must also obtain and install the FC-VI interconnect card on both systems.

Note: For details about this conversion process, see the *MetroCluster Upgrade Planning Guide*, on *www.ibm.com/storage/support/nseries/*.

Steps

1. Update Data ONTAP, storage system firmware, and disk firmware, as described in the *Data ONTAP Upgrade Guide*, 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 site.
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 *Brocade Switch Configuration Guide for Fabric-attached MetroClusters*.

   You can find this document on the N series support site.

   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 when installing and configuring your MetroCluster switches:

   - The MetroCluster Compatibility Matrix
   - The Brocade Switch Description Page
   - The *Brocade Switch Configuration Guide for Fabric-attached MetroClusters*
6. Cable the local node.
7. Install the Data ONTAP licenses in the following order:
   a. cluster
   b. syncmirror_local
   c. cluster _remote
8. Configure the local node depending on the type of active/active configuration:

If you are converting a...	Then...
Standard active/active configuration	Set up mirroring and configure the local node.
Stretch MetroCluster	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.

Related concepts

Configuring an active/active configuration on page 107
Disaster recovery using MetroCluster on page 163

Related tasks

Cabling Node A on page 86
Cabling Node B on page 91
Disabling the change_fsid option in MetroCluster configurations on page 112
Planning the fabric-attached MetroCluster installation on page 83

Upgrading an existing MetroCluster

You can upgrade an existing MetroCluster on a system using hardware-based disk ownership to a MetroCluster on a system using software-based disk ownership (N5300, N5600, or N7600, N7700, N7800, or N7900 systems). This is useful when you are upgrading to 4-Gbps cluster interconnect support, which requires software-based disk ownership.

About this task

When using the typical hardware upgrade procedure you upgrade your software on the old system and then use the `disk upgrade_ownership` command to apply software-based ownership to the disks. You then perform the hardware upgrade.

In the following procedure, you perform the hardware upgrade prior to using the `disk upgrade_ownership` command. This is because the old system hardware does not support the new features of the `disk upgrade_ownership` command. For the Data ONTAP 7.2.3 (or later) version of the `disk upgrade_ownership` command to run successfully, you must issue it on a system that supports software-based disk ownership.

Steps

1. Halt the system, and then turn off the controller and disk shelves.
2. Remove the existing controller from the rack or system cabinet and install the N5300, N5600, or N7600, N7700, N7800, or N7900 system in its place.

When replacing the controllers, use the same cabling to the Brocade switches and the disk shelves as the original controller. For the upgrade to work, you must retain the original cabling until you run the `disk upgrade_ownership` command later in this procedure.

3. Power on the disk shelves.
4. To reassign disk ownership to software-based disk ownership, complete the following substeps on both controllers:
   a. Power on the system and boot the system into Maintenance mode.

      For more information, see the *Data ONTAP System Administration Guide*.

   b. Enter the following command at the firmware prompt:

      `disk upgrade_ownership`

      This command converts the system to software-based disk ownership. Data ONTAP assigns all the disks to the same system and pool that they were assigned to for the hardware-based disk ownership.

      See the *Data ONTAP Storage Management Guide* for detailed information about software-based disk ownership.
5. Verify disk ownership information by entering the following command:

   `disk show -v`

   Disk assignment is now complete.
6. Clear the mailboxes by entering the following commands:

   `mailbox destroy local`

   `mailbox destroy partner`
7. Enter the following command to exit Maintenance mode:

   `halt`
8. Enter the following command for each required license:

   Example

   `license add xxxxxx`

   xxxxx is the license code you received for the feature.
9. Enter the following command to reboot the node:

   `reboot`
10. Configure the RLM, if applicable, as described in the *Data ONTAP Software Setup Guide*.
11. Recable the connections to the Brocade switches to conform to the virtual channel rules for the switch.

After you finish

Configure the MetroCluster.

Related concepts

Configuring an active/active configuration on page 107
Disaster recovery using MetroCluster on page 163
Switch bank rules and virtual channel rules on page 85

Related tasks

Cabling Node A on page 86
Cabling Node B on page 91
Disabling the change_fsid option in MetroCluster configurations on page 112

Cabling a stretch MetroCluster

The process to cable a stretch MetroCluster is the same as a mirrored active/active configuration. However, your systems must meet the requirements for a stretch MetroCluster.

Related concepts

Configuring an active/active configuration on page 107
Setup requirements and restrictions for stretch MetroCluster configurations on page 36
Disaster recovery using MetroCluster on page 163

Related tasks

Cabling a mirrored active/active configuration on page 57

Cabling a stretch MetroCluster between single enclosure active/active configuration systems

If you are configuring a stretch MetroCluster between single enclosure active/active configuration systems (for example, N6040, N6060, or N6070 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 MetroClusters. 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 active/active configuration 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 MetroClusters configuration.

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 formed by the top controllers as described in the procedure for cabling a mirrored active/active configuration.
5. Cable the disk shelf loops for the stretch MetroCluster formed by the bottom controllers as described in the procedure for cabling a mirrored active/active configuration.

Related concepts

Stretch MetroCluster configuration on single- enclosure active/active configurations on page 35

Related tasks

Cabling a mirrored active/active configuration on page 57

Changing the default configuration speed of a stretch MetroCluster

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

Before you begin

The stretch MetroCluster default configuration speed must conform to the Stretch MetroCluster setup requirements and restrictions.

About this task

- You enter the commands at the boot environment prompt, which might 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 system by entering the following command:

   `halt`

2. Reset the configuration speed.

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 system console displays output similar to the following:
	`Variable Name Value`
	`----------`
	`----------`
	`ispfcvi-force-2G-only? *** Undefined ***`
	d. Verify that your system is configured 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? 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 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 ***`
	d. Verify that your system is configured for 2 Gb by entering the following command:
	printenv *ispfcvi-force-2G-only?*
	If your 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 using an FC-VI adapter, you can reset it to the default speed by using the `unsetenv` command at the boot environment prompt.

About this task

- The boot environment prompt might 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, 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?***
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

Cabling a fabric-attached MetroCluster

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

Before you begin

To cable a fabric-attached MetroCluster, you must be familiar with active/active configurations, the Brocade command-line interface, and synchronous mirroring. You must also be familiar with the characteristics of fabric-attached MetroClusters. You must also have the following information:

- Correct Brocade 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.

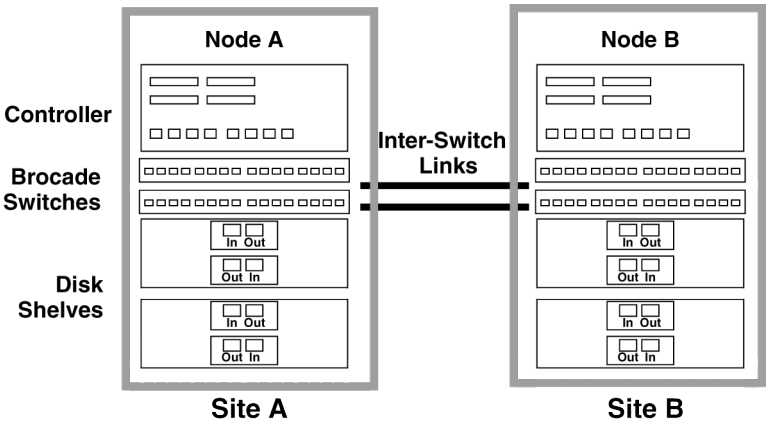
- Ethernet IP address for both the switches and nodes

  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 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.



Complete the following tasks in the order shown:

Steps

1. *Planning the fabric-attached MetroCluster installation* on page 83
2. *Configuration differences for fabric-attached MetroClusters on single enclosure active/active configuration* on page 84
3. *Configuring the switches* on page 84
4. *Cabling Node A* on page 86
5. *Cabling Node B* on page 91
6. *Assigning disk pools (if you have software-based disk ownership)* on page 96
7. *Verifying disk paths if you have software-based disk ownership* on page 97

Related concepts

Setup requirements and restrictions for fabric-attached MetroClusters on page 40
Configuring an active/active configuration on page 107
Disaster recovery using MetroCluster on page 163

Related tasks

Disabling the change_fsid option in MetroCluster configurations on page 112

Planning the fabric-attached MetroCluster installation

You must fill out the fabric-attached MetroCluster worksheet to record specific cabling information about your fabric-attached MetroCluster. 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 a 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 at each site connect to the Fibre Channel switches. Use the following table to record which switch port the disk shelves are connected to.

Disk shelf...	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 MetroClusters on single enclosure active/active configuration

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

A single enclosure active/active configuration can be connected to another active/active configuration to create two separate fabric-attached MetroClusters. The internal InfiniBand connection in each system are automatically deactivated when the FCVI card is installed in the controller.

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

Related concepts

Fabric-attached MetroCluster configuration on single enclosure active/active configuration systems on page 39

Configuring the switches

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

Step

1. To configure your Brocade switches, see the *Brocade Switch Configuration Guide for Fabric-attached MetroClusters* for your switch model. You can find this document on the MetroCluster Switch Description Page at the N series support website, which is accessed and navigated as described in *Websites* on page 13.

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 or the MetroCluster Switch Description Page, when installing and configuring your MetroCluster switches.

After you finish

Proceed to configure Node A.

Related concepts

Switch bank rules and virtual channel rules on page 85

Switch bank rules and virtual channel rules

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

If your system uses hardware-based disk ownership, you must use the switch bank rules when cabling the Brocade switch. This ensures that switch traffic is distributed across the switch quadrants to reduce potential bottlenecks.

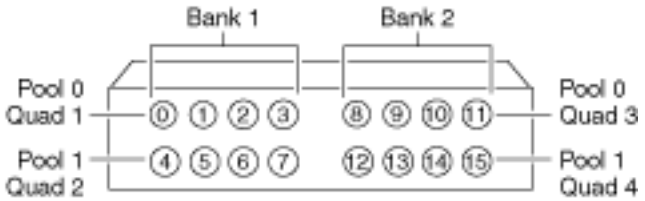
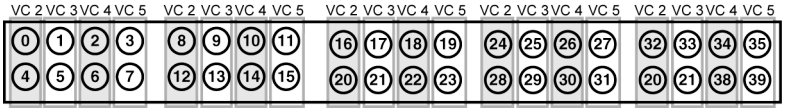


Figure 1: Brocade switch showing which ports belong to which switch banks and pools

If your system does not use hardware-based disk ownership, 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

Brocade Switch Configuration Guide for Fabric MetroCluster - www.ibm.com/storage/support/nseries

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 cluster interconnect to the switches, and ensure that the disk shelves in the configuration belong to the correct pools.

About this task

Complete the following tasks in the order shown:

Steps

1. *Cabling the controller when you have software-based disk ownership* on page 86
2. *Cabling the shelves when you have software-based disk ownership* on page 88
3. *Cabling the FC-VI adapter and inter-switch link when you have software-based disk ownership* on page 89

Cabling the controller when you have software-based disk ownership

You can use this procedure to cable the Fibre Channel ports on the controller to the Brocade switches when your system uses software-based disk ownership.

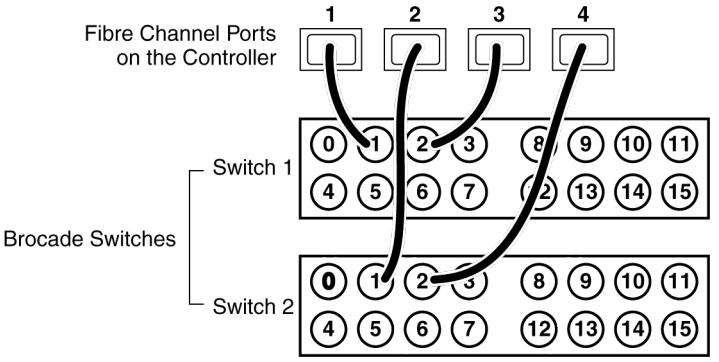
Before you begin

- Select one virtual channel on the switch for the cluster interconnect connections. The following examples use virtual channel 2, which includes ports 0, 4, 8, and 12.
- If you are using a dual-port HBA, connecting both ports of the HBA to the same switch port number can make it easier to cable and administer your MetroCluster. (However, this is not required.)

  For example, if port 1 of the HBA is connected to port 1 of Switch 1, you should connect port 2 of that HBA to port 1 of Switch 2.
- Both Fibre Channel ports on the same dual-port HBA (or adjacent pairs of onboard ports) should never be connected to the same switch. You must connect one port to one switch and the other port to the other switch.

  For example, if onboard port 0a is connected to Switch 3, you should not connect onboard port 0b to Switch 3; you must connect port 0b to Switch 4.

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.

   Again, 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 51

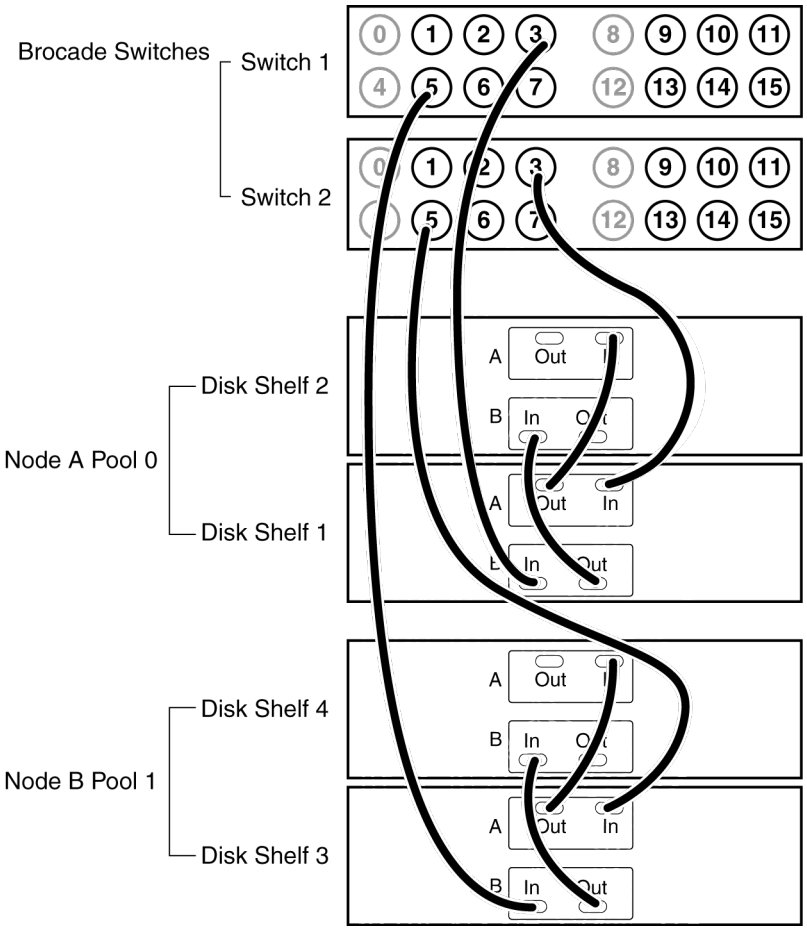
Cabling the shelves when you have software-based disk ownership

You must cable the disk shelf loops on Node A directly to the Brocade switches.

Before you begin

- You can connect the disk shelves to any ports on that are not on the virtual channel reserved for the FC-VI adapter and the inter-switch link.
- Both disk shelf modules on the same loop must be connected to the same switch port number.
  For example, if the A Channel of the first loop for the local node's disks is connected to Switch 1, port 8, then the B Channel for that loop must be connected to Switch 2, port 8.
- Both switches at a site must be the same model and have the same number of licensed ports.

About this task



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. The example uses switch port 3.
   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: ESH2 and 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 when you have software-based disk ownership

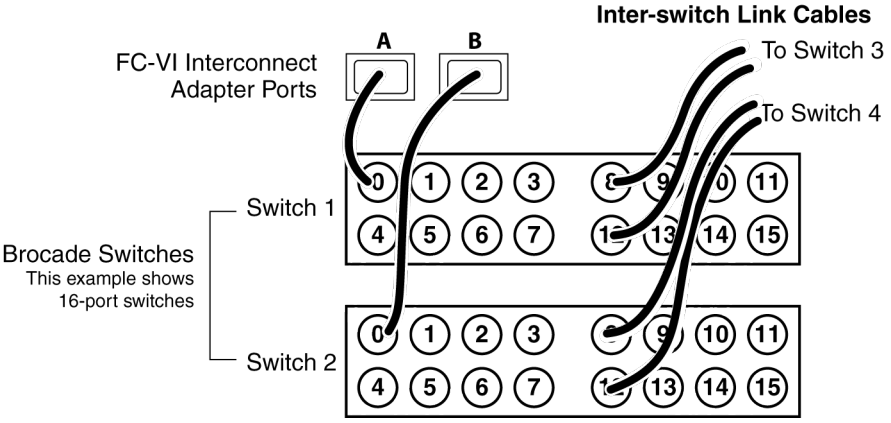
Describes how to cable the cluster interconnect and inter-switch link on Node A.

Before you begin

Each port on the Interconnect (IC) cards must be connected to the same fabric.

For example, if Port A of the IC card on the local node is connected to Switch 1, and Port A of the IC card on the remote node is connected to Switch 3, then Switch 1 and Switch 3 must be connected by the inter-switch link, thereby connecting them to the same fabric.

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 appropriate hardware and service guide.

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.

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 cluster interconnect to the switches, and ensure that the disk shelves in the configuration belong to the correct pools.

About this task

Complete the following tasks in the order shown:

Steps

1. *Cabling the controller when you have software-based disk ownership* on page 91
2. *Cabling the shelves when you have software-based disk ownership* on page 93
3. *Cabling the FC-VI adapter and inter-switch link when you have software-based disk ownership* on page 94

Cabling the controller when you have software-based disk ownership

You can use this procedure to cable the Fibre Channel ports on the controller to the Brocade switches when your system uses software-based disk ownership.

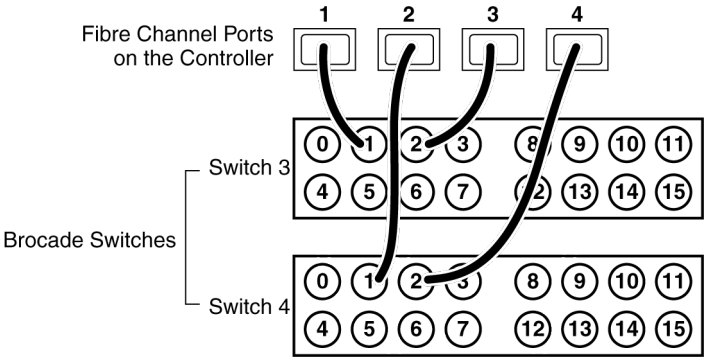
Before you begin

- Select one virtual channel on the switch for the cluster interconnect connections. The following examples use virtual channel 2, which includes ports 0, 4, 8, and 12.
- If you are using a dual-port HBA, connecting both ports of the HBA to the same switch port number can make it easier to cable and administer your MetroCluster. (However, this is not required.)

  For example, if port 1 of the HBA is connected to port 1 of Switch 1, you should connect port 2 of that HBA to port 1 of Switch 2.
- Both Fibre Channel ports on the same dual-port HBA (or adjacent pairs of onboard ports) should never be connected to the same switch. You must connect one port to one switch and the other port to the other switch.

  For example, if onboard port 0a is connected to Switch 3, you should not connect onboard port 0b to Switch 3; you must connect port 0b to Switch 4.

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.

   Again, 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 51

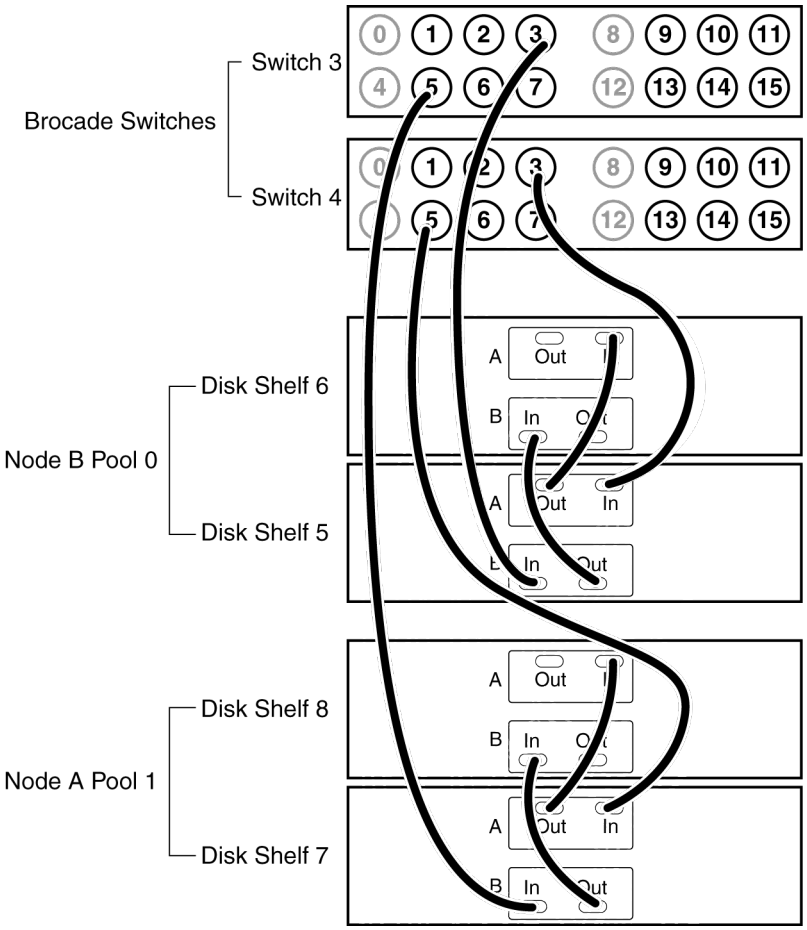
Cabling the shelves when you have software-based disk ownership

You must cable the disk shelf loops on Node B directly to the Brocade switches.

Before you begin

- You can connect the disk shelves to any ports on that are not on the virtual channel reserved for the FC-VI adapter and the inter-switch link.
- Both disk shelf modules on the same loop must be connected to the same switch port number.
  For example, if the A Channel of the first loop for the local node's disks is connected to Switch 1, port 8, then the B Channel for that loop must be connected to Switch 2, port 8.
- Both switches at a site must be the same model and have the same number of licensed ports.

About this task



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: ESH2 and 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.

Cabling the FC-VI adapter and inter-switch link when you have software-based disk ownership

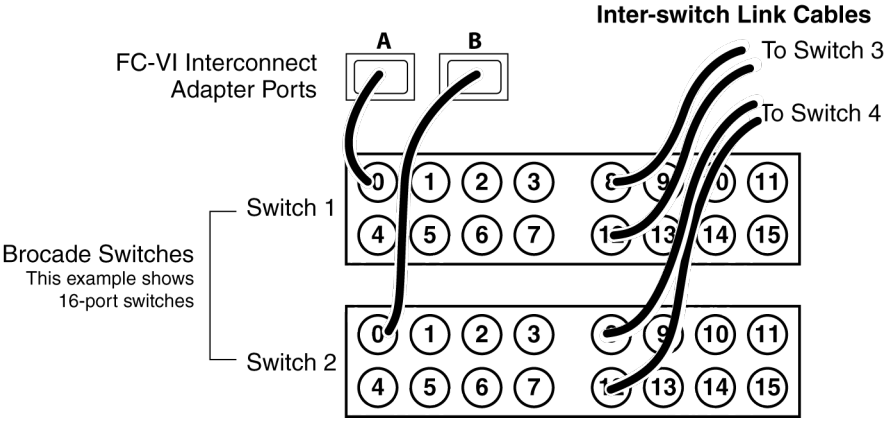
You must cable the cluster interconnect and inter-switch link on Node B.

Before you begin

Each port on the Interconnect (IC) cards must be connected to the same fabric.

For example, if port A of the IC card on the local node is connected to switch 1, and port A of the IC card on the remote node is connected to switch 3, then switch 1 and switch 3 must be connected by the inter-switch link, thereby connecting them to the same fabric.

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 appropriate hardware and service guide.

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.

   In the example we are using virtual channel 2, which includes ports 0, 4, 8, and 12, and are using 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 tasks

Assigning disk pools (if you have software-based disk ownership) on page 96

Assigning disk pools (if you have software-based disk ownership)

If your system uses software-based disk ownership, 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 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.* -p0`

   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.* -p1`

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:port12.* -p0`

   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:port0.* -p1`

   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 if you have software-based disk ownership

Use this procedure to verify your disk paths if you have software-based disk ownership.

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 information on the `aggr status` command.
3. Repeat steps 1 and 2 on Node B.

Required connections for using uninterruptible power supplies with MetroCluster configurations

You can use a UPS (Uninterruptible Power Supply) with your MetroCluster. 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 cannot shut down gracefully.

Reconfiguring an active/active configuration into two stand-alone systems

To divide an active/active configuration so that the nodes become stand-alone systems without redundancy, you must disable the active/active software features and then remove the hardware connections.

About this task

This procedure applies to all active/active configurations regardless of disk shelf type.

Steps

1. *Ensuring uniform disk ownership within disk shelves and loops in the system* on page 99
2. *Disabling the active/active software* on page 100
3. *Reconfiguring nodes using disk shelves for stand-alone operation* on page 101
4. *Requirements when changing a node using array LUNs to stand-alone* on page 103
5. *Reconfiguring nodes using array LUNs for stand-alone operation* on page 104

Ensuring uniform disk ownership within disk shelves and loops in the system

In systems using software-based disk ownership, 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.

About this task

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

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 active/active 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 active/active software.

Disabling the active/active software

You need to disable the active/active configuration in the software before reconfiguring the hardware to completely unconfigure the active/active feature.

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 can't 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 cluster license by entering the following command:

   ```
   license delete cluster
   ```

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.

Reconfiguring nodes using disk shelves for stand-alone operation

You can use this procedure to reconfigure the hardware if you want to return to a single-controller configuration.

Before you begin

You must disable the active/active software.

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 disk. 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*.

   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*.

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*.

   b. Use the `disk reassign` command and reassign all disk shelves 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. 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 cluster interconnect cables from both nodes. See the hardware documentation for your system for more details.
6. Move or remove the adapter used for the cluster interconnect:

If your system uses a...	Then...
cluster interconnect adapter or an FC-VI adapter	Remove the adapter from the system.
NVRAM5 or NVRAM6 adapter	You might need to change the slot position of the adapter. See the appropriate hardware and service 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
    ```

12. Check active/active configuration status by entering the following command:

    ```
    cf status
    ```

    If the active/active configuration is disabled, you see the following output:
    ```
    Failover monitor not initialized
    ```

13. Repeat Step 1 through Step 10 for the partner node.

Related concepts

Requirements when changing a node using array LUNs to stand-alone on page 103

Related tasks

Reconfiguring nodes using array LUNs for stand-alone operation on page 104

Requirements when changing a node using array LUNs to stand-alone

After uncoupling gateways in an active/active configuration, 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 active/active configuration.

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 active/active 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 active/active 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 the active/active software on page 100
Reconfiguring nodes using array LUNs for stand-alone operation on page 104

Reconfiguring nodes using array LUNs for stand-alone operation

After uncoupling the nodes in active/active configuration, 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

You must disable the active/active configuration software.

About this task

If you want both systems in the active/active configuration 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, will continue 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 cluster interconnect cables from both nodes. See the hardware documentation for your system for more details.
4. Move or remove the adapter used for the cluster interconnect.

If your system uses a...	Then...
cluster interconnect adapter or an FC-VI adapter	Remove the adapter from the system.
NVRAM5 or NVRAM6 adapter	You might need to change the slot position of the adapter. See the appropriate hardware and service guide and the N series Interoperability Matrices website at *www.ibm.com/systems/storage/network/interophome.html* 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`**
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*. **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`**
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 information about what you need to do to prepare for taking a system using array LUNs out of service.

Related tasks

Disabling the active/active software on page 100

Configuring an active/active configuration

Bringing up and configuring a standard or mirrored active/active configuration for the first time can require enabling licenses, setting options, configuring networking, and testing the configuration.

These tasks apply to all active/active configurations regardless of disk shelf type.

Steps

1. *Bringing up the active/active configuration* on page 107
2. *Enabling licenses* on page 110
3. *Setting options and parameters* on page 111
4. *Configuration of network interfaces* on page 117
5. *Testing takeover and giveback* on page 127

Bringing up the active/active configuration

The first time you bring up the active/active configuration, you must ensure that the nodes are correctly connected and powered up, and then use the setup program to configure the systems.

Considerations for active/active configuration setup

When the setup program runs on a storage system in an active/active configuration, it prompts you to answer some questions specific for active/active configurations.

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 virtual interfaces (VIFs) for your network interfaces?
  For information about VIFs, see the *Data ONTAP Network Management Guide*.

  Note: You are advised to use VIFs with active/active configurations 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 active/active configuration 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 concepts

Configuration of network interfaces on page 117

Related tasks

Configuring shared interfaces with setup on page 108
Configuring dedicated interfaces with setup on page 109
Configuring standby interfaces with setup on page 109

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 you are configuring.

   For 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.

   Note: The addresses for the local node and partner node can reside on different subnetworks.

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
   ```

4. Enter the IP address or interface name of the partner.

   For 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 VIF, you must use the interface name.

   Note: The addresses for the local node and partner node can reside on different subnetworks.

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.

   For 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.

   For 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.

   For 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 active/active configuration.

Before you begin

The licenses you need to add depend on the type of your active/active configuration. The following table outlines the required licenses for each configuration.

Note: If your system is a gateway system, you must enable the gateway license on each node in the active/active configuration.

Configuration type	Required licenses
Standard active/active configuration	cluster
Mirrored active/active configuration	• cluster • syncmirror_local
MetroCluster	• cluster • syncmirror_local • cluster_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
Cluster 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 active/active configurations

Some options must be the same on both nodes in the active/active configuration, while some can be different, and some are affected by failover events.

In an active/active configuration, options are one of the following types:

- Options that must be the same on both nodes for the active/active configuration 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 active/active configuration 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, using 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 comments in parentheses are set to the same value for both nodes. The comments are as follows:

```
Value might be overwritten in takeover
Same value required in local+partner
Same value in local+partner recommended
```

3. Correct any mismatched options using the following command:
   `options` *`option_name option_value`*

   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

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.

Caution: 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 on page 163

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 of N6200 series controller modules and chassis

N6200 series controller modules recognize that they are in an active/active configuration 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 active/active configuration, 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 use the hardware assisted takeover option to speed up takeover times. The option uses the remote management card to quickly communicate local status changes to the partner node, and has configurable parameters.

Hardware-assisted takeover

Hardware-assisted takeover enables systems with remote management cards to improve the speed with which takeover events are detected, thereby speeding up the takeover time.

When enabled, hardware-assisted takeover takes advantage of the remote management card capabilities to detect failures on the local machine that could require a takeover. If a failure is detected, the card sends an alert to the partner node and, depending on the type of failure, the partner performs the takeover. These alerts can speed takeover because the Data ONTAP takeover process on the partner does not have to take the time to verify that the failing system is no longer giving a heartbeat and confirm that a takeover is actually required.

The hardware-assisted takeover option (`cf.hw_assist`) is enabled by default.

Requirements for hardware-assisted takeover

The hardware-assisted takeover feature is available only on systems that support Remote LAN Modules (RLMs) and have the RLMs installed and set up. The remote management card provides remote platform management capabilities, including remote access, monitoring, troubleshooting, logging, and alerting features.

Although a system with an RLM on both nodes provides hardware-assisted takeover on both nodes, hardware-assisted takeover is also supported on active/active configurations in which only one of the

two systems has an installed RLM. The RLM does not have to be installed on both nodes in the active/active configuration. The RLM can detect failures on the system in which it is installed and provide faster takeover times if a failure occurs on the system with the RLM.

See the *Data ONTAP System Administration Guide* for information about setting up the RLM.

System events detected by remote management

A number of events can be detected by the remote management card and generate an alert. Depending on the type of alert received, the partner node initiates takeover.

Alert	Takeover initiated upon receipt?	Description
power_loss	Yes	Power loss on the node. The remote management card has a power supply that maintains power for a short period after a power loss, allowing it to report the power loss to the partner.
l2_watchdog_reset	Yes	L2 reset detected by the system watchdog hardware.
power_off_via_rlm	Yes	The remote management card was used to power off the system.
power_cycle_via_rlm	Yes	The remote management card was used to cycle the system power off and on.
reset_via_rlm	Yes	The remote management card was used to reset the system.
abnormal_reboot	No	Abnormal reboot of the node.
loss_of_heartbeat	No	Heartbeat message from the node no longer received by the remote management card. **Note:** This does not refer to the heartbeat messages between the nodes in the active/active configuration but the heartbeat between the node and its local remote management card.
periodic_message	No	Periodic message sent during normal hardware-assisted takeover operation.
test	No	Test message sent to verify hardware-assisted takeover operation.

Disabling and enabling the hardware-assisted takeover option

Hardware-assisted takeover is enabled by default on systems that use an RLM. Hardware-assisted takeover speeds the takeover process by using the RLM 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

The `cf.hw_assist.partner.address` option enables you to change the partner address used by the hardware-assisted takeover process on the remote management card. The default is the IP address on the e0a port of the partner. On an N6040, N6060, or N6070 system, if the partner's e0M interface has been configured, the IP address of the e0M interface is used. If the e0M interface has not been configured, e0a is used.

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_or_hostname`***

   Note: The hardware assisted takeover feature does not support IPv6 addresses when specifying the partner IP address in the `cf.hw_assist.partner.address` *`address_or_hostname`* option.

   If a host name is specified, the host name is resolved when this command is issued.

Setting the partner port for hardware-assisted takeover

When hardware-assisted takeover is enabled, the RLM 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`***

Configuration of network interfaces

If you didn't configure interfaces during system setup, you need to configure them manually to ensure continued connectivity during failover.

What the networking interfaces do

When a node in an active/active configuration 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 active/active configuration to maintain communication with the network if the partner fails.

See the *Data ONTAP Network Management Guide* for a description of available options and the function each performs.

Note: You should always use multiple NICs with VIFs to improve networking availability for both stand-alone storage systems and systems in an active/active configuration.

IPv6 considerations in an active/active configuration

When enabled, IPv6 provides features such as address autoconfiguration. Using these IPv6 features requires an understanding of how these features work with the active/active configuration functionality.

For additional information about IPv6, see the *Data ONTAP Administration Guide*.

Configuration requirements for using IPv6

To use IPv6 in an active/active configuration, 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.

IPv6 and hardware-assisted takeover

The hardware assisted takeover feature does not support IPv6 addresses when specifying the partner IP address in the `cf.hw_assist.partner.address` *`address_or_hostname`* option.

Configuring network interfaces for active/active configurations

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 active/active configuration

You can configure three types of interfaces on nodes in an active/active configuration.

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 active/active configuration.

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 active/active configuration 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 active/active configuration. Each type has a different role in normal and takeover mode.

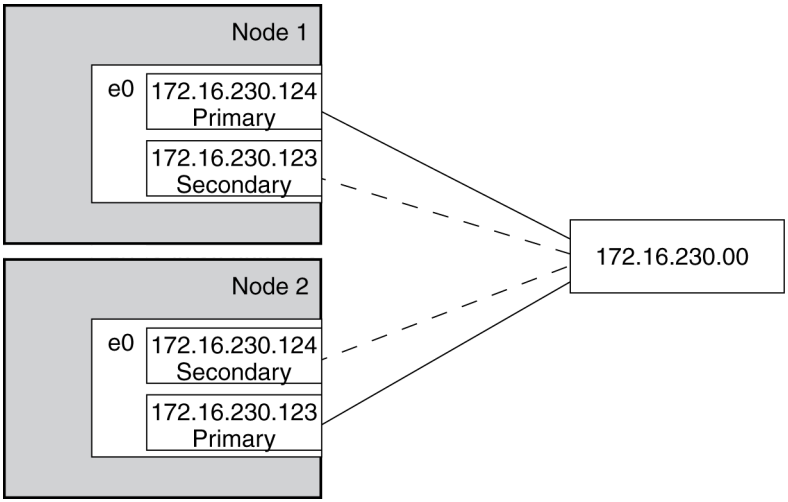
The following table shows the role of each interface type in normal and takeover mode.

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.



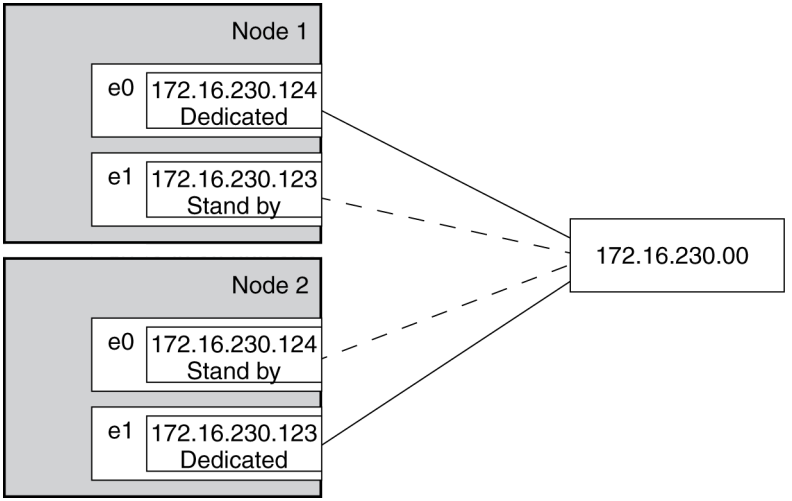
If Node 1 fails, interface e0 on Node 1 stops functioning, but the secondary address on e0 on Node 2 handles the Node 1 network connection with the 230 network.

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 active/active configuration.

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

Interface	Shared	Dedicated	Standby	Partner parameter
VLAN interface	X	X	X	IP address or interface name

Note: Some storage systems, such as the N6040, N6060, or N6070 systems, include an e0M interface that is dedicated to management traffic. This port can be partnered in an active/active configuration in the same way as a regular Ethernet interface.

Making nondisruptive changes to the VIFs

You can use the `cf takeover` and `cf giveback` commands to make changes to VIFs in the active/active configuration 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 active/active configuration, 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 VIFs.

   See the *Data ONTAP Network Management Guide* for more information about configuring VIFs.

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 VIFs.

4. Repeat these steps, making any required changes to the `/etc/rc` file on the partner node.

Configuring dedicated and standby interfaces

You can configure dedicated and standby interfaces for an active/active configuration, two on each node, so that even in the event of a takeover each node still has a dedicated interface.

Before you begin

Both nodes in the active/active configuration 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.

  Note: For MetroCluster configurations, if you use the `/etc/mcrc` file and enable the `cf.takeover.use.mcrc_file`, the addresses for the local node and partner node can reside on different subnetworks.
- 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 VIFs to improve networking availability for both stand-alone storage systems and systems in an active/active configuration.

About this task

Keep in mind that you can use interface names for specifying all interfaces.

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 active/active configuration to use FC, see the *Data ONTAP Block Access Management Guide for iSCSI and FC*.

Steps

1. On nodeA, enter the following command on the command line and also enter it in the `/etc/rc` file, so that the command is permanent:

   `ifconfig` *`interfaceA1addressA1`* `{`*`other_options`*`}`

   `interfaceA1` is the name of the dedicated local interface for nodeA.

   `addressA1` is the IP address of the dedicated local interface for nodeA.

   `other_options` denotes whatever other options are needed to correctly configure the interface in your network environment.

   The dedicated local interface for nodeA is configured.

2. Also on nodeA, enter the following command on the command line and in the `/etc/rc` file:

   `ifconfig` *`interfaceA2`* `partner` *`addressB1`*

   `interfaceA2` is the name of the standby interface for nodeA.

   `addressB1`

   Note: When you configure virtual interfaces for takeover, you must specify the interface name and not the IP address.

The standby interface for nodeA is configured to take over the dedicated interface of nodeB on takeover.

3. On nodeB, enter the following command on the command line and in the `/etc/rc` file:

   ifconfig *interfaceB1addressB1* {*other_options*}

   interfaceB1 is the name of the dedicated local interface for nodeB.

   addressB1 is the IP address of the dedicated local interface for nodeB.

   other_options denotes whatever other options are needed to correctly configure the interface in your network environment.

   The dedicated local interface for nodeB is configured.

4. Also on nodeB, enter the following command on the command line and in the `/etc/rc` file:

   ifconfig *interfaceB2* partner *addressA1*

   interfaceB2 is the name of the standby interface on nodeB.

   addressA1 is the IP address or interface name of the dedicated interface for nodeA.

   Note: When you configure virtual interfaces for takeover, you must specify the interface name and not the IP address.

   The standby interface on nodeB is configured to take over the dedicated interface of nodeA on takeover.

After you finish

If desired, configure your interfaces for automatic takeover in case of NIC failure.

Configuring partner addresses on different subnets (MetroClusters 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 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

   Enter the result of your step here (optional).

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 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 active/active configuration, verify that it operates as expected.

Steps

1. Check the cabling on the cluster 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`

   The local node takes over the partner node and gives the following output:
   `takeover completed`
4. Test communication between the local node and partner node.

   Example
   You can use the `fcstat device_map` command to ensure that one node can access the other node's disks.
5. Give back the partner node 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 got the message that giveback was completed successfully.

If takeover and giveback...	Then...
Is completed successfully	Repeat Step 2 through Step 5 on the partner node
Fails	Attempt to correct the takeover or giveback failure

Managing takeover and giveback

An active/active configuration 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 active/active configuration 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 active/active configurations regardless of disk shelf type.

How takeover and giveback work

Takeover is the process in which a node takes over the storage of its partner. Giveback is the process in which the storage is returned to the partner. You can initiate the processes in different ways. A number of things that affect the active/active configuration occur when takeover and giveback take place.

When takeovers occur

The conditions under which takeovers occur depend on how you configure the active/active configuration.

Takeovers can be initiated when one of the following conditions occur:

- A node is in an active/active configuration that is configured for immediate takeover on panic, and it undergoes a software or system failure that leads to a panic.
- A node that is in an active/active configuration 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 MetroClusters, you can initiate a forced takeover in this situation.
- There is a mismatch between the disks, array LUNs, or both that one node can see and those that the other node can see.
- 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 flag.
- You initiate a takeover manually.

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 the partner node is repaired and is operating normally, you can use the `giveback` command to return operation to the partner.

When the failed node is functioning again, the following events can occur:

- You initiate a `giveback` command that terminates the emulated node on the partner.
- The failed node resumes normal operation, serving its own data.
- The active/active configuration resumes normal operation, with each node ready to take over for its partner if the partner fails.

Management of an active/active configuration in normal mode

You manage an active/active configuration in normal mode by performing a number of management actions.

Monitoring active/active configuration 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 active/active configuration is up.

Step

1. Enter the following command:

   `cf status`

   Example

   ```
   node1>
   cf status
                            Cluster enabled, node2 is up.
   ```

   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 why failover is disabled.

   This verifies the link between the nodes and tells you that both filer1 and filer2 are functioning and available for takeover.

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 status

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 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 active/active configuration and the remote management card are operating.

Checking hardware-assisted takeover statistics

You can display statistics about hardware-assisted takeovers with the `cf hw_assist stats` command.

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 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
```

Description of active/active configuration status messages

The `cf status` command displays information about the status of the active/active configuration.

The following table shows some of the messages that the `cf status` command can display.

Message	Meaning
cluster enabled, partner_name is up.	The active/active configuration is operating normally.
partner_name_1 has taken over partner_name_2.	One node took over the other node.
Interconnect not present.	The system does not recognize the existence of a cluster interconnect adapter.
Interconnect is down.	The cluster interconnect adapter cannot access the partner. This might be due to cabling problems or the partner might be down.
Interconnect is up.	The cluster 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 and has disabled takeover by partner_name_1.	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.

Message	Meaning
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.

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 active/active configuration is new, this command returns “partner”.

Displaying disk and array LUN information on an active/active configuration

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 active/active configuration.
- 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 channelB is a file system disk, spare disk, or parity disk.

Step

1. Enter one of the following commands:

   `sysconfig -r`
   or

   `aggr status -r`

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 automatic takeover of a panicked partner

A node can be configured so it takes over immediately when its partner panics. This shortens the time between the initial failure and when service is fully restored, because the takeover can be quicker than recovery from the panic, but the subsequent giveback causes another brief outage.

Steps

1. Ensure that you enabled controller takeover by entering the following command:

   `cf enable`

2. Enter the following command to enable or disable takeover on panic:

   `options cf.takeover.on_panic [on|off]`

   on Enables immediate takeover of a failed partner or off to disable immediate takeover. This is the default value.

   off Disables immediate takeover. If you disable this option, normal takeover procedures apply. The node still takes over if its partner panics, but might take longer to do so.

   Note: If you enter this command on one node, the value applies to both nodes.

   The setting of this option is persistent across reboots.

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
   ```

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 active/active configuration 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 active/active configuration. The following table outlines which system errors can cause a takeover to occur, and whether you can configure the active/active configuration 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	Off, unless FC or iSCSI is licensed.	
There is a mismatch between the disks, array LUNs, or both that one node can see and those that the other node can see.	`cf.takeover.on_disk_shelf_miscompare` set to On	Off	

System error	Option used to configure	Default value	Notes
All the network interface cards (NICs) or VIFs 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*.
One or more of the NICs or VIFs enabled for negotiated failover failed. **Note:** If interfaces fail on both nodes in the active/active configuration, 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 VIF for negotiated failover, you use the `ifconfig if_name nfo` command. For more information, see the *Data ONTAP MultiStore Management Guide*.
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.

Related concepts

How disk shelf comparison takeover works on page 141

Related tasks

Enabling and disabling automatic takeover of a panicked partner on page 136

Commands for performing a manual takeover

Lists and describes the commands you can use when initiating a takeover. You can initiate a takeover on a node in an active/active configuration to perform maintenance on that node while still serving the data on its disks to users.

You can initiate a takeover on a node in an active/active configuration to perform maintenance on that node while still serving the data on its disks, array LUNs, or both to users. The following table lists and describes the commands you can use when initiating a takeover:

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 cluster monitor to ignore some configuration problems that would otherwise prevent a takeover, such as unsynchronized NVRAM due to a faulty cluster interconnect connection. It then initiates a takeover of the partner of the local node.

Command	Description
`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 cluster_remote license is enabled. **Attention:** Use the `-d` option only after you verify that the partner is down. **Note:** 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*.
`cf takeover -n`	Initiates a takeover for a nondisruptive upgrade. For more information, see the *Data ONTAP Upgrade Guide*.

Specifying the time period before takeover

You can specify how long (in seconds) a partner in an active/active configuration 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 active/active configuration 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.

How disk shelf comparison takeover works

A node uses disk shelf comparison with its partner node to determine if it is impaired.

When communication between nodes is first established through the cluster interconnect adapters, the nodes exchange a list of disk shelves that are visible on the A and B loops of each node. If, later, a system sees that the loop B disk shelf count on its partner is greater than its local loop A disk shelf count, the system concludes that it is impaired and prompts its partner to initiate a takeover.

Note: Disk shelf comparison does not function for active/active configurations using software-based disk ownership, or fabric-attached MetroClusters.

Configuring VIFs or interfaces for automatic takeover

After you configure your interfaces or VIFs to allow takeovers and givebacks to be completed successfully, you can also optionally configure them to trigger automatic takeover if any or all of them experience a persistent failure.

Steps

1. For every VIF or interface on which you want to enable automatic takeover, enter the following command:

   `ifconfig interface_name nfo`

2. Update the /etc/rc file with the command that you entered so that your changes persist across reboots.

3. The default policy is that takeover only occurs if all the NICs or VIFs on a node that are configured for automatic takeover fail. If you want takeover to occur if any NIC or VIF configured for automatic takeover fails, enter the following command:

   `options cf.takeover.on_network_interface_failure.policy any_nic`

4. Enter the following command to enable takeover on interface failures:

   `options cf.takeover.on_network_interface_failure enable`

   Note: If interfaces fail on both nodes in the active/active configuration, takeover won't occur.

Takeover of vFiler units and the vFiler unit limit

The vFiler limit, set with the `vfiler limit` command, determines how many vFiler units can exist on a system. In an active/active configuration, if the two systems have different vFiler limits, some vFiler units might not be taken over in the event of a takeover.

When performing a takeover, a system can take over only the number of vFiler units that were specified by that system's vFiler unit 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 had a higher vFiler limit, some vFiler units will not be successfully taken over.

For more information about setting the vFiler limit, see the *Data ONTAP MultiStore Management Guide*.

Managing an active/active configuration in takeover mode

You manage an active/active configuration 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

Describes how to 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 showing 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 with arguments

Describes how to use the `partner` command with a Data ONTAP command as an argument.

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

You can also access the emulated node remotely using a Remote Shell (rsh) connection. You cannot access the emulated node using Secure Shell (ssh) or Telnet.

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 cluster
- rdate
- reboot
- timezone

Commands with different behaviors

Command	Difference
ifconfig [interface]	Displays the following: • Emulated interface mappings based on the failed node's `/etc/rc` file rather than the takeover node interface mappings **Note:** MetroCluster nodes use the failed node's `/etc/mcrc` file if the `cf.takeover.use_mcrc_file` option is enabled. • Emulated interface names rather than the interface names of the takeover node • 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 active/active configuration.
sysconfig	When it displays hardware information, the sysconfig command 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, the `aggr` status command 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 active/active configuration.
- 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 ONTAP Data Protection Tape Backup and Recovery Guide*.

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:

```
filer1 (takeover)> partner
filer1/filer2> restore [options [arguments]]
filer1 (takeover)> partner
```

Example

Issuing a `restore` command as an argument in the `partner` command:

```
filer1 (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.

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 *Storage Management Guide*.

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

On a fabric-attached MetroCluster, before you undertake the giveback operation, you must rejoin the aggregates on the surviving node and the partner node 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 use the 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 cluster 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*.
2. Check for the message
   `cf.giveback.disk.check.fail`
   on the console. 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 active/active configuration.
3. Check the cluster 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:
   - Outstanding CIFS sessions
   - RAID disk additions
   - Volume creation (traditional volume or FlexVol volume)
   - Aggregate creation
   - Disk ownership assignment
   - Disks being added to a volume (vol add)
   - Snapshot copy creation, deletion, or renaming
   - Quota initialization
   - Advanced mode repair operations, such as wafliron
   - Storage system panics
   - Backup dump and restore operations
   - SnapMirror transfers (if the partner is a SnapMirror destination)
   - SnapVault restorations
   - Disk sanitization operations

   If any of these processes are taking place, either cancel the process or wait until it is complete, and then try the giveback operation again.

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 150

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.

Step

1. On the takeover node, enter the following command:

   ```
   cf forcegiveback
   ```

   Attention: Use `cf forcegiveback` only when you cannot get `cf giveback` 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 when 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.

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`.

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.

Option for terminating long-running processes

Describes the `cf.giveback.auto.terminate.bigjobs` option, which, when on, specifies that automatic giveback should immediately terminate long-running operations.

The `cf.giveback.auto.terminate.bigjobs` option, when on, specifies that automatic giveback should immediately terminate long-running operations (dump/restore, vol verify, and so on) when initiating an automatic giveback. When this option is off, the automatic giveback is deferred until the long-running operations are complete. This option is used only if automatic giveback is On.

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 `on` by default.

Configuring automatic giveback

You can enable automatic giveback by using the cf.giveback.auto.enable option.

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.

Step

1. Enable the following option to enable automatic giveback: `cf.giveback.auto.enable on`. The `on` value enables automatic giveback. The `off` value disables automatic giveback. This option is `off` by default.

Troubleshooting takeover or giveback failures

If takeover or giveback fails for an active/active configuration, you need to check the cluster 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 cluster adapter error	Check the cluster 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.

If the error message indicates...	Then...
A networking error	Check for network connectivity. See the *Data ONTAP MultiStore Management Guide* for more information.

3. Correct any errors or differences displayed in the output.
4. Reboot the active/active configuration and rerun the takeover and giveback tests.
5. If you still do not have takeover enabled, contact technical support.

Managing EXN1000, EXN2000, or EXN4000 units in an active/active configuration

You must follow specific procedures to add disk shelves to an active/active configuration or a MetroCluster, or to upgrade or replace disk shelf hardware in an active/active configuration.

If your configuration includes SAS disk shelves, refer to the following documents on the N series support website at *www.ibm.com/storage/support/nseries/*:

- For SAS disk shelf management, see the *Hardware and Service Guide* for your disk shelf model.
- For cabling SAS disk shelves in an active/active configuration, see the *Universal SAS and ACP Cabling Guide*.

Note: Fabric-attached MetroCluster configurations do not support SAS disk shelves.

Adding EXN1000, EXN2000, or EXN4000 units to a multipath HA loop

To add supported EXN1000 or EXN2000 unit or EXN4000 units to an active/active configuration 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 task 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 there are not two paths 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 appropriate Storage Expansion *Hardware and Service Guide*.
3. Determine whether disk shelf counting is enabled by entering the following command:

   ```
   options cf.takeover.on_disk_shelf_miscompare
   ```

4. If the disk shelf counting option is set to On, turn it off by entering the following command:

```
options cf.takeover.on_disk_shelf_miscompare off
```

5. 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 6 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.

6. 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.

7. 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.
8. Connect the cable and SFP you removed in Step 6 to the Channel A Output port of the new disk shelf.
9. If you disabled the adapter in Step 5, reenable the adapter by entering the following command:

   ```
   fcadmin config -e <adapter>
   ```

10. Repeat Step 6 through Step 9 for Channel B.

    Note: The Channel B Output port is connected to the other controller.

11. Confirm that there are two paths to every disk by entering the following command:

    ```
    storage show disk -p
    ```

    There should be two paths listed for every disk.

12. If disk shelf counting was Off, reenable it by entering the following command:

    ```
    options cf.takeover.on_disk_shelf_miscompare on
    ```

Related tasks

Determining path status for your active/active configuration on page 159

Upgrading or replacing modules in an active/active configuration

In an active/active configuration 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 disk shelves.

Note: If your configuration includes SAS disk shelves, refer to the following documents on the N series support website at *www.ibm.com/storage/support/nseries/*:

- For SAS disk shelf management, see the *Hardware and Service Guide* for your disk shelf model.
- For cabling SAS disk shelves in an active/active configuration, see the *Universal SAS and ACP Cabling Guide*.

About the disk shelf modules

A disk shelf module (ESH2, 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 ESH2, 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 active/active configuration, make sure that you understand the restrictions.

- You cannot mix ESH2 or ESH4 modules in the same loop with AT-FCX modules.
- You cannot mix ESH and ESH4 modules in the same loop.

Best practices for changing module types

If you plan to change the type of any module in your active/active configuration, make sure that you review the best practice guidelines.

- Whenever you remove a module from an active/active configuration, 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 active/active configuration, 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: ESH2 and ESH4 modules are self-terminating; this guideline does not apply to ESH2 and 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.
- ESH2 and ESH4 modules should not be on the same disk shelf loop.

Related concepts

Understanding redundant pathing in active/active configurations on page 158

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.

Understanding redundant pathing in active/active configurations

Some active/active configurations have two paths from each controller to each of their disk shelves or array LUNs; this configuration is called a redundant-path or multipath configuration. Active/active

configurations using Multipath HA are redundant-path configurations. Fabric-attached MetroClusters are also redundant-path configurations.

Determining path status for your active/active configuration

You can determine whether any module in your system provides the only path to any disk or array LUN by using the `storage show disk -p` command at your system console.

About this task

If you want to remove a module from your active/active configuration, 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. On the other hand, if that module provides the only path to any of the disks or array LUNs in your active/active configuration, 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 or array LUN in the active/active configuration:

   - 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 active/active configuration 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 or array LUN has two paths. Even in an active/active configuration configured for redundant paths, a hardware or configuration problem can cause one or more disks to have only one path. If any disk or array LUN in your active/active configuration has only one path, you must treat that loop as if it were in a single-path active/active configuration when removing modules.

The following example shows what the `storage show disk -p` command output might look like for an active/active configuration 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 active/active configuration, 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 Node A.
   b. At the Node B console, enter the following command:

      ```
      cf takeover
      ```

   c. Wait for takeover to be complete and make sure that the partner node, or Node A, reboots and is waiting for giveback.

   Any module in the loop that is attached to Node A can now be replaced.
4. Note whether the module you are replacing has a terminate switch. If it does, set the terminate switch of the new module to the same setting.

   Note: The ESH2 and ESH4 are self-terminating and do not have a terminate switch.

5. Put on the antistatic wrist strap and grounding leash.
6. Disconnect the module that you are removing from the Fibre Channel cabling.
7. 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.

8. 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.

9. Recable the disk shelf to its original location.
10. 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.

11. If you performed a takeover previously, complete the following steps:

    a. At the console of the takeover node, return control of Node B's disk shelves by entering the following command:

       ```
       cf giveback
       ```

    b. Wait for the giveback to be completed before proceeding to the next step.

12. Test the replacement module.
13. Test the configuration.

Related concepts

Best practices for changing module types on page 157

Related tasks

Determining path status for your active/active configuration on page 159

Disaster recovery using MetroCluster

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.

Note: If you are a gateway system customer, see the *Gateway MetroCluster Guide* for information about configuring and operating a gateway system in a MetroCluster configuration.

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

If you can reestablish the MetroCluster connection after fixing the problem, you should not perform the disaster recovery procedure.

Do not perform the disaster recovery procedure for the following failures:

- A failure of the cluster interconnect between the two sites. This 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 returns to normal operation.
- The storage from one site (site A) is not accessible to the node at the other site (site B). This 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 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 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 returns to normal operation.

Recovering from a disaster

After determinating 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 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 assure 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 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*.

Steps

1. *Restricting access to the disaster site node* on page 165
2. *Forcing a node into takeover mode* on page 166
3. *Remounting volumes of the failed node* on page 166
4. *Recovering LUNs of the failed node* on page 167
5. *Fixing failures caused by the disaster* on page 168
6. *Reestablishing the MetroCluster configuration* on page 169

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 165
2. *Restricting access to the node by fencing off* on page 165

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 cluster 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*.

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 112

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 Block Access Management Guide for iSCSI and FC*.

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.

Before you begin

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 MetroCluster-configured partner 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

Describes how to reestablish a MetroCluster 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

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 to normal operation

Describes how to return the MetroCluster to normal operation by re-creating the MetroCluster mirror.

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...`

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*

   For more information about the SyncMirror feature, see the *Data ONTAP Data Protection Online Backup and Recovery Guide*.

5. Re-create the mirrored aggregates by entering the following command for each aggregate that was split:

   aggr mirror *aggr_name*

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.

First, the aggregate status of the disaster site's storage after reestablishing access to the partner at the surviving site is shown.

```
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.
```

Nondisruptive operations with active/active configurations

By taking advantage of an active/active configuration'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 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 Guide*
Upgrade the controller	*Hardware System Upgrade Procedures*
Replace a disk shelf	*Hardware and Service Guide* for your disk shelf
Replace a hardware FRU component	FRU procedures for your platform

Controller failover and single-points-of-failure

A storage system has a variety of SPOFs that you can reduce by using an active/active configuration. In an active/active configuration, there are a number of failures that can cause a controller to fail over.

Benefits of controller failover

You can use controller failover, a high-availability data service solution, to further increase the uptime of storage systems. It protects against controller failure by transferring the data service from the failed node to its partner node. Controller failover can also protect against other hardware failures, such as problems with network interface cards, Fibre Channel-Arbitrated Loops (FC-AL loops), SAS loops, and disk shelf modules. Controller failover is also an effective tool for reducing planned downtime of one of the nodes.

Note: You might also see the term *cluster failover*; this is equivalent to the term *controller failover* used in this document.

Single-point-of-failure definition

Explains what a single-point-of-failure is in the context of your storage system.

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 active/active configurations

Enables you to see which hardware components are SPOFs, and how controller failover can eliminate these SPOFs to improve data availability.

Hardware components	**SPOF**		**How controller failover eliminates SPOF**
	Stand-alone	**Active/ active configuration**	
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 gracefully shuts down and automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
Multiple NICs with VIFs (virtual interfaces)	No	No	If one of the networking links within a VIF 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 a VIF fail, the node automatically fails over to its partner node, if failover is enabled for the VIF.
Single NIC	Yes	No	If a NIC fails, the node automatically fails over to its partner node, if failover is enabled for the NIC.

Hardware components	**SPOF**		**How controller failover eliminates SPOF**
	Stand-alone	**Active/ active configuration**	
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)	Yes	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	Yes	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.
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.

Hardware components	SPOF		How controller failover eliminates SPOF
	Stand-alone	Active/ active configuration	
Power supply	No	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)	No	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.
Cluster adapter	N/A	No	If a cluster adapter fails, the failover capability is disabled but both nodes continue to serve data to their respective applications and users.
Cluster interconnect cable	N/A	No	The cluster interconnect adapter supports dual cluster 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 active/active configurations. The configuration responds differently depending on the event and the type of active/active configurations.

Cause-and-effect table for standard or mirrored active/active configurations

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	
			Single Storage System	Standard or Mirrored
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, there won't be a takeover; 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?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	
			Single Storage System	Standard or Mirrored
Single HBA (initiator) failure, Loop B	No	Yes, unless you are using SyncMirror or multipath HA and the mailbox disks aren't 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.
ESH2 or 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.
ESH2 or 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?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	
			Single Storage System	**Standard or Mirrored**
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 isn't 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?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	
			Single Storage System	**Standard or Mirrored**
Controller, dual power failure	Yes	Yes, until power is restored.	No	Yes, if failover succeeds.
Cluster interconnect failure (1 port)	No	No	n/a	Yes
Cluster interconnect failure (both ports)	No	Yes	n/a	Yes
Ethernet interface failure (primary, no VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (primary, VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (secondary, VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (VIF, all ports)	Yes, if set up to do so	No	Yes	Yes
Tape interface 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?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	
			Single Storage System	Standard or Mirrored
Heat exceeds permissible amount	Yes	No	No	No
Fan failures (disk shelves or controller)	No	No	Yes	Yes
Reboot	No	No	Maybe. Depends on cause of reboot.	Maybe. Depends on cause of reboot.
Panic	No	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.

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 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.
ESH2 or AT-FCX failure (Loop A)	No	No	Yes	Yes, with no failover necessary.
ESH2 or AT-FCX failure (Loop B)	No	No	Yes	Yes
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.

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**
Cluster interconnect failure (1 port)	No	No	Yes	Yes
Cluster interconnect failure (both ports)	No	No. Failover is possible.	Yes	Yes
Ethernet interface failure (primary, no VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (primary, VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (secondary, VIF)	Yes, if set up to do so	No	Yes	Yes
Ethernet interface failure (VIF, 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

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**
Fan failures (disk shelves or controller)	No	No	Yes	Yes
Reboot	No	No	Maybe. Depends on cause of reboot.	Maybe. Depends on cause of reboot.
Panic	No	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* 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 active/active configuration information • Moved Brocade switch configuration to Brocade Switch Description Page. • Moved from *cluster* to *active/active configuration* • Added information about Multipath Storage for Active/active configurations	Data ONTAP 7.1.1	June 2006

Feature updates	Feature first implemented in	Feature release date
• Generalized standard and mirrored active/active configuration cabling instructions • Updated standard and mirrored active/active configuration cabling instructions to include N7600, N7700, N7800, or N7900 • Changed name of document from *Cluster Installation and Administration Guide* to *Active/Active Configuration Guide*. • Added N7600, N7700, N7800, or N7900 information • Updated and extended active/active configuration 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 Active/active configurations.	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 active/active configuration	Data ONTAP 7.2.4	November 2007

Feature updates	Feature first implemented in	Feature release date
• Added support for N3300/ N3600and N6040/N6070 systems • Added procedures for the hardware-assisted takeover feature • Added information about disk shelves on gateway Active/active configurations	Data ONTAP 7.3.0	August 2008
• Added gateway content • Added references to the EXN3000 disk shelf documentation • Added support for 672 disks in MetroClusters • Added MetroCluster support for the Brocade 300 and 5100 switches • Add support for MetroCluster nodes on separate subnetworks	Data ONTAP 7.3.2	October 2009
• 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 EXN3500 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

Abbreviations

A list of abbreviations and their spelled-out forms are included here for your reference.

A
ABE (Access-Based Enumeration)
ACE (Access Control Entry)
ACL (access control list)
ACP (Alternate Control Path)
AD (Active Directory)
ALPA (arbitrated loop physical address)
ALUA (Asymmetric Logical Unit Access)
AMS (Account Migrator Service)
API (Application Program Interface)
ARP (Address Resolution Protocol)
ASCII (American Standard Code for Information Interchange)
ASP (Active Server Page)
ATA (Advanced Technology Attachment)

B
BCO (Business Continuance Option)
BIOS (Basic Input Output System
BCS (block checksum type)
BLI (block-level incremental)
BMC (Baseboard Management Controller)

C
CD-ROM (compact disc read-only memory)
CDDI (Copper Distributed Data Interface)
CDN (content delivery network)
CFE (Common Firmware Environment)
CFO (controller failover)
CGI (Common Gateway Interface)
CHA (channel adapter)
CHAP (Challenge Handshake Authentication Protocol)
CHIP (Client-Host Interface Processor)
CIDR (Classless Inter-Domain Routing)
CIFS (Common Internet File System)
CIM (Common Information Model)
CLI (command-line interface)
CP (consistency point)
CPU (central processing unit)
CRC (cyclic redundancy check)
CSP (communication service provider)

D
DAFS (Direct Access File System)
DBBC (database consistency checker)
DCE (Distributed Computing Environment)
DDS (Decru Data Decryption Software)
dedupe (deduplication)
DES (Data Encryption Standard)
DFS (Distributed File System)
DHA (Decru Host Authentication)
DHCP (Dynamic Host Configuration Protocol)
DIMM (dual-inline memory module)
DITA (Darwin Information Typing Architecture)
DLL (Dynamic Link Library)
DMA (direct memory access)
DMTD (Distributed Management Task Force)
DNS (Domain Name System)
DOS (Disk Operating System)
DPG (Data Protection Guide)
DTE (Data Terminal Equipment)

E

ECC (Elliptic Curve Cryptography) or (EMC Control Center)

ECDN (enterprise content delivery network)

ECN (Engineering Change Notification)

EEPROM (electrically erasable programmable read-only memory)

EFB (environmental fault bus)

EFS (Encrypted File System)

EGA (Enterprise Grid Alliance)

EISA (Extended Infrastructure Support Architecture)

ELAN (Emulated LAN)

EMU environmental monitoring unit)

ESH (embedded switching hub)

F

FAQs (frequently asked questions)

FAS (fabric-attached storage)

FC (Fibre Channel)

FC-AL (Fibre Channel-Arbitrated Loop)

FC SAN (Fibre Channel storage area network)

FC Tape SAN (Fibre Channel Tape storage area network)

FC-VI (virtual interface over Fibre Channel)

FCP (Fibre Channel Protocol)

FDDI (Fiber Distributed Data Interface)

FQDN (fully qualified domain name)

FRS (File Replication Service)

FSID (file system ID)

FSRM (File Storage Resource Manager)

FTP (File Transfer Protocol)

G
GbE (Gigabit Ethernet)
GID (group identification number)
GMT (Greenwich Mean Time)
GPO (Group Policy Object)
GUI (graphical user interface)
GUID (globally unique identifier)

H
HA (high availability)
HBA (host bus adapter)
HDM (Hitachi Device Manager Server)
HP (Hewlett-Packard Company)
HTML (hypertext markup language)
HTTP (Hypertext Transfer Protocol)

I

IB (InfiniBand)

IBM (International Business Machines Corporation)

ICAP (Internet Content Adaptation Protocol)

ICP (Internet Cache Protocol)

ID (identification)

IDL (Interface Definition Language)

ILM (information lifecycle management)

IMS (If-Modified-Since)

I/O (input/output)

IP (Internet Protocol)

IP SAN (Internet Protocol storage area network)

IQN (iSCSI Qualified Name)

iSCSI (Internet Small Computer System Interface)

ISL (Inter-Switch Link)

iSNS (Internet Storage Name Service)

ISP (Internet storage provider)

J

JBOD (just a bunch of disks)

JPEG (Joint Photographic Experts Group)

K

KB (Knowledge Base)

Kbps (kilobits per second)

KDC (Kerberos Distribution Center)

L
LAN (local area network)
LBA (Logical Block Access)
LCD (liquid crystal display)
LDAP (Lightweight Directory Access Protocol)
LDEV (logical device)
LED (light emitting diode)
LFS (log-structured file system)
LKM (Lifetime Key Management)
LPAR (system logical partition)
LREP (logical replication tool utility)
LUN (logical unit number)
LUSE (Logical Unit Size Expansion)
LVM (Logical Volume Manager)

M
MAC (Media Access Control)
Mbps (megabits per second)
MCS (multiple connections per session)
MD5 (Message Digest 5)
MDG (managed disk group)
MDisk (managed disk)
MIB (Management Information Base)
MIME (Multipurpose Internet Mail Extension)
MMC (Microsoft Management Console)
MMS (Microsoft Media Streaming)
MPEG (Moving Picture Experts Group)
MPIO (multipath network input/output)
MRTG (Multi-Router Traffic Grapher)
MSCS (Microsoft Cluster Service
MSDE (Microsoft SQL Server Desktop Engine)
MTU (Maximum Transmission Unit)

N
NAS (network-attached storage)
NDMP (Network Data Management Protocol)
NFS (Network File System)
NHT (N series Health Trigger)
NIC (network interface card)
NMC (Network Management Console)
NMS (network management station)
NNTP (Network News Transport Protocol)
NTFS (New Technology File System)
NTLM (NetLanMan)
NTP (Network Time Protocol)
NVMEM (nonvolatile memory management)
NVRAM (nonvolatile random-access memory)

O
OFM (Open File Manager)
OFW (Open Firmware)
OLAP (Online Analytical Processing)
OS/2 (Operating System 2)
OSMS (Open Systems Management Software)
OSSV (Open Systems SnapVault)

P
PC (personal computer)
PCB (printed circuit board)
PCI (Peripheral Component Interconnect)
pcnfsd (storage daemon)
(PC)NFS (Personal Computer Network File System)
PDU (protocol data unit)
PKI (Public Key Infrastructure)
POP (Post Office Protocol)
POST (power-on self-test)
PPN (physical path name)
PROM (programmable read-only memory)
PSU power supply unit)
PVC (permanent virtual circuit)

Q
QoS (Quality of Service)
QSM (Qtree SnapMirror)

R
RAD (report archive directory)
RADIUS (Remote Authentication Dial-In Service)
RAID (redundant array of independent disks)
RAID-DP (redundant array of independent disks, double-parity)
RAM (random access memory)
RARP (Reverse Address Resolution Protocol)
RBAC (role-based access control)
RDB (replicated database)
RDMA (Remote Direct Memory Access)
RIP (Routing Information Protocol)
RISC (Reduced Instruction Set Computer)
RLM (Remote LAN Module)
RMC (remote management controller)
ROM (read-only memory)
RPM (revolutions per minute)
rsh (Remote Shell)
RTCP (Real-time Transport Control Protocol)
RTP (Real-time Transport Protocol)
RTSP (Real Time Streaming Protocol)

S
SACL (system access control list)
SAN (storage area network)
SAS (storage area network attached storage) or (serial-attached SCSI)
SATA (serial advanced technology attachment)
SCSI (Small Computer System Interface)
SFO (storage failover)
SFSR (Single File SnapRestore operation)
SID (Secure ID)
SIMM (single inline memory module)
SLB (Server Load Balancer)
SLP (Service Location Protocol)
SNMP (Simple Network Management Protocol)
SNTP (Simple Network Time Protocol)
SP (Storage Processor)
SPN (service principal name)
SPOF (single point of failure)
SQL (Structured Query Language)
SRM (Storage Resource Management)
SSD (solid state disk
SSH (Secure Shell)
SSL (Secure Sockets Layer)
STP (shielded twisted pair)
SVC (switched virtual circuit)

T

TapeSAN (tape storage area network)

TCO (total cost of ownership)

TCP (Transmission Control Protocol)

TCP/IP (Transmission Control Protocol/Internet Protocol)

TOE (TCP offload engine)

TP (twisted pair)

TSM (Tivoli Storage Manager)

TTL (Time To Live)

U

UDP (User Datagram Protocol)

UI (user interface)

UID (user identification number)

Ultra ATA (Ultra Advanced Technology Attachment)

UNC (Uniform Naming Convention)

UPS (uninterruptible power supply)

URI (universal resource identifier)

URL (uniform resource locator)

USP (Universal Storage Platform)

UTC (Universal Coordinated Time)

UTP (unshielded twisted pair)

UUID (universal unique identifier)

UWN (unique world wide number)

V

VCI (virtual channel identifier)

VCMDB (Volume Configuration Management Database)

VDI (Virtual Device Interface)

VDisk (virtual disk)

VDS (Virtual Disk Service)

VFM (Virtual File Manager)

VFS (virtual file system)

VI (virtual interface)

vif (virtual interface)

VIRD (Virtual Router ID)

VLAN (virtual local area network)

VLD (virtual local disk)

VOD (video on demand)

VOIP (voice over IP)

VRML (Virtual Reality Modeling Language)

VTL (Virtual Tape Library)

W

WAFL (Write Anywhere File Layout)

WAN (wide area network)

WBEM (Web-Based Enterprise Management)

WHQL (Windows Hardware Quality Lab)

WINS (Windows Internet Name Service)

WORM (write once, read many)

WWN (worldwide name)

WWNN (worldwide node name)

WWPN (worldwide port name)

www (worldwide web)

Z
ZCS (zoned checksum)

Glossary

Storage terms

array LUN	The storage that third-party storage arrays provide to storage systems running Data ONTAP software. One array LUN is the equivalent of one disk on a native disk shelf.
LUN (logical unit number)	A logical unit of storage identified by a number.
ESH (Embedded Switching Hub) disk shelf module	A component that provides a means of managing an FC-AL loop in an intelligent manner, such that a single drive failure does not take down the loop. It also contains the enclosure services processor, which communicates the environmental data of the disk shelf.
ESH2 disk shelf module	A second-generation ESH module.
ESH4 disk shelf module	A third-generation ESH module.
multipath HA	In an active/active configuration, a configuration in which each controller has multiple ways to connect to a disk drive. Multipath HA cabling is the most resilient and preferred configuration for active/active configurations. This is because it takes full advantage of the resiliency capability of the disk shelves, which means that the node continues to have access to disk drives in the event of cable, HBA, or shelf module failure. A single failure of a cable, HBA, or module does not result in a controller failover.
native disk	A disk that is sold as local storage for storage systems that run Data ONTAP software.
native disk shelf	A disk shelf that is sold as local storage for storage systems that run Data ONTAP software.
SAS stack	Also referred to as *stack*. A group of one or more SAS disk shelves connected (daisy-chained) together and connected to the controller through the top disk shelf in the stack and the bottom disk shelf in the stack (as needed). The maximum number of disk shelves in a stack of disk shelves and the number of disk shelf stacks supported in a configuration are dependent on the type of storage system.
storage controller	The component of a storage system that runs the Data ONTAP operating system and controls its disk subsystem. Storage controllers are also sometimes called

controllers, *storage appliances*, *appliances*, *storage engines*, *heads*, *CPU modules*, or *controller modules*.

storage system The hardware device running Data ONTAP that receives data from and sends data to native disk shelves, third-party storage, or both. Storage systems that run Data ONTAP are sometimes referred to as *filers*, *appliances*, *storage appliances*, *gateways*, or *systems*.

Note: 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.

third-party storage The back-end storage arrays, such as IBM, Hitachi Data Systems, and HP, that provide storage for storage systems running Data ONTAP.

Cluster and high-availability terms

active/active configuration

- In the Data ONTAP 7.2 and 7.3 release families, a pair of storage systems or gateways (sometimes called *nodes*) configured to serve data for each other if one of the two systems stops functioning. Also sometimes referred to as *active/active pairs*.
- In the Data ONTAP 8.*x* release family, this functionality is referred to as a *high-availability (HA) configuration* or an *HA pair*.
- In the Data ONTAP 7.1 release family, this functionality is referred to as a *cluster*.

MetroCluster A type of HA pair that provides the capability to force a takeover when an entire node is destroyed or unavailable. Not all platforms, switches, storage subsystems, or Data ONTAP versions are supported in MetroCluster configurations.

Copyright and trademark information

This section includes copyright and trademark information, and important notices.

Copyright 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 (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S.A. patents, foreign patents, or pending applications.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

Trademark information

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. A complete and current list of other IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

NetApp, the NetApp logo, Network Appliance, the Network Appliance logo, Akorri, ApplianceWatch, ASUP, AutoSupport, BalancePoint, BalancePoint Predictor, Bycast, Campaign Express, ComplianceClock, Cryptainer, CryptoShred, Data ONTAP, DataFabric, DataFort, Decru, Decru DataFort, DenseStak, Engenio, Engenio logo, E-Stack, FAServer, FastStak, FilerView, FlexCache, FlexClone, FlexPod, FlexScale, FlexShare, FlexSuite, FlexVol, FPolicy, GetSuccessful, gFiler, Go further, faster, Imagine Virtually Anything,

Lifetime Key Management, LockVault, Manage ONTAP, MetroCluster, MultiStore, NearStore, NetCache, NOW (NetApp on the Web), Onaro, OnCommand, ONTAPI, OpenKey, PerformanceStak, RAID-DP, ReplicatorX, SANscreen, SANshare, SANtricity, SecureAdmin, SecureShare, Select, Service 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

NA 210-04531_A0, Printed in USA

GC27-2208-08