IBM System Storage N series

Clustered Data ONTAP 8.2 High-Availability Configuration Guide

SC27-6276-01

Contents

Preface

About this guide

This document applies to IBM N series systems running Data ONTAP, including systems with gateway functionality. If the terms *Cluster-Mode* or *clustered Data ONTAP* are used in this document, they refer to the Data ONTAP features and functionality designed for clusters, which are different from 7-Mode and prior Data ONTAP 7.1, 7.2, and 7.3 release families.

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

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

Supported features

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

Information about supported features can also be found on the N series support website (accessed and navigated as described in *Websites* on page 7).

Websites

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

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

content is organized and navigated, and to access the N series support website, refer to the following publicly accessible web page:

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

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

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

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

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

Getting information, help, and service

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

Before you call

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

- Check all cables to make sure they are connected.
- Check the power switches to make sure the system is turned on.
- Use the troubleshooting information in your system documentation and use the diagnostic tools that come with your system.
- Refer to the N series support website (accessed and navigated as described in *Websites* on page 7) for information on known problems and limitations.

Using the documentation

The latest versions of N series software documentation, including Data ONTAP and other software products, are available on the N series support website (accessed and navigated as described in *Websites* on page 7).

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

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

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

Hardware service and support

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

www.ibm.com/planetwide/

Firmware updates

IBM N series product firmware is embedded in Data ONTAP. As with all devices, ensure that you run the latest level of firmware. Any firmware updates are posted to the N series support website (accessed and navigated as described in *Websites* on page 7).

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

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

How to send your comments

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

Be sure to include the following:

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

Understanding HA pairs

HA pairs provide hardware redundancy that is required for nondisruptive operations and fault tolerance and give each node in the pair the software functionality to *take over* its partner's storage and subsequently *give back* the storage.

What an HA pair is

An HA pair is two storage systems (nodes) whose controllers are connected to each other directly. In this configuration, one node can take over its partner's storage to provide continued data service if the partner goes down.

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

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

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

How HA pairs support nondisruptive operations and fault tolerance

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

- Fault tolerance
  When one node fails or becomes impaired and a takeover occurs, the partner node continues to serve the failed node's data.
- Nondisruptive software upgrades or hardware maintenance
  During hardware maintenance or upgrades, when you halt one node and a takeover occurs (automatically, unless you specify otherwise), the partner node continues to serve data for the halted node while you upgrade or perform maintenance on the node you halted.
  During nondisruptive upgrades of Data ONTAP, the user manually enters the `storage failover takeover` command to take over the partner node to allow the software upgrade to occur. The takeover node continues to serve data for both nodes during this operation.
  For more information about nondisruptive software upgrades, see the *Clustered Data ONTAP Upgrade and Revert/Downgrade Guide*.

Nondisruptive aggregate ownership relocation can be performed without a takeover and giveback.

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

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

  The nodes use the interconnect to perform the following tasks:

  - Continually check if the other node is functioning
  - Mirror log data for each other's NVRAM or NVMEM
- The nodes use two or more disk shelf loops, or storage arrays, in which the following conditions apply:

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

  Note: Disk ownership is established by Data ONTAP or the administrator; it is not based on which disk shelf the disk is attached to.

  For more information about disk ownership, see the *Clustered Data ONTAP Physical Storage Management Guide*.

- They own their spare disks, spare array LUNs, or both, and do not share them with the other node.
- They each have mailbox disks or array LUNs on the root volume that perform the following tasks:

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

Related concepts

Nondisruptive operations with HA pairs on page 86
Where to find procedures for nondisruptive operations with HA pairs on page 86

How the HA pair improves fault tolerance

A storage system has a variety of single points of failure, such as certain cables or hardware components. An HA pair greatly reduces the number of single points of failure because if a failure occurs, the partner can take over and continue serving data for the affected system until the failure is fixed.

Single point of failure definition

A single point of failure represents the failure of a single hardware component that can lead to loss of data access or potential loss of data.

Single point of failure 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 such as a controller or adapter fails, you can use the controller failover function to provide continuous data availability and preserve data integrity for client applications and users.

Single point of failure analysis for HA pairs

Different individual hardware components and cables in the storage system are single points of failure, but an HA configuration can eliminate these points to improve data availability.

Hardware components	Single point of failure		How storage failover eliminates single point of failure
	Stand-alone	HA pair	
Controller	Yes	No	If a controller fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
NVRAM	Yes	No	If an NVRAM adapter fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
CPU fan	Yes	No	If the CPU fan fails, the node automatically fails over to its partner node. The partner (takeover) node serves data for both of the nodes.
Multiple NICs with interface groups (virtual interfaces)	Maybe, if all NICs fail	No	If one of the networking links within an interface group fails, the networking traffic is automatically sent over the remaining networking links on the same node. No failover is needed in this situation.

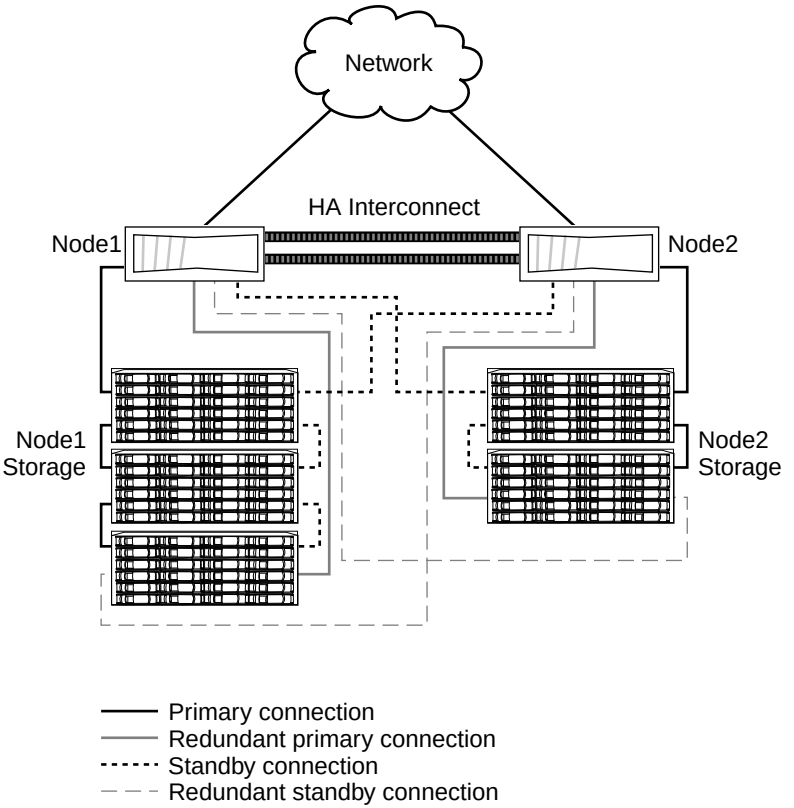
Hardware components	Single point of failure		How storage failover eliminates single point of failure
	Stand-alone	HA pair	
FC-AL adapter or SAS HBA	Yes	No	If an FC-AL adapter for the primary loop fails for a configuration without multipath HA, the partner node attempts a takeover at the time of failure. With multipath HA, no takeover is required. If the FC-AL adapter for the secondary loop fails for a configuration without multipath HA, the failover capability is disabled, but both nodes continue to serve data to their respective applications and users, with no impact or delay. With multipath HA, failover capability is not affected.
FC-AL or SAS cable (controller-to-shelf, shelf-to-shelf)	No, if dual-path cabling is used	No	If an FC-AL loop or SAS stack breaks in a configuration that does not have multipath HA, the break could lead to a failover, depending on the shelf type. The partnered nodes invoke the negotiated failover feature to determine which node is best for serving data, based on the disk shelf count. When multipath HA is used, no failover is required.
Disk shelf module	No, if dual-path cabling is used	No	If a disk shelf module fails in a configuration that does not have multipath HA, the failure could lead to a failover. The partnered nodes invoke the negotiated failover feature to determine which node is best for serving data, based on the disk shelf count. When multipath HA is used, there is no impact.
Disk drive	No	No	If a disk fails, the node can reconstruct data from the RAID4 parity disk. No failover is needed in this situation.
Power supply	Maybe, if both power supplies fail	No	Both the controller and disk shelf have dual power supplies. If one power supply fails, the second power supply automatically kicks in. No failover is needed in this situation. If both power supplies fail, the node automatically fails over to its partner node, which serves data for both nodes.

Hardware components	Single point of failure		How storage failover eliminates single point of failure
	Stand-alone	HA pair	
Fan (controller or disk shelf)	Maybe, if both fans fail	No	Both the controller and disk shelf have multiple fans. If one fan fails, the second fan automatically provides cooling. No failover is needed in this situation. If both fans fail, the node automatically fails over to its partner node, which serves data for both nodes.
HA interconnect adapter	Not applicable	No	If an HA interconnect adapter fails, the failover capability is disabled but both nodes continue to serve data to their respective applications and users.
HA interconnect cable	Not applicable	No	The HA interconnect adapter supports dual HA interconnect cables. If one cable fails, the heartbeat and NVRAM data are automatically sent over the second cable with no delay or interruption. If both cables fail, the failover capability is disabled but both nodes continue to serve data to their respective applications and users.

Connections and components of an HA pair

Each node in an HA pair requires a network connection, an HA interconnect between the controllers, and connections both to its own disk shelves as well as its partner node's shelves.

Standard HA pair



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

This diagram shows EXN3000 disk shelves. For more information about cabling SAS disk shelves, see the *Universal SAS and ACP Cabling Guide* on the N series support website (accessed and navigated as described in *Websites* on page 7).

How HA pairs relate to the cluster

HA pairs are components of the cluster, and both nodes in the HA pair are connected to other nodes in the cluster through the data and cluster networks. But only the nodes in the HA pair can takeover each other's storage.

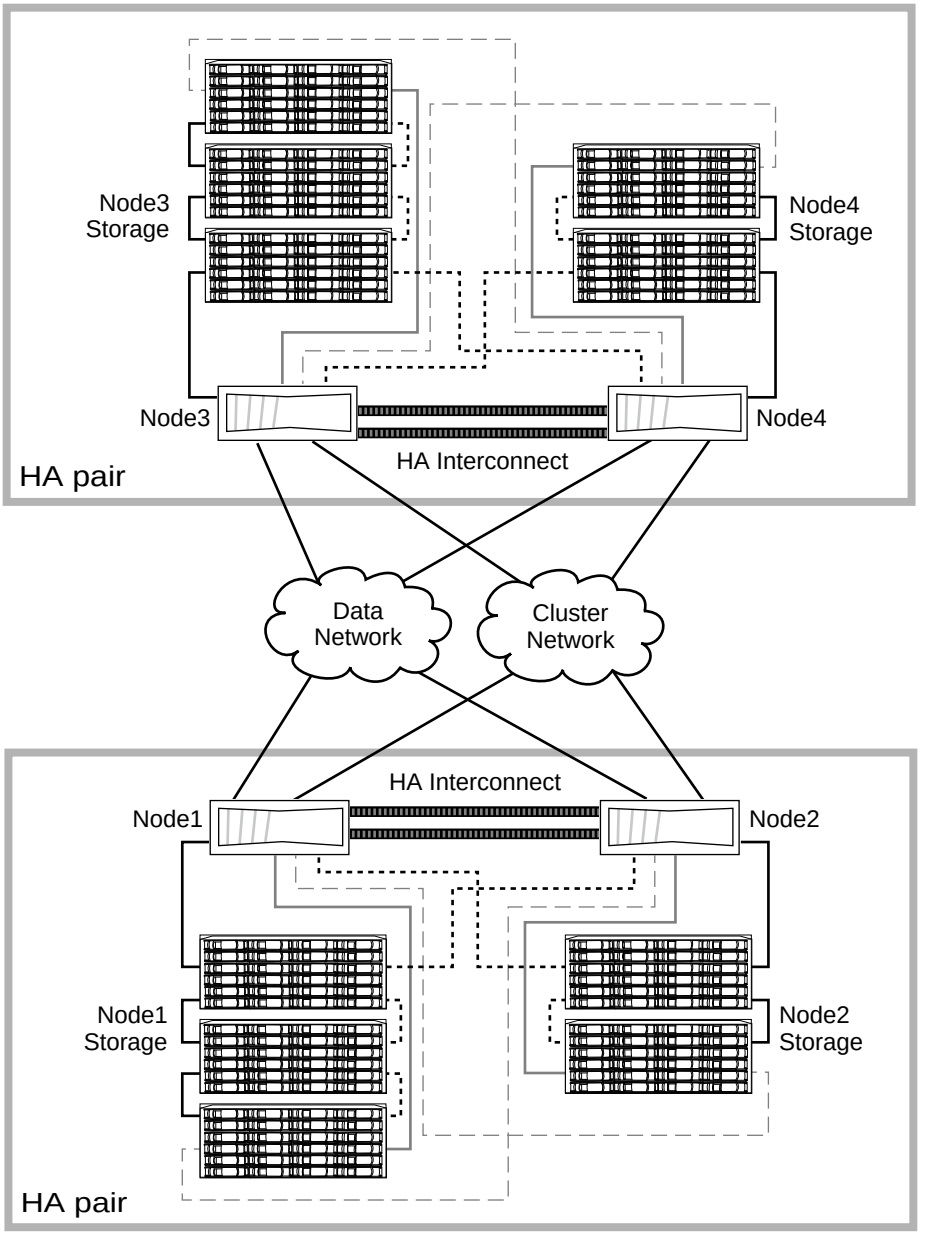
Although the controllers in an HA pair are connected to other controllers in the cluster through the cluster network, the HA interconnect and disk-shelf connections are found only between the node and its partner and their disk shelves or array LUNs.

The HA interconnect and each node's connections to the partner's storage provide physical support for high-availability functionality. The high-availability storage failover capability does not extend to other nodes in the cluster.

Note: Network failover does not rely on the HA interconnect and allows data network interfaces to failover to different nodes in the cluster outside the HA pair. Network failover is different than storage failover since it enables network resiliency across all nodes in the cluster.

Non-HA (or stand-alone) nodes are not supported in a cluster containing two or more nodes. Although single node clusters are supported, joining two separate single node clusters to create one cluster is not supported, unless you wipe clean one of the single node clusters and join it to the other to create a two-node cluster that consists of an HA pair. For information on single node clusters, see the *Clustered Data ONTAP System Administration Guide for Cluster Administrators*.

The following diagram shows two HA pairs. The multipath HA storage connections between the nodes and their storage are shown for each HA pair. For simplicity, only the primary connections to the data and cluster networks are shown.



Key to storage connections

- Primary connection
- Redundant primary connection
- Standby connection
- Redundant standby connection

Possible storage failover scenarios in this cluster are as follows:

- Node1 fails and Node2 takes over Node1's storage.
- Node2 fails and Node1 takes over Node2's storage.
- Node3 fails and Node4 takes over Node3's storage.
- Node4 fails and Node3 takes over Node4's storage.

If Node1 *and* Node2 both fail, the storage owned by Node1 and Node2 becomes unavailable to the data network. Although Node3 and Node4 are clustered with Node1 and Node2, they do not have direct connections to Node1 and Node2's storage and cannot take over their storage.

If you have a two-node switchless cluster

In a two-node switchless cluster configuration, you do not need to connect the nodes in the HA pair to cluster network switches. Instead, you install cluster network connections directly between the two storage controllers.

In a two-node switchless cluster, the two nodes can only be an HA pair. For cabling details, see the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) and the *Installation and Setup Instructions* for your system. The switchless cluster feature cannot be used with more than two nodes. If you plan to add more nodes, you must connect each node in the cluster to cluster network switches.

Related concepts

If your cluster consists of a single HA pair on page 32

Related tasks

Enabling cluster HA and switchless-cluster in a two-node cluster on page 49

Related references

Halting or rebooting a node without initiating takeover on page 71

Understanding takeover and giveback

Takeover and giveback are the operations that let you take advantage of the HA configuration to perform nondisruptive operations and avoid service interruptions. 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.

When takeovers occur

Takeovers can be initiated manually or occur automatically when a failover event happens, depending on how you configure the HA pair. In some cases, takeovers occur automatically, regardless of configuration.

Takeovers can occur under the following conditions:

- A takeover is manually initiated with the `storage failover takeover` command.
- A node is in an HA pair with the default configuration for immediate takeover on panic, and that node undergoes a software or system failure that leads to a panic.

  By default, the node automatically performs a giveback, returning the partner to normal operation after the partner has recovered from the panic and booted up.
- A node that is in an HA pair undergoes a system failure (for example, a loss of power) and cannot reboot.

  Note: If the storage for a node also loses power at the same time, a standard takeover is not possible.
- A node does not receive heartbeat messages from its partner.

  This could happen if the partner experienced a hardware or software failure that did not result in a panic but still prevented it from functioning correctly.
- You halt one of the nodes without using the `-f` or `-inhibit-takeover true` parameter.

  Note: In a two-node cluster with cluster HA enabled, halting or rebooting a node using the `-inhibit-takeover true` parameter will cause both nodes to stop serving data unless you first disable cluster HA and then assign epsilon to the node that you want to remain online.
- You reboot one of the nodes without using the `-inhibit-takeover true` parameter.

  The `-onreboot` parameter of the `storage failover` command is enabled by default.
- Hardware-assisted takeover is enabled and triggers a takeover when the remote management device (RLM or Service Processor) detects failure of the partner node.

Related references

Halting or rebooting a node without initiating takeover on page 71

Failover event cause-and-effect table

Failover events cause a controller failover in HA pairs. The storage system responds differently depending on the event and the type of HA pair.

Cause-and-effect table for HA pairs

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	HA pair
Single disk failure	No	No	Yes	Yes
Double disk failure (2 disks fail in same RAID group)	Yes, unless you are using 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, then yes	No, unless you are using RAID-DP, then yes
Triple disk failure (3 disks fail in same RAID group)	Yes	Maybe. If root volume has triple disk failure, no failover is possible	No	No
Single HBA (initiator) failure, Loop A	Maybe. If 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 is being used	Yes, if multipath HA is being used, or if failover succeeds
Single HBA (initiator) failure, Loop B	No	Yes, unless you are using multipath HA and the mailbox disks are not affected, then no	Yes, if multipath HA is being used	Yes, if multipath HA is being used, or if failover succeeds

Event	**Does the event trigger failover?**	**Does the event prevent a future failover from occurring, or a failover from occurring successfully?**	**Is data still available on the affected volume after the event?**	
			Single storage system	**HA pair**
Single HBA initiator failure (both loops at the same time)	Yes, unless multipath HA is in use, then no takeover needed	Maybe. If multipath HA is being used and the mailbox disks are not affected, then no; otherwise, yes	No, unless multipath HA is in use, then yes	No failover needed if multipath HA is in use
AT-FCX failure (Loop A)	Only if multidisk volume failure or open loop condition occurs, and multipath HA is not in use	Maybe. If root volume has double disk failure, no failover is possible	No	Yes, if failover succeeds
AT-FCX failure (Loop B)	No	Maybe. If multipath HA is in use, then no; otherwise, yes	Yes, if multipath HA is in use	Yes
IOM failure (Loop A)	Only if multidisk volume failure or open loop condition occurs, and multipath HA is not 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 multipath HA is in use, then no; otherwise, yes	Yes, if multipath HA is in use	Yes

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	HA pair
Shelf (backplane) failure	Only if multidisk volume failure or open loop condition occurs	Maybe. If root volume has double disk failure or if the mailboxes are affected, no failover is possible	No	No
Shelf, single power failure	No	No	Yes	Yes
Shelf, dual power failure	Only if multidisk volume failure or open loop condition occurs	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	No
Controller, single power failure	No	No	Yes	Yes
Controller, dual power failure	Yes	Yes, until power is restored	No	Yes, if failover succeeds
HA interconnect failure (1 port)	No	No	Not applicable	Yes
HA interconnect failure (both ports)	No	Yes	Not applicable	Yes

Event	Does the event trigger failover?	Does the event prevent a future failover from occurring, or a failover from occurring successfully?	Is data still available on the affected volume after the event?	
			Single storage system	HA pair
Tape interface failure	No	No	Yes	Yes
Heat exceeds permissible amount	Yes	No	No	No
Fan failures (disk shelves or controller)	No	No	Yes	Yes
Reboot	Yes	No	No	Yes, if failover occurs
Panic	Yes	No	No	Yes, if failover occurs

How hardware-assisted takeover speeds up takeover

Hardware-assisted takeover speeds up the takeover process by using a node's remote management device (SP or RLM) to detect failures and quickly initiate the takeover rather than waiting for Data ONTAP to recognize that the partner's heartbeat has stopped.

Without hardware-assisted takeover, if a failure occurs, the partner waits until it notices that the node is no longer giving a heartbeat, confirms the loss of heartbeat, and then initiates the takeover.

The hardware-assisted takeover feature uses the following process to take advantage of the remote management device and avoid that wait:

1. The remote management device monitors the local system for certain types of failures.
2. If a failure is detected, the remote management device immediately sends an alert to the partner node.
3. Upon receiving the alert, the partner initiates takeover.

Hardware-assisted takeover is enabled by default.

What happens during takeover

When a node takes over its partner, it continues to serve and update data in the partner's aggregates and volumes. To do this, the node takes ownership of the partner's aggregates, and the partner's LIFs migrate according to network interface failover rules. Except for specific SMB 3.0 connections, existing SMB (CIFS) sessions are disconnected when the takeover occurs.

The following steps occur when a node takes over its partner:

1. If the negotiated takeover is user-initiated, aggregate relocation is performed to move data aggregates one at a time from the partner node to the node that is performing the takeover.
   The current owner of each aggregate (except for the root aggregate) is changed from the target node to the node that is performing the takeover. There is a brief outage for each aggregate as ownership changes. This outage is briefer than an outage that occurs during a takeover without aggregate relocation.
   - You can monitor the progress using the `storage failover show-takeover` command.
   - The aggregate relocation can be avoided during this takeover instance by using the `-bypass-optimization` parameter with the `storage failover takeover` command. To bypass aggregate relocation during all future planned takeovers, set the `-bypass-takeover-optimization` parameter of the `storage failover modify` command to *`true`*.

     Note: Aggregates are relocated serially during planned takeover operations to reduce client outage. If aggregate relocation is bypassed, longer client outage occurs during planned takeover events.
2. If the user-initiated takeover is a negotiated takeover, the target node gracefully shuts down, followed by takeover of the target node's root aggregate and any aggregates which were not relocated in Step 1.
3. Before the storage takeover begins, data LIFs migrate from the target node to the node performing the takeover or to any other node in the cluster based on LIF failover rules.
   The LIF migration can be avoided by using the `-skip-lif-migration` parameter with the `storage failover takeover` command.
   For details on LIF configuration and operation, see the *Clustered Data ONTAP File Access Management Guide for CIFS*, the *Clustered Data ONTAP File Access Management Guide for NFS*, and the *Clustered Data ONTAP Network Management Guide*.
4. Existing SMB (CIFS) sessions are disconnected when takeover occurs.

   Note: Due to the nature of the SMB protocol, all SMB sessions except for SMB 3.0 sessions connected to shares with the *`Continuous Availability`* property set will be disruptive. SMB 1.0 and SMB 2.x sessions cannot reconnect after a takeover event. Therefore, takeover is disruptive and some data loss could occur.

5. SMB 3.0 sessions established to shares with the `Continuous Availability` property set can reconnect to the disconnected shares after a takeover event.
   If your site uses SMB 3.0 connections to Microsoft Hyper-V and the `Continuous Availability` property is set on the associated shares, takeover will be nondisruptive for those sessions.
   For more information about SMB configurations that support nondisruptive takeover, see the *Clustered Data ONTAP File Access Management Guide for CIFS*.

If the node doing the takeover panics

If the node that is performing the takeover panics within 60 seconds of initiating takeover, the following events occur:

- The node that panicked reboots.
- After it reboots, the node performs self-recovery operations and is no longer in takeover mode.
- Failover is disabled.
- If the node still owns some of the partner's aggregates, after enabling storage failover, return these aggregates to the partner using the `storage failover giveback` command.

Related information

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

What happens during giveback

The local node returns ownership of the aggregates and volumes to the partner node after any issues on the partner node are resolved or maintenance is complete. In addition, the local node returns ownership when the partner node has booted up and giveback is initiated either manually or automatically.

The following process takes place in a normal giveback. In this discussion, *Node A* has taken over *Node B*. Any issues on Node B have been resolved and it is ready to resume serving data.

1. Any issues on Node B have been resolved and it displays the following message:
   `Waiting for giveback`
2. The giveback is initiated by the `storage failover giveback` command or by automatic giveback if the system is configured for it.
   This initiates the process of returning ownership of Node B's aggregates and volumes from Node A back to Node B.
3. Node A returns control of the root aggregate first.
4. Node B proceeds to complete the process of booting up to its normal operating state.
5. As soon as Node B is at the point in the boot process where it can accept the non-root aggregates, Node A returns ownership of the other aggregates one at a time until giveback is complete.

You can monitor the progress of the giveback with the `storage failover show-giveback` command.

I/O resumes for each aggregate once giveback is complete for that aggregate, therefore reducing the overall outage window of each aggregate.

Background disk firmware update and takeover, giveback, and aggregate relocation

Background disk firmware updates affect HA pair takeover, giveback, and aggregate relocation operations differently, depending on how those operations are initiated.

The following list describes how background disk firmware update affects takeover, giveback, and aggregate relocation:

- If a background disk firmware update is occurring on a disk on either node, manually initiated takeover operations are delayed until the disk firmware update completes on that disk. If the background disk firmware update takes longer than 120 seconds, takeover operations are aborted and must be restarted manually after the disk firmware update completes. If the takeover was initiated with the `-bypass-optimization` parameter of the `storage failover takeover` command set to *`true`*, the background disk firmware update occurring on the destination node does not affect the takeover.
- If a background disk firmware update is occurring on a disk on the source (or takeover) node and the takeover was initiated manually with the `-options` parameter of the `storage failover takeover` command set to *`immediate`*, takeover operations are delayed until the disk firmware update completes on that disk.
- If a background disk firmware update is occurring on a disk on a node and it panics, takeover of the panicked node begins immediately.
- If a background disk firmware update is occurring on a disk on either node, giveback of data aggregates is delayed until the disk firmware update completes on that disk. If the background disk firmware update takes longer than 120 seconds, giveback operations are aborted and must be restarted manually after the disk firmware update completes.
- If a background disk firmware update is occurring on a disk on either node, aggregate relocation operations are delayed until the disk firmware update completes on that disk. If the background disk firmware update takes longer than 120 seconds, aggregate relocation operations are aborted and must be restarted manually after the disk firmware update completes. If aggregate relocation was initiated with the `-override-destination-checks` of the `storage aggregate relocation` command set to *`true`*, background disk firmware update occurring on the destination node does not affect aggregate relocation.

HA policy and giveback of the root aggregate and volume

Aggregates are automatically assigned an HA policy of CFO or SFO that determines how the aggregate and its volumes are given back.

Aggregates created on clustered Data ONTAP systems (*except* for the root aggregate containing the root volume) have an HA policy of SFO. During the giveback process, they are given back one at a time after the taken-over system boots.

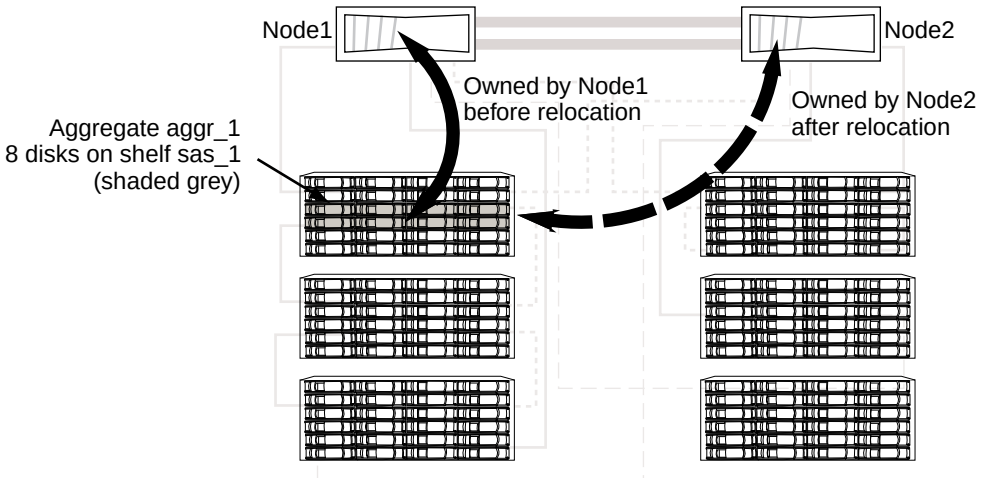
The root aggregate always has an HA policy of CFO and is given back at the start of the giveback operation. This is necessary to allow the taken-over system to boot. The other aggregates are given back one at a time after the taken-over node completes the boot process.

The HA policy of an aggregate cannot be changed from SFO to CFO in normal operation.

How aggregate relocation works

Aggregate relocation operations take advantage of the HA configuration to move the ownership of storage aggregates within the HA pair. Aggregate relocation occurs automatically during manually initiated takeovers to reduce downtime during planned failover events such as nondisruptive software upgrades, and can be initiated manually for load balancing, maintenance, and nondisruptive controller upgrades. Aggregate relocation cannot move ownership of the root aggregate.

The following illustration shows the relocation of the ownership of aggregate aggr_1 from Node1 to Node2 in the HA pair:



The aggregate relocation operation can relocate the ownership of one or more SFO aggregates if the destination node can support the number of volumes in the aggregates. There is only a short interruption of access to each aggregate. Ownership information is changed one by one for the aggregates.

During takeover, aggregate relocation happens automatically when the takeover is initiated manually. Before the target controller is taken over, ownership of the aggregates belonging to that controller are moved one at a time to the partner controller. When giveback is initiated, the ownership is automatically moved back to the original node. The `-bypass-optimization` parameter can be used with the `storage failover takeover` command to suppress aggregate relocation during the takeover.

The aggregate relocation requires additional steps if the aggregate is currently used by an Infinite Volume with SnapDiff enabled.

Aggregate relocation and Infinite Volumes with SnapDiff enabled

The aggregate relocation requires additional steps if the aggregate is currently used by an Infinite Volume with SnapDiff enabled. You must ensure that the destination node has a namespace mirror constituent, and make decisions about relocating aggregates that include namespace constituents.

For information about Infinite Volumes, see the *Clustered Data ONTAP Infinite Volumes Management Guide*.

Planning your HA pair configuration

As you plan your HA pair, you must consider recommended best practices, the requirements, and the possible variations.

Best practices for HA pairs

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

- Do not use the root aggregate for storing data.
- Do not create new volumes on a node when takeover, giveback, or aggregate relocation operations are in progress or pending. This includes the *partial giveback* state, which can occur when giveback operations are incomplete. If you create volumes on a node to which a giveback is being or will be performed, you might exceed the volume limit maximum for the node, which prevents bringing any additional volumes online.
- Make sure that each power supply unit in the storage system is on a different power grid so that a single power outage does not affect all power supply units.
- Use LIFs (logical interfaces) with defined failover policies to provide redundancy and improve availability of network communication.
- Follow the documented procedures in the *Clustered Data ONTAP Upgrade and Revert/Downgrade Guide* when upgrading your HA pair.
- Maintain consistent configuration between the two nodes.

  An inconsistent configuration is often the cause of failover problems.
- Make sure that each node has sufficient resources to adequately support the workload of both nodes during takeover mode.
- If your system supports remote management (through an RLM), make sure that you configure it properly, as described in the *Clustered Data ONTAP System Administration Guide for Cluster Administrators*.
- Follow recommended limits for FlexVol volumes, dense volumes, Snapshot copies, and LUNs to reduce the takeover or giveback time.

  When adding traditional or FlexVol volumes to an HA pair, consider testing the takeover and giveback times to ensure that they fall within your requirements.
- For systems using disks, check for and remove any failed disks, as described in the *Clustered Data ONTAP Physical Storage Management Guide*.
- Multipath HA is required on all HA pairs except for some N3150, N3220 and N3240 system configurations, which use single-path HA and lack the redundant standby connections.
- To ensure that you receive prompt notification if takeover capability becomes disabled, configure your system for automatic email notification for the `takeover impossible` EMS messages:
  - `ha.takeoverImpVersion`

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

- Avoid using the `-only-cfo-aggregates` parameter with the `storage failover giveback` command.

Setup requirements and restrictions for HA pairs

You must follow certain requirements and restrictions when setting up a new HA pair. These requirements help you ensure the data availability benefits of the HA pair design.

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

- Architecture compatibility

  Both nodes must have the same system model and be running the same Data ONTAP software and system firmware versions. See the *Clustered Data ONTAP Release Notes* for the list of supported systems.
- Nonvolatile memory (NVRAM or NVMEM) size and version compatibility

  The size and version of the system's nonvolatile memory must be identical on both nodes in an HA pair.
- Storage capacity

  The number of disks or array LUNs must not exceed the maximum configuration capacity. If your system uses both native disks and array LUNs, 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, array LUNs, or both, see the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7).

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

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

  - Two disks are required if the root volume is on a disk shelf.
  - One array LUN is required if the root volume is on a storage array.
- HA interconnect adapters and cables must be installed unless the system has two controllers in the chassis and an internal interconnect.
- Nodes must be attached to the same network and the Network Interface Cards (NICs) must be configured correctly.
- The same system software, such as Common Internet File System (CIFS) or Network File System (NFS), must be licensed and enabled on both nodes.

  Note: If a takeover occurs, the takeover node can provide only the functionality for the licenses installed on it. If the takeover node does not have a license that was being used by the partner node to serve data, your HA pair loses functionality after a takeover.
- For an HA pair using array LUNs, both nodes in the pair must be able to detect the same array LUNs.

  However, only the node that is the configured owner of a LUN has read-and-write access to the LUN. During takeover operations, the emulated storage system maintains read-and-write access to the LUN.

Requirements for hardware-assisted takeover

The hardware-assisted takeover feature is available on systems where the RLM or SP module is configured for remote management. Remote management provides remote platform management capabilities, including remote access, monitoring, troubleshooting, logging, and alerting features.

Although a system with remote management on both nodes provides hardware-assisted takeover for both, hardware-assisted takeover is also supported on HA pairs in which only one of the two systems has remote management configured. Remote management does not have to be configured on both nodes in the HA pair. Remote management can detect failures on the system in which it is installed and provide faster takeover times if a failure occurs on the system with remote management.

See the *Clustered Data ONTAP System Administration Guide for Cluster Administrators* for information about setting up remote management.

If your cluster consists of a single HA pair

Cluster high availability (HA) is activated automatically when you enable storage failover on clusters that consist of two nodes, and you should be aware that automatic giveback is enabled by default. On clusters that consist of more than two nodes, automatic giveback is disabled by default, and cluster HA is disabled automatically.

A cluster with only two nodes presents unique challenges in maintaining a quorum, the state in which a majority of nodes in the cluster have good connectivity. In a two-node cluster, neither node holds *epsilon*, the value that designates one of the nodes as the master. *Epsilon* is required in clusters with more than two nodes. Instead, both nodes are polled continuously to ensure that if takeover occurs,

the node that is still up and running has full read-write access to data as well as access to logical interfaces and management functions. This continuous polling function is referred to as *cluster high availability* or *cluster HA*.

Cluster HA is different than and separate from the high availability provided by HA pairs and the `storage failover` commands. While crucial to full functional operation of the cluster after a failover, cluster HA does not provide the failover capability of the storage failover functionality.

See the *Clustered Data ONTAP System Administration Guide for Cluster Administrators* for information about quorum and epsilon.

Related concepts

If you have a two-node switchless cluster on page 18

Related tasks

Enabling cluster HA and switchless-cluster in a two-node cluster on page 49
Configuring a node for non-HA (stand-alone) use on page 51

Related references

Halting or rebooting a node without initiating takeover on page 71

Possible storage configurations in the HA pairs

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

Symmetrical configurations	In a symmetrical HA pair, each node has the same amount of storage.
Asymmetrical configurations	In an asymmetrical standard HA pair, one node has more storage than the other. This is supported as long as neither node exceeds the maximum capacity limit for the node.
Active/passive configurations	In this configuration, the passive node has only a root volume, and the active node has all the remaining storage in addition to serving 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	In this configuration, shared loops or stacks between the nodes are particularly useful for active/passive configurations, as described in the preceding bullet.

HA pairs and storage system model types

Different model storage systems support some different HA configurations. This includes the physical configuration of the HA pair and the manner in which the system recognizes that it is in an HA pair.

Note: The physical configuration of the HA pair does not affect the cluster cabling of the nodes in the HA pair.

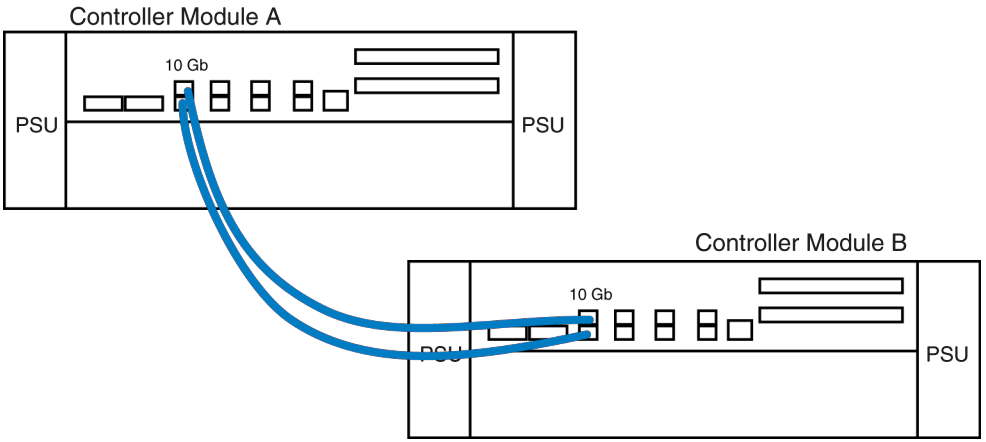
Single-chassis and dual-chassis HA pairs

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

The following example shows a single-chassis HA pair:

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

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



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

Interconnect cabling for systems with variable HA configurations

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

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

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

HA configuration and the HA state PROM value

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

Table of storage system models and HA configuration differences

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

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

Storage system model	HA configuration (single-chassis, dual-chassis, or either)	Interconnect type (internal InfiniBand, external InfiniBand, or external 10-Gb Ethernet)	Uses HA state PROM value?
N7950T	Single-chassis or dual-chassis	Dual-chassis: External InfiniBand using NVRAM adapter Single-chassis: Internal InfiniBand	Yes
N7550T	Single-chassis	Internal InfiniBand	Yes

Storage system model	HA configuration (single-chassis, dual-chassis, or either)	Interconnect type (internal InfiniBand, external InfiniBand, or external 10-Gb Ethernet)	Uses HA state PROM value?
N6250	Dual-chassis	External 10-Gb Ethernet using onboard ports c0a and c0b These ports are dedicated HA interconnect ports. Regardless of the system configuration, these ports cannot be used for data or other purposes.	Yes
N6220	Single-chassis or dual-chassis	Dual-chassis: External 10-Gb Ethernet using onboard ports c0a and c0b These ports are dedicated HA interconnect ports. Regardless of the system configuration, these ports cannot be used for data or other purposes. Single-chassis: Internal InfiniBand	Yes
N3150, N3220 and N3240	Single-chassis	Internal InfiniBand	Yes

Installing and cabling an HA pair

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

The specific procedure you use depends on whether you are using FC or SAS disk shelves.

Note: If your configuration includes SAS Storage Expansion Units, see the *Universal SAS and ACP Cabling Guide* on the N series support website (accessed and navigated as described in *Websites* on page 7) for information about cabling. For cabling the HA interconnect between the nodes, use the procedures in this guide.

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

System cabinet or equipment rack installation

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

HA pairs in an equipment rack

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

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

HA pairs in a system cabinet

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

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

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

If the HA pair you ordered has more than one cabinet, you must complete the cabling by cabling the local node to the partner node's disk shelves and the partner node to the local node's disk shelves.

You must also cable the nodes together by cabling the NVRAM HA interconnects. If the HA pair uses switches, you must install the switches, as described in the accompanying switch documentation. The system cabinets might also need to be connected to each other. See your *System Cabinet Guide* for information about connecting your system cabinets together.

Required documentation

Installation of an HA pair requires the correct documentation.

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

Manual name	Description
IBM System Storage N series Introduction and Planning Guide	This guide describes the physical requirements that your site must meet to install IBM N series equipment.
The appropriate system cabinet guide	This guide describes how to install IBM N series equipment into a system cabinet.
The appropriate disk shelf guide	These guides describe how to cable a disk shelf to a storage system.
The appropriate hardware documentation for your storage system model	These guides describe how to install the storage system, connect it to a network, and bring it up for the first time.
Diagnostics Guide	This guide describes the diagnostics tests that you can run on the storage system.
Clustered Data ONTAP Network Management Guide	This guide describes how to perform network configuration for the storage system.
Clustered Data ONTAP Upgrade and Revert/ Downgrade Guide	This guide describes how to upgrade storage system and disk firmware, and how to upgrade storage system software.
Clustered Data ONTAP System Administration Guide for Cluster Administrators	This guide describes general storage system administration, including tasks such as adding nodes to a cluster.
Clustered 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 HA pair with third-party storage, see the *Installation Requirements and Reference Guide* for information about cabling gateways to storage arrays. You

can also refer to the *Implementation Guide for Third-Party Storage* for information about configuring storage arrays to work with gateways.

Required tools

Installation of an HA pair requires the correct tools.

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

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

Required equipment

When you receive your HA pair, you should receive the equipment listed in the following table. See the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) to confirm your storage system type, storage capacity, and so on.

Required equipment	Details
Storage system	Two of the same type of storage systems
Storage	See the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7)
HA interconnect adapter (for controller modules that do not share a chassis)	InfiniBand (IB) HA adapter (The NVRAM adapter functions as the HA interconnect adapter on N5000 series and later storage systems)
For SAS disk shelves: SAS HBAs, if applicable For EXN1000 or EXN2000 unit disk shelves: FC-AL or FC HBA (FC HBA for disk) adapters, if applicable	Minimum of two SAS HBAs, two FC-AL adapters, or their equivalent in onboard ports
Fibre Channel switches, if applicable	Not applicable
SFP (small form-factor pluggable) modules, if applicable	Not applicable
NVRAM HA adapter media converter	Only if using fiber cabling

Required equipment	Details
Cables (provided with shipment unless otherwise noted)	• For systems using FC disk shelf connections, two optical controller-to-disk shelf cables per loop • For systems using SAS disk shelf connections, two SAS controller-to-disk shelf cables per stack • Multiple disk shelf-to-disk shelf cables, if applicable • For systems using the IB HA interconnect adapter, two 4xIB copper cables, two 4xIB optical cables, or two optical cables with media converters **Note:** You must purchase longer optical cables separately for cabling distances greater than 30 meters.

Preparing your equipment

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

Installing the nodes in equipment racks

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

Steps

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

Result

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

After you finish

Proceed to cable the HA pair.

Installing the nodes in a system cabinet

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

Steps

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

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

Result

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

After you finish

Proceed to cable the HA pair.

Cabling an HA pair

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

About this task

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

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

Note: If you are installing an HA pair between gateways using array LUNs, see the *Installation Requirements and Reference Guide* for information about cabling gateways to storage arrays. See the *Implementation Guide for Third-Party Storage* for information about configuring storage arrays to work with Data ONTAP.

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

Steps

1. *Determining which Fibre Channel ports to use for Fibre Channel disk shelf connections* on page 42

2. *Cabling Node A to EXN1000, EXN2000, or EXN4000 unit disk shelves* on page 43
3. *Cabling Node B to EXN1000, EXN2000, or EXN4000 unit disk shelves* on page 45
4. *Cabling the HA interconnect* on page 47

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

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

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

- Every disk shelf loop in the HA pair requires two ports on the node, one for the primary connection and one for the redundant multipath HA connection.

  A standard HA pair with one loop for each node uses four ports on each node.
- Onboard Fibre Channel ports should be used before using ports on expansion adapters.
- Always use the expansion slots in the order shown in the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) for your platform for an HA pair.
- If using Fibre Channel HBAs, insert the adapters in the same slots on both systems.

See the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) to obtain all slot assignment information for the various adapters you use to cable your HA pair.

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

Cabling guidelines for a quad-port Fibre Channel HBA

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

- Disk shelf loops using ESH4 modules must be cabled to the quad-port HBA first.
- Disk shelf loops using AT-FCX modules must be cabled to dual-port HBA ports or onboard ports before using ports on the quad-port HBA.
- Port A of the HBA must be cabled to the In port of Channel A of the first disk shelf in the loop.

  Port A of the partner node's HBA must be cabled to the In port of Channel B of the first disk shelf in the loop. This ensures that disk names are the same for both nodes.
- Additional disk shelf loops must be cabled sequentially with the HBA's ports.

  Port A is used for the first loop, port B for the second loop, and so on.
- If available, ports C or D must be used for the redundant multipath HA connection after cabling all remaining disk shelf loops.
- All other cabling rules described in the documentation for the HBA and the *IBM System Storage N series Introduction and Planning Guide* must be observed.

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

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

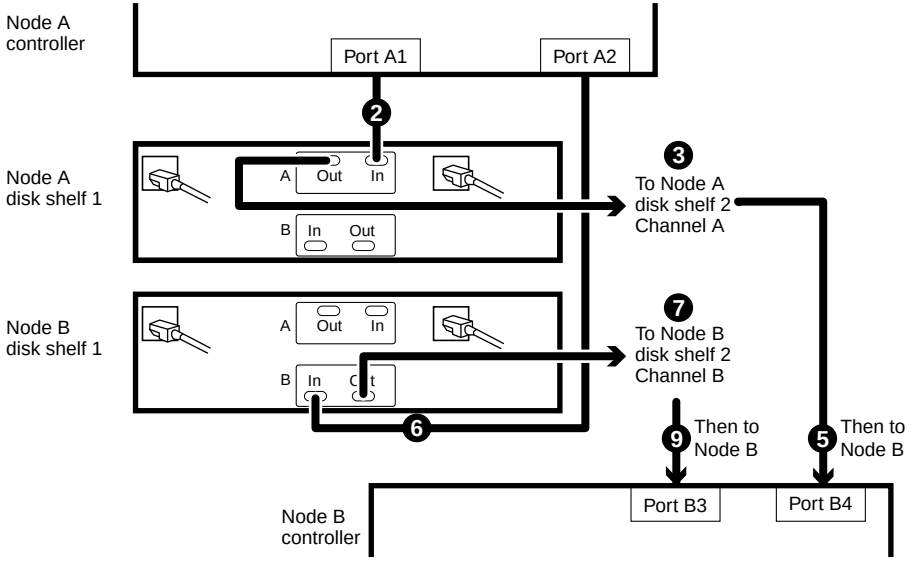
About this task

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

Note: You can find additional cabling diagrams in your system's *Installation and Setup Instructions* on the N series support website (accessed and navigated as described in *Websites* on page 7).

Steps

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



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

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

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

Result

Node A is completely cabled.

After you finish

Cable Node B.

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

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

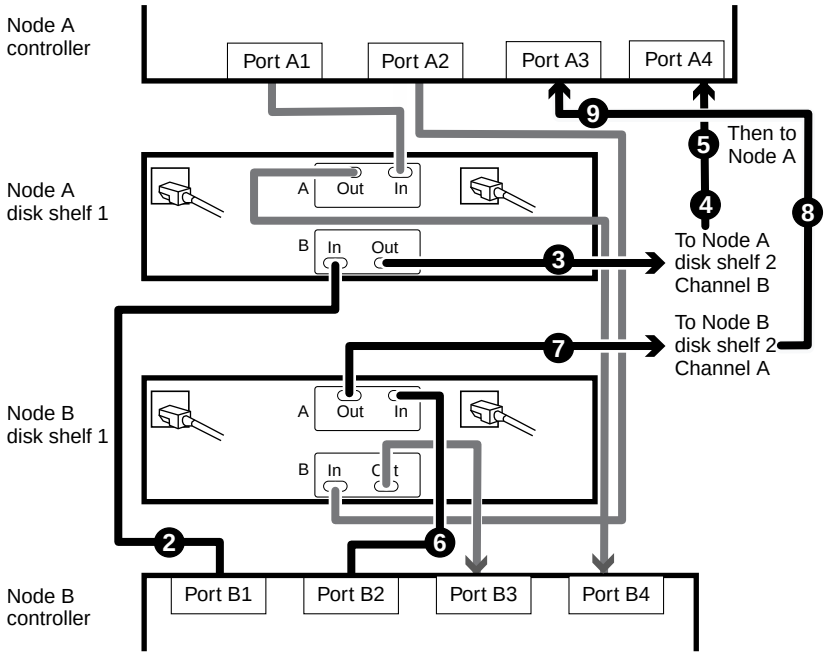
About this task

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

Note: You can find additional cabling diagrams in your system's *Installation and Setup Instructions* on the N series support website (accessed and navigated as described in *Websites* on page 7).

Steps

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



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

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

3. Cable the disk shelf Channel B Output port to the Channel B Input port of the next disk shelf in loop 1.

4. Repeat Step 3 for any remaining disk shelves in loop 1.

5. Cable the Channel B Output port of the last disk shelf in the loop to Fibre Channel port A4 of Node A.

   This provides the redundant multipath HA connection for Channel B.

6. Cable Fibre Channel port B2 of Node B to the Channel A Input port of the first disk shelf of Node B loop 1.

7. Cable the disk shelf Channel A Output port to the Channel A Input port of the next disk shelf in loop 1.

8. Repeat Step 7 for any remaining disk shelves in loop 1.

9. Cable the Channel A Output port of the last disk shelf in the loop to Fibre Channel port A3 of Node A.

   This provides the redundant multipath HA connection for Channel A.

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

Result

Node B is completely cabled.

After you finish

Cable the HA interconnect.

Cabling the HA interconnect

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

About this task

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

Steps

1. See the *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) to ensure that your interconnect adapter is in the correct slot for your system in an HA pair.

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

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

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

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

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

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

Result

The nodes are connected to each other.

After you finish

Configure the system.

Required connections for using uninterruptible power supplies with HA pairs

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

To gain the full benefit of the UPS, you must ensure that all the required equipment is connected to the UPS.

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

Configuring an HA pair

Bringing up and configuring an HA pair for the first time can require enabling HA mode capability and failover, setting options, configuring network connections, and testing the configuration.

These tasks apply to all HA pairs regardless of disk shelf type.

Steps

1. *Enabling cluster HA and switchless-cluster in a two-node cluster* on page 49
2. *Enabling the HA mode and storage failover* on page 50
3. *Configuring hardware-assisted takeover* on page 53
4. *Configuring automatic takeover* on page 54
5. *Configuring automatic giveback* on page 56
6. *Testing takeover and giveback* on page 58

Enabling cluster HA and switchless-cluster in a two-node cluster

A cluster consisting of only two nodes requires special configuration settings. Cluster high availability (HA) differs from the HA provided by storage failover, and is required in a cluster if it contains only two nodes. Also, if you have a switchless configuration, the switchless-cluster option must be enabled.

About this task

In a two-node cluster, cluster HA ensures that the failure of one node does not disable the cluster. If your cluster contains only two nodes:

- Enabling cluster HA requires and automatically enables storage failover and auto-giveback.
- Cluster HA is enabled automatically when you enable storage failover.

Note: If the cluster contains or grows to more than two nodes, cluster HA is not required and is disabled automatically.

A two-node cluster can be configured using direct-cable connections between the nodes instead of a cluster interconnect switch. If you have a two-node switchless configuration, the `switchless-cluster` network option must be enabled to ensure proper cluster communication between the nodes.

Steps

1. Enter the following command to enable cluster HA:

```
cluster ha modify -configured true
```

If storage failover is not already enabled, you are prompted to confirm enabling of both storage failover and auto-giveback.

2. If you have a two-node switchless cluster, enter the following commands to verify that the switchless-cluster option is set:

   a) Enter the following command to change to the advanced privilege level:

   ```
   set -privilege advanced
   ```

   Confirm when prompted to continue into advanced mode. The advanced mode prompt appears (*>).

   b) Enter the following command:

   ```
   network options switchless-cluster show
   ```

   If the output shows that the value is false, you must issue the following command:

   ```
   network options switchless-cluster modify true
   ```

   c) Enter the following command to return to the admin privilege level:

   ```
   set -privilege admin
   ```

Related concepts

How HA pairs relate to the cluster on page 16
If your cluster consists of a single HA pair on page 32
If you have a two-node switchless cluster on page 18

Related references

Halting or rebooting a node without initiating takeover on page 71

Enabling the HA mode and storage failover

You need to enable the HA mode and storage failover functionality to get the benefits of an HA pair.

Commands for enabling and disabling storage failover

There are specific Data ONTAP commands for enabling the storage failover functionality.

If you want to...	Use this command...
Enable takeover	`storage failover modify -enabled true -node` *nodename*
Disable takeover	`storage failover modify -enabled false -node` *nodename*

See the man page for each command for more information.

Commands for setting the HA mode

The HA license is no longer required in Data ONTAP 8.2, yet there are specific Data ONTAP commands for setting the HA mode. The system must be physically configured for HA before HA mode is selected. A reboot is required to implement the mode change.

If you want to...	Use this command...
Set the mode to HA	`storage failover modify -mode ha -node` *nodename*
Set the mode to non-HA	`storage failover modify -mode non_ha -node` *nodename* **Note:** You must disable storage failover before disabling ha_mode.

See the man page for each command for more information.

Related references

Connections and components of an HA pair on page 15
Description of node states displayed by storage failover show-type commands on page 60

Configuring a node for non-HA (stand-alone) use

By default, storage controllers are configured for use in HA mode. To use a controller as a single node cluster, you must disable the controller failover functionality and change the node to non-HA mode.

Before you begin

You must determine the current configuration of the storage controller because the controller failover and HA mode states can vary. You can use the `storage failover show` command to determine the current configuration.

You must also confirm that all loops and disk shelves in the system contain disks that belong to only one of the two nodes you intend to isolate. If any disk shelves or loops contain a mix of disks belonging to both nodes, you must move data.

About this task

When a storage controller is shipped from the factory or when Data ONTAP is reinstalled using option four of the Data ONTAP boot menu (`Clean configuration and initialize all disks`), HA mode is enabled by default, and the system's nonvolatile memory (NVRAM or NVMEM) is split. If you plan to use the controller as a single node cluster, you must configure the node as non-HA. Reconfiguring the node as non-HA mode enables full use of the system's nonvolatile memory.

Note: Configuring the node as a single node cluster removes the availability benefits of the HA configuration and creates a single point of failure.

For information on single node clusters, see the *Clustered Data ONTAP System Administration Guide for Cluster Administrators*.

For information on using the boot menu to manage the storage system, see the *Clustered Data ONTAP System Administration Guide for Cluster Administrators*.

Choices

- If the `storage failover show` output displays `Non-HA mode` in the `State Description` column, then the node is configured for non-HA mode and you are finished:

  Example

```
cluster01::> storage failover show
                                Takeover
Node           Partner          Possible State Description
-------------- -------------- -------- -----------------
node1          -              false    Non-HA mode
```

- If the `storage failover show` output directs you to reboot, you must reboot the node to enable full use of the system's nonvolatile memory:

  Example

```
cluster01::> storage failover show
                                Takeover
Node           Partner          Possible State Description
-------------- -------------- -------- -----------------
node1          -              false    Non-HA mode, reboot to use
full NVRAM
```

  a) Reboot the node using the following command:

  `cluster01::> reboot –node` *`nodename`*

  After the node reboots, you are finished.

- If the `storage failover show` output does not display `Non-HA mode` in the `State Description` column, you must disable both storage failover and HA mode and then reboot the node to enable full use of the system's nonvolatile memory:

  Example

```
cluster01::> storage failover show
                                Takeover
Node           Partner          Possible State Description
```

```
-------------- -------------- -------- -----------------
node1          -              true     Connected to partner_name
```

a) If you have a two-node cluster, disable cluster HA by using the following command:

cluster ha modify -configured false

b) Disable storage failover using the following command:

cluster01::> storage failover modify -enabled false -node *nodename*

c) Set the mode to non-HA by using the following command:

cluster01::> storage failover modify -mode non_ha -node *nodename*

d) Reboot the node by using the following command:

cluster01::> reboot –node *nodename*

After the node reboots, you are finished.

After you finish

Reconfigure the hardware.

Related concepts

If your cluster consists of a single HA pair on page 32
If you have a two-node switchless cluster on page 18

Related tasks

Enabling cluster HA and switchless-cluster in a two-node cluster on page 49

Configuring hardware-assisted takeover

You can configure hardware-assisted takeover to speed up takeover times. Hardware-assisted takeover uses the remote management device to quickly communicate local status changes to the partner node.

Commands for configuring hardware-assisted takeover

There are specific Data ONTAP commands for configuring the hardware-assisted takeover feature.

If you want to...	Use this command...
Disable or enable hardware-assisted takeover	`storage failover modify hwassist`
Set the partner address	`storage failover modify hwassist-partner-ip`
Set the partner port	`storage failover modify hwassist-partner-port`

If you want to...	Use this command...
Specify the interval between heartbeats	`storage failover modify hwassist-health-check-interval`
Specify the number of times the hardware-assisted takeover alerts are sent	`storage failover modify hwassist-retry-count`

See the man page for each command for more information. For a mapping of the `cf` options and commands used in Data ONTAP operating in 7-Mode to the `storage failover` commands, refer to the *Data ONTAP 7-Mode to Clustered Data ONTAP Command Map*. When in clustered Data ONTAP, you should always use the `storage failover` commands rather than issuing an equivalent 7-Mode command via the nodeshell (using the `system node run` command).

Configuring automatic takeover

Automatic takeover is enabled by default. You can control when automatic takeovers occur by using specific commands.

Commands for controlling automatic takeover

There are specific Data ONTAP commands you can use to change the default behavior and control when automatic takeovers occur.

If you want takeover to occur automatically when the partner node...	Use this command...
Reboots	`storage failover modify -node` *`nodename`* `-onreboot true`
Panics	`storage failover modify -node` *`nodename`* `-onpanic true`

See the man page for each command for more information. For a mapping of the `cf` options and commands used in Data ONTAP operating in 7-Mode to the `storage failover` commands, refer to the *Data ONTAP 7-Mode to Clustered Data ONTAP Command Map*. When in clustered Data ONTAP, you should always use the `storage failover` commands rather than issuing an equivalent 7-Mode command via the nodeshell (using the `system node run` command).

System events that always result in an automatic takeover

Some events always lead to an automatic takeover if storage failover is enabled. These takeovers cannot be avoided through configuration.

The following system events cause an automatic and unavoidable takeover of the node:

- The node cannot send heartbeat messages to its partner due to events such as loss of power or watchdog reset.
- You halt the node without using the `-f` or `-inhibit-takeover` parameter.
- The node panics.

System events that trigger hardware-assisted takeover

Many events can be detected by the remote management device (either a Remote LAN Module or Service Processor) and generate alerts. The partner node might initiate takeover, depending on the type of alert received.

Alert	Takeover initiated upon receipt?	Description
`power_loss`	Yes	Power loss on the node. The remote management 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. The remote management detected a lack of response from the system CPU and reset the system.
`power_off_via_rlm`	Yes	The remote management was used to power off the system.
`power_cycle_via_rlm`	Yes	The remote management was used to cycle the system power off and on.
`reset_via_rlm`	Yes	The remote management was used to reset the system.
`abnormal_reboot`	No	Abnormal reboot of the node.
`loss_of_heartbeat`	No	Heartbeat message from the node was no longer received by the remote management device. **Note:** This alert does not refer to the heartbeat messages between the nodes in the HA pair; it refers to the heartbeat between the node and its local remote management device.
`periodic_message`	No	Periodic message sent during normal hardware-assisted takeover operation.
`test`	No	Test message sent to verify hardware-assisted takeover operation.

Configuring automatic giveback

You can configure automatic giveback so that when a node that has been taken over boots up to the Waiting for Giveback state, giveback automatically occurs.

How automatic giveback works

The automatic takeover and automatic giveback operations can work together to reduce and avoid client outages. They occur by default in the case of a panic or reboot, or if the cluster contains only a single HA pair. However, these operations require configuration for other cases.

With the default settings, if one node in the HA pair panics or reboots, the partner node automatically takes over and then automatically gives back storage when the node that suffered the panic or reboot eventually reboots. This returns the HA pair to a normal operating state.

The automatic giveback after panic or reboot occurs by default. You can set the system to always attempt an automatic giveback (for cases other than panic or reboot), although you should do so with caution:

- The automatic giveback causes a second unscheduled interruption (after the automatic takeover). Depending on your client configurations, you might want to initiate the giveback manually to plan when this second interruption occurs.
- The takeover might have been due to a hardware problem that can recur without additional diagnosis, leading to additional takeovers and givebacks.

Note: Automatic giveback is enabled by default if the cluster contains only a single HA pair Automatic giveback is disabled by default during nondisruptive Data ONTAP upgrades.

Before performing the automatic giveback (regardless of what triggered it), the partner node waits for a fixed amount of time as controlled by the -delay-seconds parameter of the `storage failover modify` command. The default delay is 300 seconds. By delaying the giveback, the process results in two brief outages:

1. One outage during the takeover operation
2. One outage during the giveback operation

This process avoids a single, prolonged outage that includes:

1. The time for the takeover operation
2. The time it takes for the taken-over node to boot up to the point at which it is ready for the giveback
3. The time for the giveback operation

If the automatic giveback fails for any of the non-root aggregates, the system automatically makes two additional attempts to complete the giveback.

Commands for configuring automatic giveback

There are specific Data ONTAP commands for enabling or disabling automatic giveback.

If you want to...	Use this command...
Enable automatic giveback so that giveback occurs as soon as the taken-over node boots, reaches the Waiting for Giveback state, and the Delay before Auto Giveback period has expired The default setting is `false`, except for two-node clusters, where the default setting is `true`.	`storage failover modify -node` *nodename* `-auto-giveback true`
Disable automatic giveback The default setting is `false`, except for two-node clusters, where the default setting is `true` **Note:** Setting this parameter to false does not disable automatic giveback after takeover on panic and takeover on reboot; automatic giveback after takeover on panic must be disabled by setting the `-auto-giveback-after-panic` parameter to false	`storage failover modify -node` *nodename* `-auto-giveback false`
Disable automatic giveback after takeover on panic (this setting is enabled by default)	`storage failover modify -node` *nodename* `-auto-giveback-after-panic false`
Delay automatic giveback for a specified number of seconds (default is 300) This option determines the minimum time that a node will remain in takeover before performing an automatic giveback.	`storage failover modify -node` *nodename* `-delay-seconds` *seconds*
Change the number of times the automatic giveback is attempted within ten minutes (default is three)	`storage failover modify -node` *nodename* `-attempts` *integer*
Change the time period (in minutes) used by the -attempts parameter (default is 10 minutes)	`storage failover modify -node` *nodename* `-attempts-time` *integer*

If you want to...	Use this command...
Override any vetos of the giveback **Note:** Some vetos cannot be overridden.	`storage failover modify -node` *nodename* `-auto-giveback-override-vetoes true`

See the man page for each command for more information.

For a mapping of the `cf` options and commands used in Data ONTAP operating in 7-Mode to the `storage failover` commands, refer to the *Data ONTAP 7-Mode to Clustered Data ONTAP Command Map*. When in clustered Data ONTAP, you should always use the `storage failover` commands rather than issuing an equivalent 7-Mode command via the nodeshell (using the `system node run` command).

Testing takeover and giveback

After you configure all aspects of your HA pair, you need to verify that it is operating as expected in maintaining uninterrupted access to both nodes' storage during takeover and giveback operations. Throughout the takeover process, the local (or takeover) node should continue serving the data normally provided by the partner node. During giveback, control and delivery of the partner's storage should return transparently to the partner node.

Steps

1. Check the cabling on the HA interconnect cables to make sure that they are secure.
2. Verify that you can create and retrieve files on both nodes for each licensed protocol.
3. Enter the following command:

   `storage failover takeover -ofnode` *`partner_node`*

   See the man page for command details.
4. Enter either of the following commands to confirm that takeover occurred:

   `storage failover show-takeover`

   `storage failover show`
5. Enter the following command to display all disks belonging to the partner node (Node2) that the takeover node (Node1) can detect:

   `storage disk show -disk node1:* -home node2 -ownership`

   You can use the wildcard (*) character to display all the disks visible from a node. The following command displays all disks belonging to Node2 that Node1 can detect:

```
cluster::> storage disk show -disk node1:* -home node2 -ownership

Disk     Aggregate Home  Owner DR Home Home ID    Owner ID   DR Home ID
Reserver
```

```
-------- --------- ----- ----- ------- ---------- ---------- ----------
----------
node1:0c.3
         -         node2 node2 -       4078312453 4078312453 -
4078312452
node1:0d.3
         -         node2 node2 -       4078312453 4078312453 -
4078312452
.
.
.
```

6. Enter the following command to confirm that the takeover node (node1) controls the partner node's (Node2) aggregates:

   `aggr show -fields home-id,home-name,is-home`

```
cluster::> aggr show -fields home-name,is-home
aggregate home-name is-home
--------- --------- ---------
aggr0_1   node1     true
aggr0_2   node2     false
aggr1_1   node1     true
aggr1_2   node2     false
4 entries were displayed.
```

   During takeover, the *`is-home`* value of the partner node's aggregates is false.

7. Give back the partner node's data service after it displays the `Waiting for giveback` message by entering the following command:

   `storage failover giveback -ofnode` ***`partner_node`***

8. Enter either of the following commands to observe the progress of the giveback operation:

   `storage failover show-giveback`

   `storage failover show`

9. Proceed depending on whether you saw the message that giveback was completed successfully:

If takeover and giveback...	Then...
Is completed successfully	Repeat Step 2 through Step 8 on the partner node.
Fails	Correct the takeover or giveback failure and then repeat this procedure.

Monitoring an HA pair

You can use a variety of commands to monitor the status of the HA pair. If a takeover occurs, you can also determine what caused the takeover.

Commands for monitoring an HA pair

There are specific Data ONTAP commands for monitoring the HA pair.

If you want to check...	Use this command...
Whether failover is enabled or has occurred, or reasons why failover is not currently possible	`storage failover show`
Whether hardware-assisted takeover is enabled	`storage failover hwassist show`
The history of hardware-assisted takeover events that have occurred	`storage failover hwassist stats show`
The progress of a takeover operation as the partner's aggregates are moved to the node doing the takeover	`storage failover show-takeover`
The progress of a giveback operation in returning aggregates to the partner node	`storage failover show-giveback`
Whether an aggregate is home during takeover or giveback operations	`aggr show -fields home-id,owner-id,home-name,owner-name,is-home`
The HA state of the components of an HA pair (on systems that use the HA state)	`ha-config show` **Note:** This is a Maintenance mode command.

See the man page for each command for more information.

Description of node states displayed by storage failover show-type commands

You can use the `storage failover show`, `storage failover show-takeover`, and `storage failover show-giveback` commands to check the status of the HA pair and to troubleshoot issues.

The following table shows the node states that the `storage failover show` command displays:

State	Meaning
`Connected to partner_name.`	The HA interconnect is active and can transmit data to the partner node.
`Connected to partner_name, Partial giveback.`	The HA interconnect is active and can transmit data to the partner node. The previous giveback to the partner node was a partial giveback, or is incomplete. **Note:** Do not create new volumes on a node when takeover, giveback, or aggregate relocation operations are in progress or pending. If volumes are created during the partial giveback state, the volume limit maximum may be exceeded for the node to which a giveback is being or will be performed. This prevents bringing any additional volumes online.
`Connected to partner_name, Takeover of partner_name is not possible due to reason(s): reason1, reason2,....`	The HA interconnect is active and can transmit data to the partner node, but takeover of the partner node is not possible. A detailed list of reasons explaining why takeover is not possible is provided in the section following this table.
`Connected to partner_name, Partial giveback, Takeover of partner_name is not possible due to reason(s): reason1, reason2,....`	The HA interconnect is active and can transmit data to the partner node, but takeover of the partner node is not possible. The previous giveback to the partner was a partial giveback.
`Connected to partner_name, Waiting for cluster applications to come online on the local node.`	The HA interconnect is active and can transmit data to the partner node and is waiting for cluster applications to come online. This waiting period can last several minutes.
`Waiting for partner_name, Takeover of partner_name is not possible due to reason(s): reason1, reason2,....`	The local node cannot exchange information with the partner node over the HA interconnect. Reasons for takeover not being possible are displayed under *reason1*, *reason2*,…

State	Meaning
`Waiting for partner_name, Partial giveback, Takeover of partner_name is not possible due to reason(s): reason1, reason2,....`	The local node cannot exchange information with the partner node over the HA interconnect. The previous giveback to the partner was a partial giveback. Reasons for takeover not being possible are displayed under *reason1*, *reason2*,...
`Pending shutdown.`	The local node is shutting down. Takeover and giveback operations are disabled.
`In takeover.`	The local node is in takeover state and automatic giveback is disabled.
`In takeover, Auto giveback will be initiated in number of seconds seconds.`	The local node is in takeover state and automatic giveback will begin in *number of seconds* seconds.
`In takeover, Auto giveback deferred.`	The local node is in takeover state and an automatic giveback attempt failed because the partner node was not in waiting for giveback state.
`Giveback in progress, module module name.`	The local node is in the process of giveback to the partner node. Module *module name* is being given back. • Run the `storage failover show-giveback` command for more information.
`Normal giveback not possible: partner missing file system disks.`	The partner node is missing some of its own file system disks.
`Normal giveback not possible: disk inventory not yet received.`	The partner node has not sent disk inventory information to the local node.
`Previous giveback failed in module module name.`	Giveback to the partner node by the local node failed due to an issue in *module name*. • Run the `storage failover show-giveback` command for more information.
`Previous giveback failed. Auto giveback disabled due to exceeding retry counts.`	Giveback to the partner node by the local node failed. Automatic giveback is disabled because of excessive retry attempts.

State	Meaning
`Takeover scheduled in seconds seconds.`	Takeover of the partner node by the local node is scheduled due to the partner node shutting down or an operator-initiated takeover from the local node. The takeover will be initiated within the specified number of seconds.
`Takeover in progress, module module name.`	The local node is in the process of taking over the partner node. Module *module name* is being taken over.
`Takeover in progress.`	The local node is in the process of taking over the partner node.
firmware-status.	The node is not reachable and the system is trying to determine its status from firmware updates to its partner. A detailed list of possible firmware statuses is provided after this table.
`Node unreachable.`	The node is unreachable and its firmware status cannot be determined.
`Takeover failed, reason: reason.`	Takeover of the partner node by the local node failed due to reason *reason*.
`Previous giveback failed in module: module name. Auto giveback disabled due to exceeding retry counts.`	Previously attempted giveback failed in module *module name*. Automatic giveback is disabled. • Run the `storage failover show-giveback` command for more information.
`Previous giveback failed in module: module name.`	Previously attempted giveback failed in module *module name*. Automatic giveback is not enabled by the user. • Run the `storage failover show-giveback` command for more information.
`Connected to partner_name, Giveback of one or more SFO aggregates failed.`	The HA interconnect is active and can transmit data to the partner node. Giveback of one or more SFO aggregates failed and the node is in partial giveback state.
`Waiting for partner_name, Partial giveback, Giveback of one or more SFO aggregates failed.`	The local node cannot exchange information with the partner node over the HA interconnect. Giveback of one or more SFO aggregates failed and the node is in partial giveback state.

State	Meaning
`Connected to partner_name, Giveback of SFO aggregates in progress.`	The HA interconnect is active and can transmit data to the partner node. Giveback of SFO aggregates is in progress. • Run the `storage failover show-giveback` command for more information.
`Waiting for partner_name, Giveback of SFO aggregates in progress.`	The local node cannot exchange information with the partner node over the HA interconnect. Giveback of SFO aggregates is in progress. • Run the `storage failover show-giveback` command for more information.
`Waiting for partner_name. Node owns aggregates belonging to another node in the cluster.`	The local node cannot exchange information with the partner node over the HA interconnect, and owns aggregates that belong to the partner node.
`Connected to partner_name, Giveback of partner spare disks pending.`	The HA interconnect is active and can transmit data to the partner node. Giveback of SFO aggregates to the partner is done, but partner spare disks are still owned by the local node. • Run the `storage failover show-giveback` command for more information.
`Connected to partner_name, Automatic takeover disabled.`	The HA interconnect is active and can transmit data to the partner node. Automatic takeover of the partner is disabled.
`Waiting for partner_name, Giveback of partner spare disks pending.`	The local node cannot exchange information with the partner node over the HA interconnect. Giveback of SFO aggregates to the partner is done, but partner spare disks are still owned by the local node. • Run the `storage failover show-giveback` command for more information.
`Waiting for partner_name. Waiting for partner lock synchronization.`	The local node cannot exchange information with the partner node over the HA interconnect, and is waiting for partner lock synchronization to occur.

State	Meaning
`Waiting for partner_name. Waiting for cluster applications to come online on the local node.`	The local node cannot exchange information with the partner node over the HA interconnect, and is waiting for cluster applications to come online.
`Takeover scheduled. target node relocating its SFO aggregates in preparation of takeover.`	Takeover processing has started. The target node is relocating ownership of its SFO aggregates in preparation for takeover.
`Takeover scheduled. target node has relocated its SFO aggregates in preparation of takeover.`	Takeover processing has started. The target node has relocated ownership of its SFO aggregates in preparation for takeover.
`Takeover scheduled. Waiting to disable background disk firmware updates on local node. A firmware update is in progress on the node.`	Takeover processing has started. The system is waiting for background disk firmware update operations on the local node to complete.
`Relocating SFO aggregates to taking over node in preparation of takeover.`	The local node is relocating ownership of its SFO aggregates to the taking-over node in preparation for takeover.
`Relocated SFO aggregates to taking over node. Waiting for taking over node to takeover.`	Relocation of ownership of SFO aggregates from the local node to the taking-over node has completed. The system is waiting for takeover by the taking-over node.
`Relocating SFO aggregates to partner_name. Waiting to disable background disk firmware updates on the local node. A firmware update is in progress on the node.`	Relocation of ownership of SFO aggregates from the local node to the taking-over node is in progress. The system is waiting for background disk firmware update operations on the local node to complete.
`Relocating SFO aggregates to partner_name. Waiting to disable background disk firmware updates on partner_name. A firmware update is in progress on the node.`	Relocation of ownership of SFO aggregates from the local node to the taking-over node is in progress. The system is waiting for background disk firmware update operations on the partner node to complete.

State	Meaning
`Connected to` *`partner_name`*`. Previous takeover attempt was aborted because` *`reason`*`. Local node owns some of partner's SFO aggregates.` `Reissue a takeover of the partner with the "-bypass-optimization" parameter set to true to takeover remaining aggregates, or issue a giveback of the partner to return the relocated aggregates.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt was aborted because of the reason displayed under *`reason`*. The local node owns some of its partner's SFO aggregates. • Either reissue a takeover of the partner node, setting the `-bypass-optimization` parameter to `true` to takeover the remaining SFO aggregates, or perform a giveback of the partner to return relocated aggregates.
`Connected to` *`partner_name`*`. Previous takeover attempt was aborted. Local node owns some of partner's SFO aggregates.` `Reissue a takeover of the partner with the "-bypass-optimization" parameter set to true to takeover remaining aggregates, or issue a giveback of the partner to return the relocated aggregates.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt was aborted. The local node owns some of its partner's SFO aggregates. • Either reissue a takeover of the partner node, setting the `-bypass-optimization` parameter to `true` to takeover the remaining SFO aggregates, or perform a giveback of the partner to return relocated aggregates.
`Waiting for` *`partner_name`*`. Previous takeover attempt was aborted because` *`reason`*`. Local node owns some of partner's SFO aggregates.` `Reissue a takeover of the partner with the "-bypass-optimization" parameter set to true to takeover remaining aggregates, or issue a giveback of the partner to return the relocated aggregates.`	The local node cannot exchange information with the partner node over the HA interconnect. The previous takeover attempt was aborted because of the reason displayed under *`reason`*. The local node owns some of its partner's SFO aggregates. • Either reissue a takeover of the partner node, setting the `-bypass-optimization` parameter to `true` to takeover the remaining SFO aggregates, or perform a giveback of the partner to return relocated aggregates.

State	Meaning
`Waiting for partner_name. Previous takeover attempt was aborted. Local node owns some of partner's SFO aggregates.` `Reissue a takeover of the partner with the "-bypass-optimization" parameter set to true to takeover remaining aggregates, or issue a giveback of the partner to return the relocated aggregates.`	The local node cannot exchange information with the partner node over the HA interconnect. The previous takeover attempt was aborted. The local node owns some of its partner's SFO aggregates. • Either reissue a takeover of the partner node, setting the `-bypass-optimization` parameter to `true` to takeover the remaining SFO aggregates, or perform a giveback of the partner to return relocated aggregates.
`Connected to partner_name. Previous takeover attempt was aborted because failed to disable background disk firmware update (BDFU) on local node.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt was aborted because the background disk firmware update on the local node was not disabled.
`Connected to partner_name. Previous takeover attempt was aborted because reason.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt was aborted because of the reason displayed under *reason*.
`Waiting for partner_name. Previous takeover attempt was aborted because reason.`	The local node cannot exchange information with the partner node over the HA interconnect. The previous takeover attempt was aborted because of the reason displayed under *reason*.
`Connected to partner_name. Previous takeover attempt by partner_name was aborted because reason.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt by the partner node was aborted because of the reason displayed under *reason*.
`Connected to partner_name. Previous takeover attempt by partner_name was aborted.`	The HA interconnect is active and can transmit data to the partner node. The previous takeover attempt by the partner node was aborted.
`Waiting for partner_name. Previous takeover attempt by partner_name was aborted because reason.`	The local node cannot exchange information with the partner node over the HA interconnect. The previous takeover attempt by the partner node was aborted because of the reason displayed under *reason*.

State	Meaning
`Previous giveback failed in module: module name. Auto giveback will be initiated in number of seconds seconds.`	The previous giveback attempt failed in module *module_name*. Auto giveback will be initiated in *number of seconds* seconds. • Run the `storage failover show-giveback` command for more information.
`Node owns partner's aggregates as part of the non-disruptive controller upgrade procedure.`	The node owns its partner's aggregates due to the non-disruptive controller upgrade procedure currently in progress.
`Connected to partner_name. Node owns aggregates belonging to another node in the cluster.`	The HA interconnect is active and can transmit data to the partner node. The node owns aggregates belonging to another node in the cluster.
`Connected to partner_name. Waiting for partner lock synchronization.`	The HA interconnect is active and can transmit data to the partner node. The system is waiting for partner lock synchronization to complete.
`Connected to partner_name. Waiting for cluster applications to come online on the local node.`	The HA interconnect is active and can transmit data to the partner node. The system is waiting for cluster applications to come online on the local node.
`Non-HA mode, reboot to use full NVRAM.`	Storage failover is not possible. The HA mode option is configured as *non_ha*. • You must reboot the node to use all of its NVRAM.
`Non-HA mode, remove HA interconnect card from HA slot to use full NVRAM.`	Storage failover is not possible. The HA mode option is configured as *non_ha*. • You must move the HA interconnect card from the HA slot to use all of the node's NVRAM.
`Non-HA mode, remove partner system to use full NVRAM.`	Storage failover is not possible. The HA mode option is configured as *non_ha*. • You must remove the partner controller from the chassis to use all of the node's NVRAM.

State	Meaning
`Non-HA mode. Reboot node to activate HA.`	Storage failover is not possible. • The node must be rebooted to enable HA capability.
`Non-HA mode. See documentation for procedure to activate HA.`	Storage failover is not possible. The HA mode option is configured as *non_ha*. • You must run the `storage failover modify -mode ha -node` *nodename* command on both nodes in the HA pair and then reboot the nodes to enable HA capability.

Possible reasons automatic takeover is not possible

If automatic takeover is not possible, the reasons are displayed in the `storage failover show` command output. The output has the following form:
`Takeover of` *partner_name* `is not possible due to reason(s):` *reason1*`,` *reason2*`, ...`

Possible values for *reason* are as follows:

- Automatic takeover is disabled
- Disk shelf is too hot
- Disk inventory not exchanged
- Failover partner node is booting
- Failover partner node is performing software revert
- Local node about to halt
- Local node has encountered errors while reading the storage failover partner's mailbox disks
- Local node is already in takeover state
- Local node is performing software revert
- Local node missing partner disks
- Low memory condition
- NVRAM log not synchronized
- Storage failover interconnect error
- Storage failover is disabled
- Storage failover is disabled on the partner node
- Storage failover is not initialized
- Storage failover mailbox disk state is invalid
- Storage failover mailbox disk state is uninitialized
- Storage failover mailbox version mismatch
- Takeover disabled by operator

- The size of NVRAM on each node of the SFO pair is different
- The version of software running on each node of the SFO pair is incompatible
- Partner node attempting to take over this node
- Partner node halted after disabling takeover
- Takeover disallowed due to unknown reason
- Waiting for partner node to recover

Possible firmware states

- Boot failed
- Booting
- Dumping core
- Dumping sparecore and ready to be taken-over
- Halted
- In power-on self test
- In takeover
- Initializing
- Operator completed
- Rebooting
- Takeover disabled
- Unknown
- Up
- Waiting
- Waiting for cluster applications to come online on the local node
- Waiting for giveback
- Waiting for operator input

Halting or rebooting a node without initiating takeover

You can prevent an automatic storage failover takeover when you halt or reboot a node. This ability enables specific maintenance and reconfiguration operations.

Commands for halting or rebooting a node without initiating takeover

Inhibiting automatic storage failover takeover when halting or rebooting a node requires specific commands. If you have a two-node cluster, you must perform additional steps to ensure continuity of service.

To prevent the partner from taking over when you...	Use this command...
Halt the node	`system node halt -node` *`node`* `-inhibit-takeover` **Note:** If you have a two-node cluster, this command will cause all data LIFs to go offline unless you first disable cluster HA and then assign epsilon to the node that you intend to keep online.
Reboot the node Including the `-inhibit-takeover` parameter overrides the `-takeover-on-reboot` setting of the partner node to prevent it from initiating takeover.	`system node reboot -node` *`node`* `-inhibit-takeover` **Note:** If you have a two-node cluster, this command will cause all data LIFs to go offline unless you first disable cluster HA and then assign epsilon to the node that you intend to keep online.
Reboot the node By default, a node automatically takes over for its partner if the partner reboots. You can change the `-onreboot` parameter of the `storage failover` command to change this behavior.	`storage failover modify -node` *`node`* `-onreboot false` **Note:** Takeover can still occur if the partner exceeds the user-configurable expected time to reboot even when the `-onreboot` parameter is set to *`false`*.

For more information, see the man page for each command.

Halting or rebooting a node without initiating takeover in a two-node cluster

In a two-node cluster, cluster HA ensures that the failure of one node does not disable the cluster. If you halt or reboot a node in a two-node cluster without takeover by using the `-inhibit-takeover true` parameter, both nodes will stop serving data unless you change specific configuration settings.

About this task

Before a node in a cluster configured for cluster HA is rebooted or halted using the `-inhibit-takeover true` parameter, you must first disable cluster HA and then assign epsilon to the node that you want to remain online.

Steps

1. Enter the following command to disable cluster HA:

   `cluster ha modify -configured false`

   Note that this operation does not disable storage failover.

2. Because disabling cluster HA automatically assigns epsilon to one of the two nodes, you must determine which node holds it, and if necessary, reassign it to the node that you wish to remain online.

   a) Enter the following command to change to the advanced privilege level:

   `set -privilege advanced`

   Confirm when prompted to continue into advanced mode. The advanced mode prompt appears (*>).

   b) Determine which node holds epsilon by using the following command:

   `cluster show`

   In the following example, *Node1* holds epsilon:

   ```
   cluster::*> cluster show
   Node                 Health  Eligibility   Epsilon
   -------------------- ------- ------------  ------------
   Node1                true    true          true
   Node2                true    true          false
   ```

   If the node you wish to halt or reboot does not hold epsilon, proceed to step 3.

   c) If the node you wish to halt or reboot holds epsilon, you must remove it from the node by using the following command:

   `cluster modify -node Node1 -epsilon false`

   At this point, neither node holds epsilon.

d) Assign epsilon to the node that you wish to remain online (in this example, *Node2*) by using the following command:

```
cluster modify -node Node2 -epsilon true
```

3. Halt or reboot and inhibit takeover of the node that does not hold epsilon (in this example, *Node2*) by using either of the following commands as appropriate:

```
system node halt -node Node2 -inhibit-takeover true
```

```
system node reboot -node Node2 -inhibit-takeover true
```

4. After the halted or rebooted node is back online, you must enable cluster HA by using the following command:

```
cluster ha modify -configured true
```

Enabling cluster HA automatically removes epsilon from both nodes.

5. Enter the following command to return to the admin privilege level:

```
set -privilege admin
```

About manual takeover

You can perform a takeover manually when maintenance is required on the partner, and in other similar situations. Depending on the state of the partner, the command you use to perform the takeover varies.

Commands for performing and monitoring a manual takeover

You can manually initiate the takeover of a node in an HA pair to perform maintenance on that node while it is 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:

If you want to...	Use this command...
Take over the partner node	`storage failover takeover`
Take over the partner node before the partner has time to close its storage resources gracefully **Note:** If possible, migrate the data LIFs to another node before specifying this option. • If this option is specified, migration of data LIFs from the partner is delayed significantly even if the `skip-lif-migration` option is not specified. • Similarly, if this option is specified, negotiated takeover optimization is bypassed even if the `bypass-optimization` option is set to *`false`*.	`storage failover takeover -option immediate`
Take over the partner node without migrating LIFs	`storage failover takeover -skip-lif-migration` *`true`*
Take over the partner node even if there is a disk mismatch	`storage failover takeover -allow-disk-inventory-mismatch`

If you want to...	Use this command...
Take over the partner node even if there is a Data ONTAP version mismatch **Note:** This option is only used during the nondisruptive Data ONTAP upgrade process.	`storage failover takeover -option allow-version-mismatch`
Take over the partner node without performing aggregate relocation	`storage failover takeover -bypass-optimization` *`true`*
Monitor the progress of the takeover as the partner's aggregates are moved to the node doing the takeover	`storage failover show-takeover`

See the man page for each command for more information. For a mapping of the `cf` options and commands used in Data ONTAP operating in 7-Mode to the `storage failover` commands, refer to the *Data ONTAP 7-Mode to Clustered Data ONTAP Command Map*. When in clustered Data ONTAP, you should always use the `storage failover` commands rather than issuing an equivalent 7-Mode command via the nodeshell (using the `system node run` command).

About 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 the failed drives in the taken-over system as described in the *Clustered Data ONTAP Physical Storage Management Guide*.

If giveback is interrupted

If the takeover node experiences a failure or a power outage during the giveback process, that process stops and the takeover node returns to takeover mode until the failure is repaired or the power is restored.

However, this depends upon the stage of giveback in which the failure occurred. If the node encountered failure or a power outage during partial giveback state (after it has given back the root aggregate), it will not return to takeover mode. Instead, the node returns to partial-giveback mode. If this occurs, complete the process by repeating the giveback operation.

If giveback is vetoed

If giveback is vetoed, you must check the EMS messages to determine the cause. Depending on the reason or reasons, you can decide whether you can safely override the vetoes.

The `storage failover show-giveback` command displays the giveback progress and shows which subsystem vetoed, if any. Soft vetoes can be overridden, whereas hard vetoes cannot be, even if forced. The following tables summarize the soft vetoes that should not be overridden, along with recommended workarounds.

Giveback of the root aggregate

These vetoes do not apply to aggregate relocation operations:

Vetoing subsystem module	Workaround
vfiler_low_level	Terminate the CIFS sessions causing the veto, or shutdown the CIFS application that established the open sessions. Overriding this veto may cause the application using CIFS to disconnect abruptly and lose data.

Vetoing subsystem module	Workaround
Disk Check	All failed or bypassed disks should be removed before attempting giveback. If disks are sanitizing, the user should wait until the operation completes. Overriding this veto may cause an outage caused by aggregates or volumes going offline due to reservation conflicts or inaccessible disks.

Giveback of SFO aggregates

Vetoing subsystem module	Workaround
Lock Manager	Gracefully shutdown the CIFS applications that have open files, or move those volumes to a different aggregate. Overriding this veto will result in loss of CIFS lock state, causing disruption and data loss.
Lock Manager NDO	Wait until the locks are mirrored. Overriding this veto will cause disruption to Microsoft Hyper-V virtual machines.
RAID	Check the EMS messages to determine the cause of the veto: • If the veto is due to nvfile, bring the offline volumes and aggregates online. • If disk add or disk ownership reassignment operations are in progress, wait until they complete. • If the veto is due to an aggregate name or UUID conflict, troubleshoot and resolve the issue. • If the veto is due to mirror resync, mirror verify, or offline disks, the veto can be overridden and the operation will be restarted after giveback.
Disk Inventory	Troubleshoot to identify and resolve the cause of the problem. The destination node may be unable to see disks belonging to an aggregate being migrated. Inaccessible disks can result in inaccessible aggregates or volumes.
SnapMirror	Troubleshoot to identify and resolve the cause of the problem. This veto is due to failure to send an appropriate message to SnapMirror, preventing SnapMirror from shutting down.

Related references

Description of node states displayed by storage failover show-type commands on page 60

Commands for performing a manual giveback

You can manually initiate a giveback on a node in an HA pair to return storage to the original owner after completing maintenance or resolving any issues that caused the takeover.

If you want to...	Use this command...
Give back storage to a partner node	`storage failover giveback -ofnode` *`nodename`*
Give back storage even if the partner is not in the waiting for giveback mode	`storage failover giveback -ofnode` *`nodename`* `-require-partner-waiting false` This option should be used only if a longer client outage is acceptable.
Give back storage even if processes are vetoing the giveback operation (force the giveback)	`storage failover giveback -ofnode` *`nodename`* `-override-vetoes true`
Give back only the CFO aggregates (the root aggregate)	`storage failover giveback -ofnode` *`nodename`* `-only-cfo-aggregates true`
Monitor the progress of giveback after you issue the giveback command	`storage failover show-giveback`

See the man page for each command for more information.

For a mapping of the `cf` options and commands used in Data ONTAP operating in 7-Mode to the `storage failover` commands, refer to the *Data ONTAP 7-Mode to Clustered Data ONTAP Command Map*. When in clustered Data ONTAP, you should always use the `storage failover` commands rather than issuing an equivalent 7-Mode command via the nodeshell (using the `system node run` command).

Managing EXN1000, EXN2000, or EXN4000 unit disk shelves in an HA pair

You must follow specific procedures to add disk shelves to an HA pair or to upgrade or replace disk shelf hardware in an HA pair.

If your configuration includes SAS disk shelves, see the following documents on the N series support website (accessed and navigated as described in *Websites* on page 7):

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

Adding EXN1000, EXN2000, or EXN4000 unit disk shelves to a multipath HA loop

To add supported EXN1000, EXN2000, or EXN4000 unit disk shelves to an HA pair configured for multipath HA, you need to add the new disk shelf to the end of a loop, ensuring that it is connected to the previous disk shelf and to the controller.

About this task

This procedure does not apply to SAS disk shelves.

Steps

1. Confirm that there are two paths to every disk by entering the following command:
   `storage disk show -port`

   Note: If two paths are not listed for every disk, this procedure could result in a data service outage. Before proceeding, address any issues so that all paths are redundant. If you do not have redundant paths to every disk, you can use the nondisruptive upgrade method (failover) to add your storage.

2. Install the new disk shelf in your cabinet or equipment rack, as described in the *EXN2000 and EXN4000 Hardware and Service Guide*.
3. Find the last disk shelf in the loop to which you want to add the new disk shelf.

   Note: The Channel A Output port of the last disk shelf in the loop is connected back to one of the controllers.

   Note: In Step 4 you disconnect the cable from the disk shelf. When you do this, the system displays messages about adapter resets and eventually indicates that the loop is down. These messages are normal within the context of this procedure. However, to avoid them, you can optionally disable the adapter prior to disconnecting the disk shelf.

If you choose to, disable the adapter attached to the Channel A Output port of the last disk shelf by entering the following command:

run node *nodename* fcadmin config -d *adapter*

adapter identifies the adapter by name. For example: 0a.

4. Disconnect the SFP and cable coming from the Channel A Output port of the last disk shelf.

   Note: Leave the other ends of the cable connected to the controller.

5. Using the correct cable for a shelf-to-shelf connection, connect the Channel A Output port of the last disk shelf to the Channel A Input port of the new disk shelf.
6. Connect the cable and SFP you removed in Step 4 to the Channel A Output port of the new disk shelf.
7. If you disabled the adapter in Step 3, reenable the adapter by entering the following command:

   run node *nodename* fcadmin config -e *adapter*

8. Repeat Step 4 through Step 7 for Channel B.

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

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

   storage disk show -port

   Two paths should be listed for every disk.

Upgrading or replacing modules in an HA pair

In an HA pair with redundant pathing, you can upgrade or replace disk shelf modules without interrupting access to storage.

About this task

These procedures are for EXN1000, EXN2000, or EXN4000 unit disk shelves.

If your configuration includes SAS disk shelves, refer to the following documents on the N series support website (accessed and navigated as described in *Websites* on page 7):

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

About the disk shelf modules

A disk shelf module (ESH4 or AT-FCX) in an EXN4000 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 EXN4000 or EXN1000 unit disk shelves support the ESH4 or AT-FCX modules.

There are two modules in the middle of the rear of the disk shelf, one for Channel A and one for Channel B.

Note: The Input and Output ports on module B on the EXN4000 unit are the reverse of module A.

Restrictions for changing module types

If you plan to change the type of any module in your HA pair, make sure that you understand the restrictions.

You cannot mix ESH4 modules in the same loop with AT-FCX modules.

Best practices for changing module types

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

- Whenever you remove a module from an HA pair, you need to know whether the path you will disrupt is redundant.
  If it is, you can remove the module without interfering with the storage system's ability to serve data. However, if that module provides the only path to any disk in your HA pair, you must take action to ensure that you do not incur system downtime.
- When you replace a module, make sure that the replacement module's termination switch is in the same position as the module it is replacing.

  Note: ESH4 modules are self-terminating; this guideline does not apply to ESH4 modules.
- If you replace a module with a different type of module, make sure that you also change the cables, if necessary.
  For more information about supported cable types, see the hardware documentation for your disk shelf.
- Always wait 30 seconds after inserting any module before reattaching any cables in that loop.

Testing the modules

You should test your disk shelf modules after replacing or upgrading them to ensure that they are configured correctly and operating.

Steps

1. Verify that all disk shelves are functioning properly by entering the following command:

```
run -node nodename environ shelf
```

2. Verify that there are no missing disks by entering the following command:

   ```
   run -node nodename aggr status -r
   ```

   Local disks displayed on the local node should be displayed as partner disks on the partner node, and vice-versa.

3. Verify that you can create and retrieve files on both nodes for each licensed protocol.

Determining path status for your HA pair

If you want to remove a module from your HA pair, you need to know whether the path you will disrupt is redundant. You can use the `storage disk show -port` command to indicate whether the disks have redundant paths.

About this task

If the disks have redundant paths, you can remove the module without interfering with the storage system's ability to serve data. However, if that module provides the only path to any of the disks in your HA pair, you must take action to ensure that you do not incur system downtime.

Step

1. Use the `storage disk show -port` command at your system console.

   This command displays the following information for every disk in the HA pair:

   - Primary port
   - Secondary port
   - Disk type
   - Disk shelf
   - Bay

Examples for configurations with and without redundant paths

The following example shows what the `storage disk show -port` command output might look like for a redundant-path HA pair consisting of filers:

```
Primary         Port Secondary       Port Type   Shelf Bay
--------------- ---- --------------- ---- ------ ----- ---
Clustr-1:0a.16  A    Clustr-1:0b.16  B    FCAL       1   0
Clustr-1:0a.17  A    Clustr-1:0b.17  B    FCAL       1   1
Clustr-1:0a.18  A    Clustr-1:0b.18  B    FCAL       1   2
Clustr-1:0a.19  A    Clustr-1:0b.19  B    FCAL       1   3
Clustr-1:0a.20  A    Clustr-1:0b.20  B    FCAL       1   4
Clustr-1:0a.21  A    Clustr-1:0b.21  B    FCAL       1   5
Clustr-1:0a.22  A    Clustr-1:0b.22  B    FCAL       1   6
Clustr-1:0a.23  A    Clustr-1:0b.23  B    FCAL       1   7
Clustr-1:0a.24  A    Clustr-1:0b.24  B    FCAL       1   8
```

```
Clustr-1:0a.25  A    Clustr-1:0b.25  B    FCAL       1   9
Clustr-1:0a.26  A    Clustr-1:0b.26  B    FCAL       1  10
Clustr-1:0a.27  A    Clustr-1:0b.27  B    FCAL       1  11
Clustr-1:0a.28  A    Clustr-1:0b.28  B    FCAL       1  12
Clustr-1:0a.29  A    Clustr-1:0b.29  B    FCAL       1  13
```

Notice that every disk (for example, 0a.16/0b.16) has two ports active: one for A and one for B. The presence of the redundant path means that you do not need to fail over one system before removing modules from the system.

Attention: Make sure that every disk has two paths. Even in an HA pair configured for redundant paths, a hardware or configuration problem can cause one or more disks to have only one path. If any disk in your HA pair has only one path, you must treat that loop as if it were in a single-path HA pair when removing modules.

The following example shows what the `storage disk show -port` command output might look like for an HA pair consisting of filers that do not use redundant paths:

```
Clustr::> storage disk show -port
Primary         Port Secondary       Port Type   Shelf Bay
--------------- ---- --------------- ---- ------ ----- ---
Clustr-1:0a.16  A    -               -    FCAL       1   0
Clustr-1:0a.17  A    -               -    FCAL       1   1
Clustr-1:0a.18  A    -               -    FCAL       1   2
Clustr-1:0a.19  A    -               -    FCAL       1   3
Clustr-1:0a.20  A    -               -    FCAL       1   4
Clustr-1:0a.21  A    -               -    FCAL       1   5
Clustr-1:0a.22  A    -               -    FCAL       1   6
Clustr-1:0a.23  A    -               -    FCAL       1   7
Clustr-1:0a.24  A    -               -    FCAL       1   8
Clustr-1:0a.25  A    -               -    FCAL       1   9
Clustr-1:0a.26  A    -               -    FCAL       1  10
Clustr-1:0a.27  A    -               -    FCAL       1  11
Clustr-1:0a.28  A    -               -    FCAL       1  12
Clustr-1:0a.29  A    -               -    FCAL       1  13
```

For this HA pair, there is only one path to each disk. This means that you cannot remove a module from the configuration, thereby disabling that path, without first performing a takeover.

Hot-swapping a module

You can hot-swap a faulty disk shelf module, removing the faulty module and replacing it without disrupting data availability.

About this task

When you hot-swap a disk shelf module, you must ensure that you never disable the only path to a disk; disabling that single path results in a system outage.

Attention: If there is newer firmware in the `/mroot/etc/shelf_fw` directory than that on the replacement module, the system automatically runs a firmware update. This firmware update causes a service interruption on non-multipath HA AT-FCX installations, multipath HA configurations running versions of Data ONTAP prior to 7.3.1, and systems with non-RoHS AT-FCX modules.

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. Complete the following steps if any disks use this loop as their only path to a controller:

   a) Follow the cables from the module you want to replace back to one of the nodes, called NodeA.
   b) Enter the following command at the NodeB console:

      ```
      storage failover takeover -ofnode NodeA
      ```

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

   Any module in the loop that is attached to NodeA can now be replaced.

4. Put on the antistatic wrist strap and grounding leash.
5. Disconnect the module that you are removing from the Fibre Channel cabling.
6. Using the thumb and index fingers of both hands, press the levers on the CAM mechanism on the module to release it and pull it out of the disk shelf.
7. Slide the replacement module into the slot at the rear of the disk shelf and push the levers of the cam mechanism into place.

   Attention: Do not use excessive force when sliding the module into the disk shelf; you might damage the connector.

Wait 30 seconds after inserting the module before proceeding to the next step.

8. Recable the disk shelf to its original location.
9. Check the operation of the new module by entering the following command from the console of the node that is still running:

   ```
   run -node nodename
   ```

   The node reports the status of the modified disk shelves.
10. Complete the following steps if you performed a takeover previously:
    a) Return control of NodeA's disk shelves by entering the following command at the console of the takeover node:

       ```
       storage failover giveback -ofnode NodeA
       ```

    b) Wait for the giveback to be completed before proceeding to the next step.
11. Test the replacement module.
12. Test the configuration.

Related concepts

Best practices for changing module types on page 81

Related tasks

Determining path status for your HA pair on page 82
Hot-removing disk shelves in systems running Data ONTAP 8.2.1 or later on page 86

Nondisruptive operations with HA pairs

By taking advantage of an HA pair's takeover and giveback operations, you can change hardware components and perform software upgrades in your configuration without disrupting access to the system's 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. Aggregate relocation extends the range of nondisruptive capabilities by enabling storage controller upgrade and replacement operations.

Where to find procedures for nondisruptive operations with HA pairs

An HA pair allows you to perform an extensive range of nondisruptive system maintenance and upgrade operations. You can refer to the specific documents for the required procedures.

The following table lists where you can find information on nondisruptive operations:

If you want to perform this task nondisruptively...	**See the...**
Upgrade Data ONTAP	*Clustered Data ONTAP Upgrade and Revert/ Downgrade Guide*
Replace a hardware FRU component	FRU procedures for your platform

Hot-removing disk shelves in systems running Data ONTAP 8.2.1 or later

If your system is running Data ONTAP 8.2.1 or later, you can hot-remove disk shelves—physically remove disk shelves that have had the aggregates and ownership removed from the disk drives—in a clustered ONTAP multipath HA configuration with EXN4000 and EXN1000 units that is up and serving data. You can hot-remove one or more disk shelves from anywhere within a loop of disk shelves or remove a loop of disk shelves.

Before you begin

- Your storage system must be running Data ONTAP 8.2.1 or later.
- Your storage system must be a multipath HA system.
- If you have a N3220 and N3240 configuration, the external storage must be cabled as multipath HA.

- You must have already removed all aggregates and ownership from the disk drives in the disk shelves you are removing.

  Attention: Do not attempt this procedure unless you have already removed all aggregates from the disk shelves you are removing. If you attempt this procedure with aggregates on the disk shelf you are removing, you could fail the system with a multi-disk panic.

  The *Clustered Data ONTAP Physical Storage Management Guide* has information about deleting aggregates in the "Commands for managing aggregates" section and a procedure for removing disk drive ownership in the "Removing ownership from a disk" section. This document is available on the N series support website (accessed and navigated as described in *Websites* on page 7).

  Note: The procedure for removing ownership from disk drives requires you to disable disk autoassignment. You reenable disk autoassignment when prompted at the end of this shelf hot-remove procedure.
- Multipath HA configurations cannot be in a takeover state.
- If you are removing one or more disk shelves from within a loop, you must have factored the distance to bypass the disk shelves you are removing; therefore, if the current cables are not long enough, you need to have longer cables available.

  The *IBM System Storage N series Introduction and Planning Guide* at the N series support website (accessed and navigated as described in *Websites* on page 7) contains information about supported cables.

About this task

- This procedure follows cabling best practices; therefore, references to modules and module input and output ports align with the best practices. If your storage system is cabled differently than what is prescribed as best practice, the modules and/or module input and output ports might be different.
- *Path A* refers to the A-side disk shelf module or module A located in the top location in the disk shelf.
- *Path B* refers to the B-side disk shelf module or module B located in the bottom location in the disk shelf.
- The *first* disk shelf in the loop is the disk shelf with the input ports directly connected to the controllers.
- The *interim* disk shelf in the loop is the disk shelf directly connected to other disk shelves in the loop.
- The *last* disk shelf in the loop is the disk shelf with output ports directly connected to the controllers.
- The *next* disk shelf is the disk shelf downstream of the disk shelf being removed, in depth order.
- The *previous* disk shelf is the disk shelf upstream of the disk shelf being removed, in depth order.

Steps

1. Verify your system configuration is `Multi-Path HA` by entering the following command from the nodeshell of either controller:

   `sysconfig`

   It might take up to a minute for the system to complete discovery.

   The configuration is listed in the **System Storage Configuration** field.

   Note: If you have a N3220 and N3240 system with external storage, the output shows as `Mixed-Path HA` because the internal storage is cabled as single-path HA and the external storage is cabled as multipath HA.

   Attention: If your non N3220 and N3240 system is shown as something other than `Multi-Path HA`, you cannot continue with this procedure.

2. Verify that the disk drives in the disk shelves you are removing have no aggregates (are spares) and ownership is removed, by completing the following substeps:

   a) Enter the following command from the clustershell of either controller:

      `storage disk show -shelf` ***`shelf_number`***

   b) Check the output to verify there are no aggregates on the disk drives in the disk shelves you are removing.

      Disk drives with no aggregates have a dash in the **Aggregate** column.

      Attention: If disk drives in the disk shelves you are removing have aggregates, you cannot continue with this procedure.

   c) Check the output to verify ownership is removed from the disk drives on the disk shelves you are removing or that the disk drives are failed.

If the output shows...	Then...
`unassigned` or `broken` for all disk drives	Go to the next step. The disk drives in the disk shelves you are removing are in the correct state to continue with this procedure.
Any disk drives in the disk shelves you are removing have ownership	You must remove ownership from the disk drives. You can use the “Removing ownership from a disk” procedure referenced in the “Before you begin” section of this procedure. **Attention:** You cannot continue with this procedure until the disk drives in the disk shelves you are removing are failed or ownership is removed.

Example

The following output for the **`storage disk show -shelf 3`** command shows disk drives on the disk shelf being removed (disk shelf 3). All of the disk drives in disk shelf 3 have a dash in the **Aggregate** column. Two disk drives have the ownership removed; therefore, `unassigned` appears in the **Container Type** column. And two disk drives are failed; therefore, `broken` appears in the **Container Type** column:

```
fas6200::> storage disk show -shelf 3

                  Usable                Container
Disk              Size       Shelf Bay  Type          Position
Aggregate Owner
----------------- ---------- ----- ---  -----------   ----------
--------- --------
...
ie6200-1:0a.17         -     3    6    unassigned    present
-         -
ie6200-1:0a.19         -     3    8    unassigned    present
-         -
ie6200-1:0a.21         -     3    5    broken        present
-         -
ie6200-1:0a.22         -     3    4    broken        data
-         -
...
```

3. Turn on the LEDs for each disk drive in the disk shelves you are removing so the disk shelves are physically identifiable by completing the following substeps:

   You need to be certain of which disk shelves you are removing so that you can correctly recable path A and path B later in this procedure.

   You enter the commands from the nodeshell of either controller.

   a) Identify the disk drives in each disk shelf you are removing by entering the following command:

      `fcadmin device_map`

      Example

      In this output, the shelf mapping shows three disk shelves in a loop and their respective 14 disk drives. If disk shelf 3 is being removed, disk drives 45 44 43 42 41 40 39 38 37 36 35 34 33 32 are applicable.

```
fas6200> fcadmin device_map

Loop Map for channel 0c:
...
Shelf mapping:
Shelf 3:  45  44  43  42  41  40  39  38  37  36  35  34  33  32
Shelf 4:  77  76  75  74  73  72  71  70  69  68  67  66  65  64
```

```
Shelf 5:  93  92  91  90  89  88  87  86  85  84  83  82  81  80
...
```

b) Turn on the LEDs for the disk drives you identified in Substep a, by entering the following command:

led_on *disk_name*

You must be in advanced privilege level to enter this command.

The fault LED on the front of the disk drive illuminates—solid. Additionally, if you have any failed disk drives in the disk shelves you are removing, the activity LED on the front of those disk drives blinks.

It is recommended that you turn on the LED for a minimum of four disk drives so that the disk shelves you are removing can be visually identified. You must repeat the command for each disk drive.

Example

To turn on the fault LED for disk drive 0c.45 in disk shelf 3 identified in Substep a, you enter **led_on *0c.45***

4. If you are removing an entire loop of disk shelves, complete the following substeps; otherwise, go to the next step.

a) Remove all cables on path A and path B.

This includes controller-to-shelf cables and shelf-to-shelf cables for all disk shelves in the loop you are removing.

b) Go to Step 8.

5. If you are removing one or more disk shelves from a loop (but keeping the loop), recable the applicable path A loop connections to bypass the disk shelves you are removing by completing the applicable set of substeps:

If you are removing more than one disk shelf, complete the applicable set of substeps one disk shelf at a time.

If you need a graphical system cabling reference, use the platform specific Installation and Setup Instructions document that ships with each platform, or access these documents on the IBM N series Support Site at *www.ibm.com/storage/support/nseries/*. Once at this site, you can search on specific platform. For example, to find the Installation and Setup Instructions document for N6220 systems, search on “N6220 series”.

If you are removing...	Then...
The first disk shelf in a loop	**a.** Remove the cable connecting the module A output port of the first disk shelf and the module A input port of the second disk shelf in the loop and set it aside. **b.** Move the cable connecting the controller to the module A input port of the first disk shelf to the module A input port of the second disk shelf in the loop
An interim disk shelf in a loop	**a.** Remove the cable connecting the module A output port of the disk shelf being removed and the module A input port of the next disk shelf in the loop and set it aside. **b.** Move the cable connecting the module A input port of the disk shelf being removed to the module A input port of the next disk shelf in the loop
The last disk shelf in a loop	**a.** Remove the cable connecting the module A input port of the last disk shelf and the module A output port of the previous disk shelf in the loop and set it aside. **b.** Move the cable connecting the controller to the module A output port of the last disk shelf to the module A output port of the previous disk shelf in the loop

6. Verify the cabling on path A has successfully bypassed the disk shelves you are removing and all disk drives on the disk shelves you are removing are still connected through path B, by entering the following command from the nodeshell of either controller:
`storage show disk -p`

It might take up to a minute for the system to complete discovery.

Example

In this example of how the output should appear, the disk shelf being removed is disk shelf 3. One line item appears for each disk drive connected through path B (now the primary path); therefore, the disk drives are listed in the **PRIMARY** column and `B` appears in the first **PORT** column. There is no connectivity through path A for any of the disk drives in the disk shelf being removed; therefore, no information is shown in the **SECONDARY** or second **PORT** columns:

```
fas6200> storage show disk -p

PRIMARY    PORT  SECONDARY  PORT SHELF BAY
---------- ----  ---------- ---- ---------
...
0d.64       B                     3    0
0d.65       B                     3    1
0d.66       B                     3    2
0d.67       B                     3    3
0d.68       B                     3    4
0d.69       B                     3    5
0d.70       B                     3    6
0d.71       B                     3    7
...
```

Attention: If the output shows anything other than all the disk drives connected only through path B, you must correct the cabling by repeating Step 5.

7. Complete the following substeps:
   a) Repeat Step 5 and Step 6 for path B.
   b) Repeat Step 1 to confirm that your system configuration is the same as before you began this procedure.
   c) Go to the next step.
8. If, when you unowned disk drives as part of the preparation for this procedure, you disabled disk autoassignment, then reenable disk autoassignment by entering the following command; otherwise, go to the next step:

   ```
   storage disk option modify -autoassign on
   ```

   Enter the applicable command from the clustershell of each controller.
9. Power off the disk shelves you disconnected and unplug the power cords from the disk shelves.
10. Remove the disk shelves from the rack or cabinet.

    To make disk shelves lighter and easier to maneuver, remove the power supplies and modules. Avoid removing the disk drives if possible, because excessive handling can lead to internal damage.

Related concepts

Best practices for changing module types on page 81

Related tasks

Hot-swapping a module on page 84
Determining path status for your HA pair on page 82

Relocating aggregate ownership within an HA pair

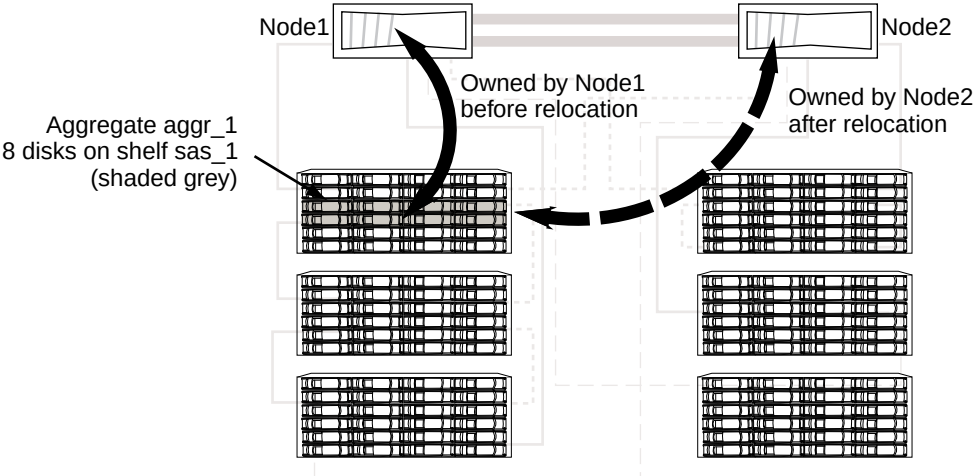
You can change the ownership of aggregates among the nodes in an HA pair without interrupting service from the aggregates.

Both nodes in an HA pair are physically connected to each other's disks or array LUNs. Each disk or array LUN is owned by one of the nodes. While ownership of disks temporarily changes when a takeover occurs, the aggregate relocation operations either permanently (for example, if done for load balancing) or temporarily (for example, if done as part of takeover) change the ownership of all disks or array LUNs within an aggregate from one node to the other. The ownership changes without any data-copy processes or physical movement of the disks or array LUNs.

How aggregate relocation works

Aggregate relocation operations take advantage of the HA configuration to move the ownership of storage aggregates within the HA pair. Aggregate relocation occurs automatically during manually initiated takeovers to reduce downtime during planned failover events such as nondisruptive software upgrades, and can be initiated manually for load balancing, maintenance, and nondisruptive controller upgrades. Aggregate relocation cannot move ownership of the root aggregate.

The following illustration shows the relocation of the ownership of aggregate aggr_1 from Node1 to Node2 in the HA pair:



The aggregate relocation operation can relocate the ownership of one or more SFO aggregates if the destination node can support the number of volumes in the aggregates. There is only a short interruption of access to each aggregate. Ownership information is changed one by one for the aggregates.

During takeover, aggregate relocation happens automatically when the takeover is initiated manually. Before the target controller is taken over, ownership of the aggregates belonging to that controller are moved one at a time to the partner controller. When giveback is initiated, the ownership is automatically moved back to the original node. The `-bypass-optimization` parameter can be used with the `storage failover takeover` command to suppress aggregate relocation during the takeover.

The aggregate relocation requires additional steps if the aggregate is currently used by an Infinite Volume with SnapDiff enabled.

Aggregate relocation and Infinite Volumes with SnapDiff enabled

The aggregate relocation requires additional steps if the aggregate is currently used by an Infinite Volume with SnapDiff enabled. You must ensure that the destination node has a namespace mirror constituent, and make decisions about relocating aggregates that include namespace constituents.

For information about Infinite Volumes, see the *Clustered Data ONTAP Infinite Volumes Management Guide*.

Relocating aggregate ownership

You can change the ownership of an aggregate only between the nodes within an HA pair.

About this task

- Because volume count limits are validated programmatically during aggregate relocation operations, it is not necessary to check for this manually. If the volume count exceeds the supported limit, the aggregate relocation operation will fail with a relevant error message.
- You should not initiate aggregate relocation when system-level operations are in progress on either the source or the destination node; likewise, you should not start these operations during the aggregate relocation. These operations can include:
  - Takeover
  - Giveback
  - Shutdown
  - Another aggregate relocation operation
  - Disk ownership changes
  - Aggregate or volume configuration operations
  - Storage controller replacement
  - Data ONTAP upgrade
  - Data ONTAP revert
- You should not initiate aggregate relocation on aggregates that are corrupt or undergoing maintenance.
- If the source node is used by an Infinite Volume with SnapDiff enabled, you must perform additional steps before initiating the aggregate relocation and then perform the relocation in a

specific manner. You must ensure that the destination node has a namespace mirror constituent and make decisions about relocating aggregates that include namespace constituents.
For information about Infinite Volumes, see the *Clustered Data ONTAP Infinite Volumes Management Guide*.

- Before initiating the aggregate relocation, save any core dumps on the source and destination nodes.

Steps

1. View the aggregates on the node to confirm which aggregates to move and ensure they are online and in good condition:

   storage aggregate show -node *source-node*

   #### Example

   The following command shows six aggregates on the four nodes in the cluster. All aggregates are online. Node1 and Node 3 form an HA pair and Node2 and Node4 form an HA pair.

   ```
   node1::> storage aggregate show
   Aggregate     Size Available Used% State   #Vols  Nodes  RAID Status
   --------- -------- --------- ----- ------- ------ ------ -----------
   aggr_0     239.0GB   11.13GB   95% online       1 node1  raid_dp,
                                                             normal
   aggr_1     239.0GB   11.13GB   95% online       1 node1  raid_dp,
                                                             normal
   aggr_2     239.0GB   11.13GB   95% online       1 node2  raid_dp,
                                                             normal
   aggr_3     239.0GB   11.13GB   95% online       1 node2  raid_dp,
                                                             normal
   aggr_4     239.0GB   238.9GB    0% online       5 node3  raid_dp,
                                                             normal
   aggr_5     239.0GB   239.0GB    0% online       4 node4  raid_dp,
                                                             normal
   6 entries were displayed.
   ```

2. Issue the command to start the aggregate relocation:

   storage aggregate relocation start -aggregate-list *aggregate-1, aggregate-2...* -node *source-node* -destination *destination-node*

   The following command moves the aggregates aggr_1 and aggr_2 from Node1 to Node3. Node3 is Node1's HA partner. The aggregates can only be moved within the HA pair.

   ```
   node1::> storage aggregate relocation start -aggregate-list aggr_1,
   aggr_2 -node node1 -destination node3
   Run the storage aggregate relocation show command to check relocation
   status.
   node1::storage aggregate>
   ```

3. Monitor the progress of the aggregate relocation with the `storage aggregate relocation show` command:

```
storage aggregate relocation show -node source-node
```

Example

The following command shows the progress of the aggregates that are being moved to Node3:

```
node1::> storage aggregate relocation show -node node1
Source Aggregate   Destination    Relocation Status
------ ----------- -------------  ------------------------
node1
       aggr_1      node3          In progress, module: wafl
       aggr_2      node3          Not attempted yet
2 entries were displayed.
node1::storage aggregate>
```

When the relocation is complete, the output of this command shows each aggregate with a relocation status of `Done`.

Related concepts

Background disk firmware update and takeover, giveback, and aggregate relocation on page 26

Commands for aggregate relocation

There are specific Data ONTAP commands for relocating aggregate ownership within an HA pair.

If you want to...	Use this command...
Start the aggregate relocation process.	`storage aggregate relocation start`
Monitor the aggregate relocation process	`storage aggregate relocation show`

See the man page for each command for more information.

Key parameters of the storage aggregate relocation start command

The `storage aggregate relocation start` command includes several key parameters used when relocating aggregate ownership within an HA pair.

Parameter	Meaning
-node *nodename*	Specifies the name of the node that currently owns the aggregate

Parameter	Meaning
-destination *nodename*	Specifies the destination node where aggregates are to be relocated
-aggregate-list *aggregate name*	Specifies the list of aggregate names to be relocated from source node to destination node (This parameter accepts wildcards)
-override-vetoes *true*|*false*	Specifies whether to override any veto checks during the relocation operation
-relocate-to-higher-version *true*|*false*	Specifies whether the aggregates are to be relocated to a node that is running a higher version of Data ONTAP than the source node
-override-destination-checks *true*|*false*	Specifies if the aggregate relocation operation should override the check performed on the destination node

See the man page for more information.

Veto and destination checks during aggregate relocation

In aggregate relocation operations, Data ONTAP determines whether aggregate relocation can be completed safely. If aggregate relocation is vetoed, you must check the EMS messages to determine the cause. Depending on the reason or reasons, you can decide whether you can safely override the vetoes.

The `storage aggregate relocation show` command displays the aggregate relocation progress and shows which subsystem, if any, vetoed the relocation. Soft vetoes can be overridden, whereas hard vetoes cannot be, even if forced. The following tables summarize the soft and hard vetoes, along with recommended workarounds.

Veto checks during aggregate relocation

Vetoing subsystem module	Workaround
Vol Move	Relocation of an aggregate is vetoed if any volumes hosted by the aggregate are participating in a volume move that has entered the cutover state. Wait for volume move to complete. If this veto is overridden, cutover will resume automatically once the aggregate relocation completes. If aggregate relocation causes the move operation to exceed the number of retries (the default is 3), then the user needs to manually initiate cutover using the `volume move trigger-cutover` command.
Backup	Relocation of an aggregate is vetoed if a dump or restore job is in progress on a volume hosted by the aggregate. Wait until the dump or restore operation in progress is complete. If this veto is overridden, the backup or restore operation will be aborted and must be restarted by the backup application.
Lock manager	To resolve the issue, gracefully shut down the CIFS applications that have open files, or move those volumes to a different aggregate. Overriding this veto will result in loss of CIFS lock state, causing disruption and data loss.
Lock Manager NDO	Wait until the locks are mirrored. This veto cannot be overridden; doing so will cause disruption to Microsoft Hyper-V virtual machines.
RAID	Check the EMS messages to determine the cause of the veto: If disk add or disk ownership reassignment operations are in progress, wait until they complete. If the veto is due to mirror resync, mirror verify, or offline disks, the veto can be overridden and the operation will be restarted after giveback.

Destination checks during aggregate relocation

Vetoing subsystem module	Workaround
Disk Inventory	Relocation of an aggregate will fail if the destination node is unable to see one or more disks belonging to the aggregate. Check storage for loose cables and verify that the destination can access disks belonging to the aggregate being relocated. This check cannot be overridden.
WAFL	Relocation of an aggregate will fail if allowing the relocation to proceed would cause the destination to exceed its limits for maximum volume count or maximum volume size. This check cannot be overridden.
Lock Manager NDO	Relocation of an aggregate will fail if: • The destination does not have sufficient lock manager resources to reconstruct locks for the relocating aggregate. • The destination node is reconstructing locks. Retry aggregate relocation after a few minutes. This check cannot be overridden.
Lock Manager	Permanent relocation of an aggregate will fail if the destination does not have sufficient lock manager resources to reconstruct locks for the relocating aggregate. Retry aggregate relocation after a few minutes. This check cannot be overridden.
RAID	Check the EMS messages to determine the cause of the failure: • If the failure is due to an aggregate name or UUID conflict, troubleshoot and resolve the issue. This check cannot be overridden. Relocation of an aggregate will fail if allowing the relocation to proceed would cause the destination to exceed its limits for maximum aggregate count, system capacity, or aggregate capacity. You should avoid overriding this check.

Copyright and trademark information

References in this documentation to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program, or service is not intended to state or imply that only IBM's product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any of IBM's or NetApp's intellectual property rights may be used instead of the IBM or NetApp product, program, or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM and NetApp, are the user's responsibility.

No part of this document covered by copyright may be reproduced in any form or by any means— graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT

(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, CyberSnap, Data Center Fitness, Data ONTAP, DataFabric, DataFort, Decru, Decru DataFort, DenseStak, Engenio, Engenio logo, E-Stack, ExpressPod, FAServer, FastStak, FilerView, Flash Accel, Flash Cache, Flash Pool, FlashRay, FlexCache, FlexClone, FlexPod, FlexScale, FlexShare, FlexSuite, FlexVol, FPolicy, GetSuccessful, gFiler, Go further, faster, Imagine Virtually Anything, Lifetime Key Management, LockVault, Mars, 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, Snap Creator, SnapDirector, SnapDrive, SnapFilter, SnapIntegrator, SnapLock, SnapManager, SnapMigrator, SnapMirror, SnapMover, SnapProtect, SnapRestore, Snapshot, SnapSuite, SnapValidator, SnapVault, StorageGRID, StoreVault, the StoreVault logo, SyncMirror, Tech OnTap, The evolution of storage, Topio, VelocityStak, 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

D

E

F

G

H

S

T

U

V

NA 210-06381_A0, Printed in USA

SC27-6276-01