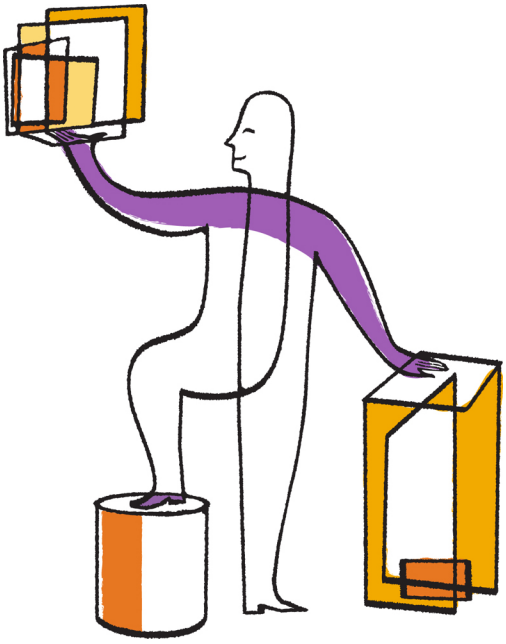
Clustered Data ONTAP® 8.3

MetroCluster Management and Disaster Recovery Guide



NetApp, Inc.
495 East Java Drive
Sunnyvale, CA 94089
U.S.

Telephone: +1 (408) 822-6000
Fax: +1 (408) 822-4501
Support telephone: +1 (888) 463-8277
Web: www.netapp.com
Feedback: doccomments@netapp.com

Part number: 215-09187_B0
January 2015

Contents

Where to find MetroCluster documentation

There are a number of documents that can help you configure, operate, and monitor a MetroCluster configuration.

MetroCluster and Data ONTAP libraries

Library	Content
NetApp Documentation: MetroCluster in clustered Data ONTAP	• All MetroCluster guides
NetApp Documentation: Clustered Data ONTAP Express Guides	• All Data ONTAP express guides
NetApp Documentation: Data ONTAP 8 (current releases)	• All Data ONTAP guides

MetroCluster and miscellaneous guides

Guide	Content
Clustered Data ONTAP 8.3 MetroCluster Installation Express Guide	How to install a MetroCluster system that has been received from the factory. You should use this guide only if the following is true: • The MetroCluster configuration has been received from the factory. • The configuration is using Brocade FC storage switches. This guide does not document configuration of the Cisco FC storage switches. • The configuration is not using array LUNs (FlexArray Virtualization). • The configuration is not sharing existing FC fabrics with a 7-Mode fabric MetroCluster during transition.

Guide	Content
Clustered Data ONTAP 8.3 MetroCluster Installation and Configuration Guide	• MetroCluster architecture • Cabling the configuration • Configuring the FC-to-SAS bridges • Configuring the FC switches • Configuring the MetroCluster in Data ONTAP • Configuring the MetroCluster tie-breaker application
MetroCluster Service Guide	• Guidelines for maintenance in a MetroCluster configuration • Hardware replacement and firmware upgrade procedures for FC-to-SAS bridges and FC switches • Hot-adding a disk shelf • Hot-removing a disk shelf • Replacing hardware at a disaster site
MetroCluster Tiebreaker Software Installation and Configuration Guide	• Monitoring the MetroCluster configuration with the MetroCluster Tiebreaker software
Clustered Data ONTAP 8.3 Data Protection Guide	• How mirrored aggregates work • SyncMirror • SnapMirror • SnapVault
NetApp Documentation: OnCommand Unified Manager Core Package (current releases)	• Monitoring the MetroCluster configuration
NetApp Documentation: OnCommand Performance Manager for Clustered Data ONTAP	• Monitoring MetroCluster performance

Guide	Content
7-Mode Transition Tool 2.0 Data and Configuration Transition Guide	• Transitioning data from 7-Mode storage systems to clustered storage systems

Understanding MetroCluster data protection and disaster recovery

It is helpful to understand how MetroCluster protects data and provides transparent recovery from failures so that you can manage your switchover and switchback activities easily and efficiently.

MetroCluster uses mirroring to protect the data in a cluster. It provides disaster recovery through a single MetroCluster command that activates a secondary on the survivor site to serve the mirrored data originally owned by a primary site affected by disaster.

How MetroCluster configurations provide local failover and switchover

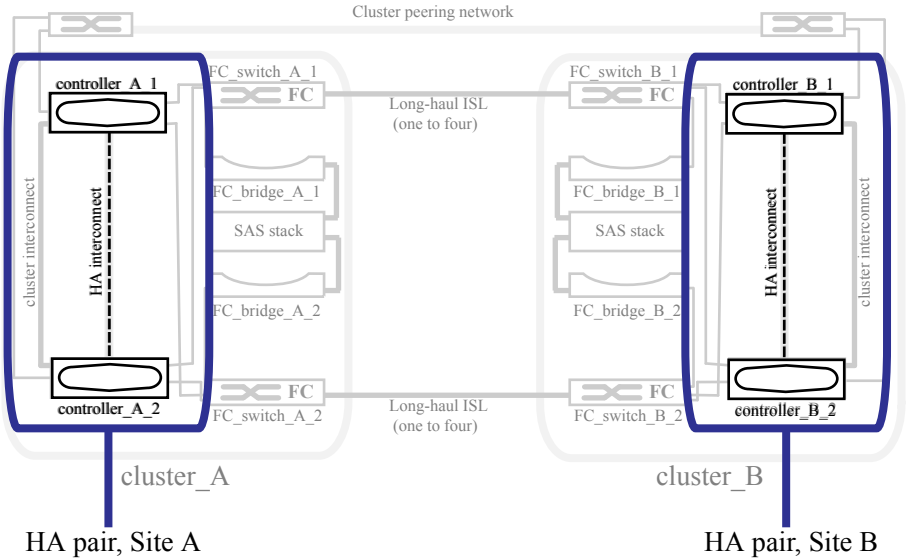
MetroCluster configurations protect data on both a local level and cluster level. If you're setting up a MetroCluster configuration, you need to know how MetroCluster configurations protect your data.

MetroCluster configurations protect data by using two physically separated, mirrored clusters. Each cluster synchronously mirrors the data and Storage Virtual Machine (SVM) configuration of the other. When a disaster occurs at one site, an administrator can activate the mirrored SVM and begin serving the mirrored data from the surviving site. Additionally, the nodes in each cluster are configured as an HA pair, providing a level of local failover.

How local HA data protection works

The two clusters in the peered network provide bidirectional disaster recovery, where each cluster can be the source and backup of the other cluster. Each cluster includes two nodes, which are

configured as an HA pair. In the case of a failure or required maintenance within a single node's configuration, storage failover can transfer that node's operations to its local HA partner.



Related information

Clustered Data ONTAP 8.3 High-Availability Configuration Guide

Replication of SVMs and switchover

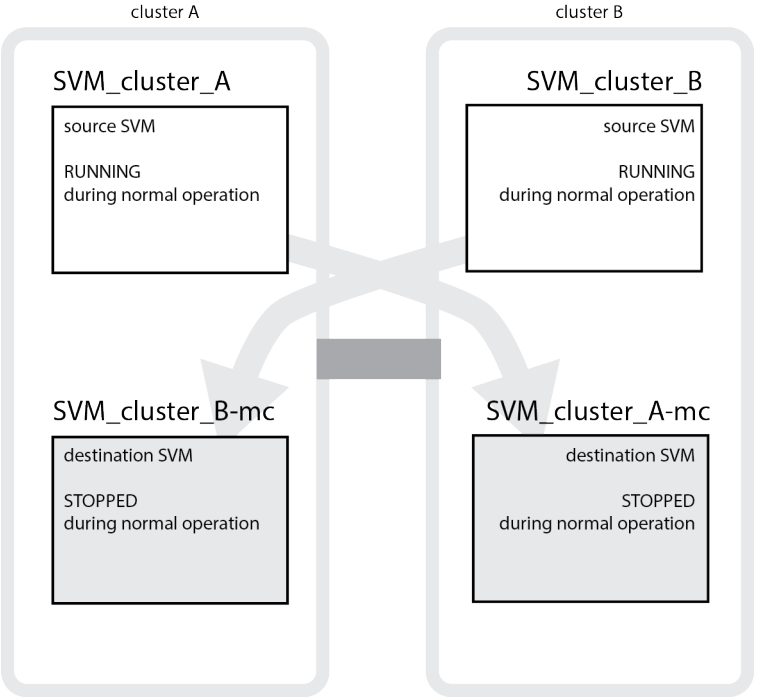
SVM mirroring provides redundant data server configuration and mirroring of data volumes that belong to the SVM. If a switchover occurs, the source SVM is brought down and the destination SVM, located on the surviving cluster, becomes active.

The following example shows the SVMs for a MetroCluster configuration, where vs1 is a SVM on the source site and vs-1mc is a sync-destination on the disaster recovery site (MetroCluster appends `-mc` to the name of the destination SVMs):

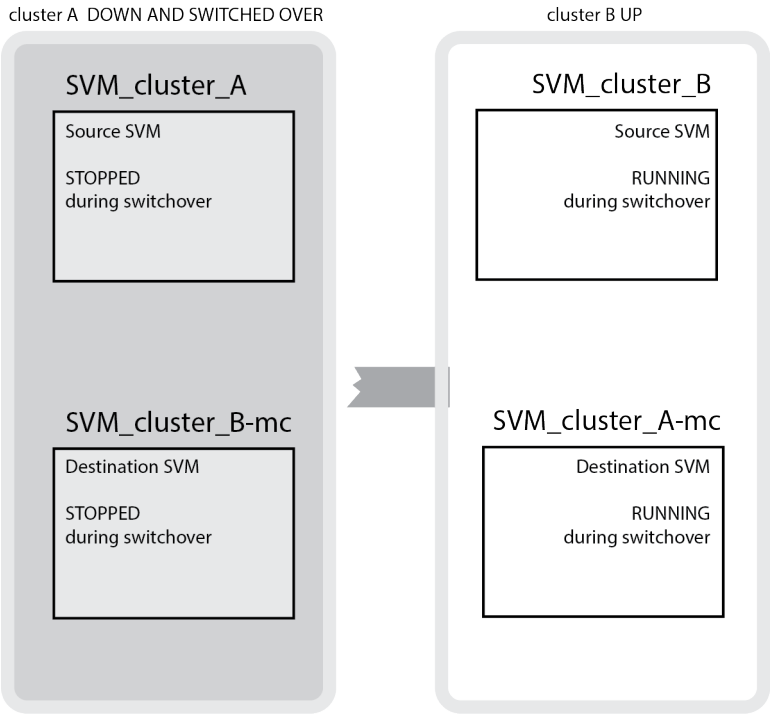
- vs1 serves data on cluster A.
  It is a sync-source SVM that replicates the SVM configuration (LIFs, protocols, and services) and data in volumes belonging to the SVM. The configuration and data are replicated to vs1-mc, a sync-destination SVM located on cluster B.
- vs2 serves data on cluster B.
  It is a sync-source SVM that replicates configuration and data to vs2-mc located on cluster A.
- vs2-mc is a sync-destination that is stopped during normal, healthy operation of the MetroCluster configuration.
  In a successful switchover from cluster B to cluster A, vs2 is stopped and vs2-mc is activated and begins serving data from cluster A.

- vs1-mc is a sync-destination that is stopped during normal, healthy operation of the MetroCluster configuration.

  In a successful switchover from cluster A to cluster B, vs1 is stopped and vs1-mc is activated and begins serving data from cluster B.



If a switchover occurs, the remote plex on the surviving cluster comes online and the secondary SVM begins serving the data.



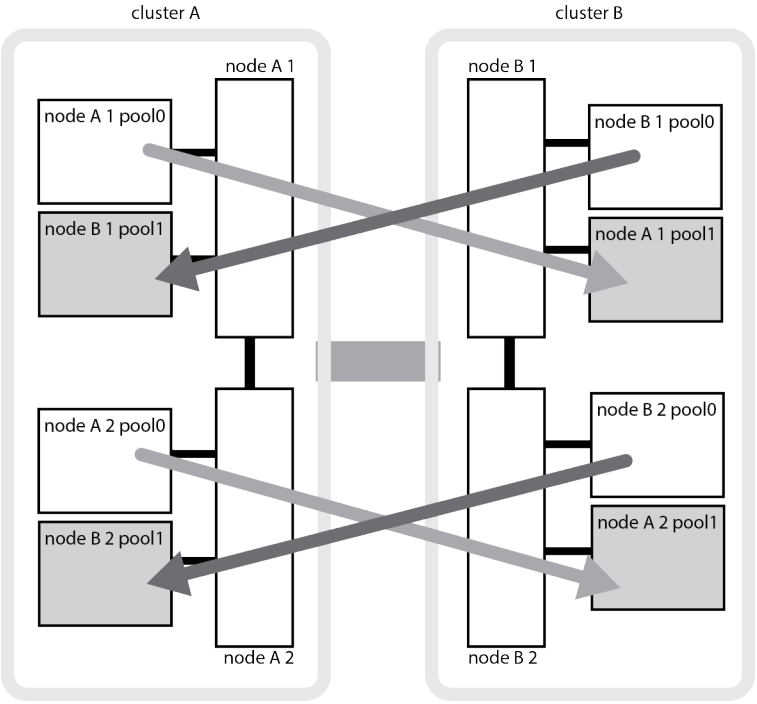
Related information

Clustered Data ONTAP 8.3 System Administration Guide for Cluster Administrators

Aggregate mirroring with SyncMirror

Mirrored aggregates using SyncMirror functionality provide data redundancy and contain the volumes owned by the source and destination SVM.

The following illustration shows how the disk pools are mirrored between the partner clusters. Data in the local plexes (in pool0) is replicated to the remote plexes (in pool1).



Related information

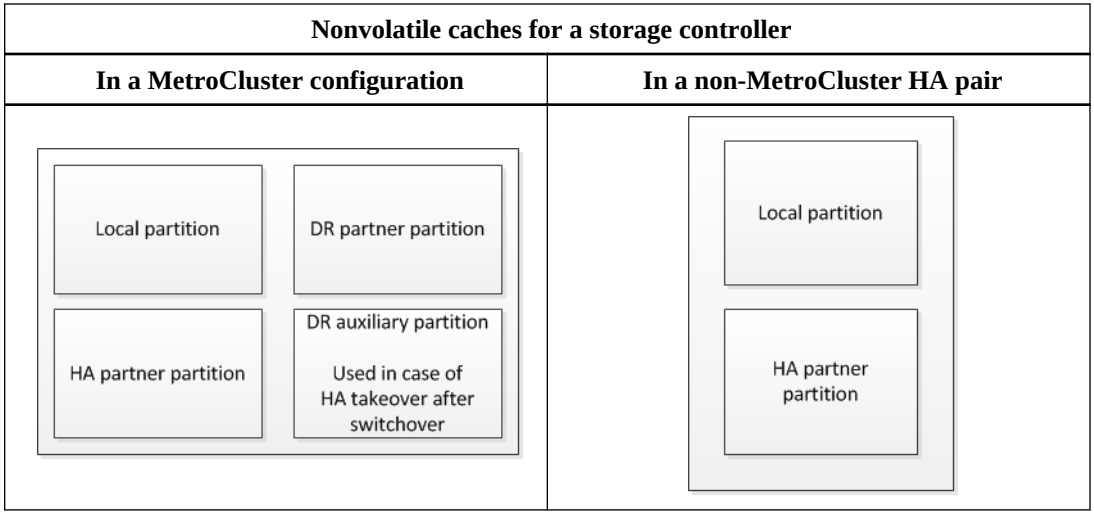
Clustered Data ONTAP 8.3 Data Protection Guide

NVRAM and NVMEM cache mirroring in a MetroCluster configuration

The nonvolatile memory (NVRAM or NVMEM, depending on the platform model) in the storage controllers is mirrored both locally to a local HA partner and remotely to a remote DR partner on the partner site. In the event of local failover or switchover, this ensures that data in the hardware cache is preserved.

In an HA pair that is not part of a clustered MetroCluster configuration, each controller maintains two caches, one for itself, and one for its HA partner.

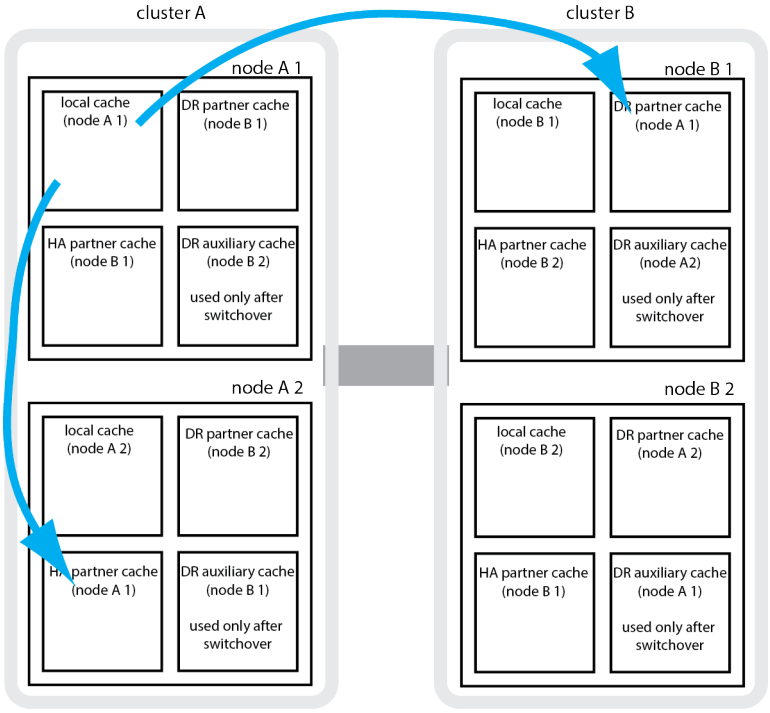
In a MetroCluster configuration, the nonvolatile cache of each storage controller is divided into four partitions.



The caches store the following content:

- *Local partition* holds data that the storage controller has not yet written to disk.
- *HA partner partition* holds a copy of the local cache of the storage controller's HA partner.
- *DR partner partition* holds a copy of the local cache of the storage controller's DR partner.
  The DR partner is a node in the partner cluster that is paired with the local node.
- *DR auxiliary partner partition* holds a copy of the local cache of the storage controller's DR auxiliary partner.
  The DR partner is the HA partner of the local node's DR partner. This cache is needed if there is an HA takeover after a MetroCluster switchover.

For example, the local cache of a node (node_A_1) is mirrored both locally and remotely in the MetroCluster sites:



Types of disasters and recovery methods

You need to be familiar with different types of failures and disasters so that you can use the MetroCluster configuration to respond appropriately.

- Single node failure

  A single component in the local HA pair fails.

  This failure might lead to an automatic or negotiated takeover of the impaired node, depending on the component that failed. Data recovery is described in the following guide: *Clustered Data ONTAP 8.3 High-Availability Configuration Guide*

- Site failure

  Both controller modules fail at a site due to loss of power, replacement of equipment, or disaster. Typically, MetroCluster configurations cannot differentiate between failures and disasters; however, witness software, such as the MetroCluster Tiebreaker software, can differentiate between them.

  This guide describes how to recover from site failures that do not include controller failures as well as failures that include of one or more controllers.

- ISL failure

A failure occurs between sites when links are lost, whereby the MetroCluster configuration takes no action. Each node continues to serve out data normally, but the mirrors are not written to the respective disaster recovery sites because access to them is lost.

- Multiple sequential failures
  Multiple components fail in a sequence. For example, a controller module fails, followed by a switch fabric failure, followed by the shelf failure results in storage failover, fabric redundancy, and SyncMirror sequentially protecting against downtime and data loss.

The following table shows failure types, and the corresponding DR mechanism and recovery method:

Failure type	DR mechanism	Recovery method
Single node failure	Local HA failover	Not required if automatic failover is enabled.
Site failure	MetroCluster forced switchover	Switch over from the surviving site using the `metrocluster switchover -forced-on-disaster` command.
ISL failure	No MetroCluster switchover; the two clusters independently serve their data	Not required for this type of failure. After you restore connectivity, the storage will resync automatically.
Multiple sequential failures	Local HA failover followed by MetroCluster switchover	Switch over using the `metrocluster switchover -forced-on-disaster` command **Note:** Depending on the component that failed, a forced switchover might not be required.

Related tasks

Performing a forced switchover after a disaster on page 34

How MetroCluster provides nondisruptive operations

In the case of an issue limited to a single node, failover and giveback within the local HA pair provides continued nondisruptive operation. In this case, the MetroCluster configuration does not require switchover to the remote site.

Because the MetroCluster configuration consists of one HA pair at each site, each site can withstand local failures and perform nondisruptive operations without requiring a switchover to the partner site. The operation of the HA pair is the same as HA pairs in non-MetroCluster configurations.

Clustered Data ONTAP 8.3 High-Availability Configuration Guide

Consequences of local failover after switchover

If a MetroCluster switchover occurs, and then an issue arises at the surviving site, a local failover can provide continued, nondisruptive operation. However, the system is at risk because it is no longer in a redundant configuration.

If a local failover occurs after a switchover has occurred, a single controller serves data for all storage systems in the MetroCluster configuration, leading to possible resource issues, and is vulnerable to additional failures.

Overview of the switchover process

In the case of a site-wide issue, the MetroCluster switchover operation allows immediate resumption of service by moving storage and client access from the source cluster to the disaster site. The DR partner nodes begin serving data from the mirrored plexes and the sync destination SVM.

During the switchover operation, the system takes the following actions:

- The ownership of disks that belong to the disaster site is changed to the DR partner.
  This is similar to the case of a local failover within an HA pair, in which ownership of the disks belonging to the down partner is changed to the healthy partner.
- The surviving plexes located on the surviving site but belonging to the nodes in the disaster cluster are brought online on the cluster at the surviving site.
- The sync-source SVM belonging to the disaster site is brought down.

  Note: This is only applicable to negotiated switchover.

- The sync-destination SVM belonging to the disaster site is brought up.

Root aggregates of the DR partners, while being switched over, are not brought online during switchover.

If you are only switching over services to the remote site, you should perform a negotiated switchover without fencing the site. However, for power maintenance, you should first take the plexes offline at the disaster site, fence the site, and then bring the plexes online. If storage or equipment is unreliable, you should fence the disaster site and then perform a negotiated switchover. Fencing prevents RAID reconstructions when disks power up in a staggered manner.

Related tasks

Fencing off the disaster site on page 34

Disk ownership changes during HA takeover and MetroCluster switchover

The ownership of disks temporarily changes automatically during high availability and MetroCluster operations. It is helpful to know how the system tracks which node owns which disks.

In Data ONTAP, a controller module's unique system ID (obtained from a node's NVRAM card or NVMEM board) is used to identify which node owns a specific disk. Depending on the HA or DR state of the system, the ownership of the disk might temporarily change. If the ownership changes because of an HA takeover or a DR switchover, the system records which node is the original (called “home”) owner of the disk, so that it can return the ownership after HA giveback or DR switchback. The system uses the following fields to track disk ownership:

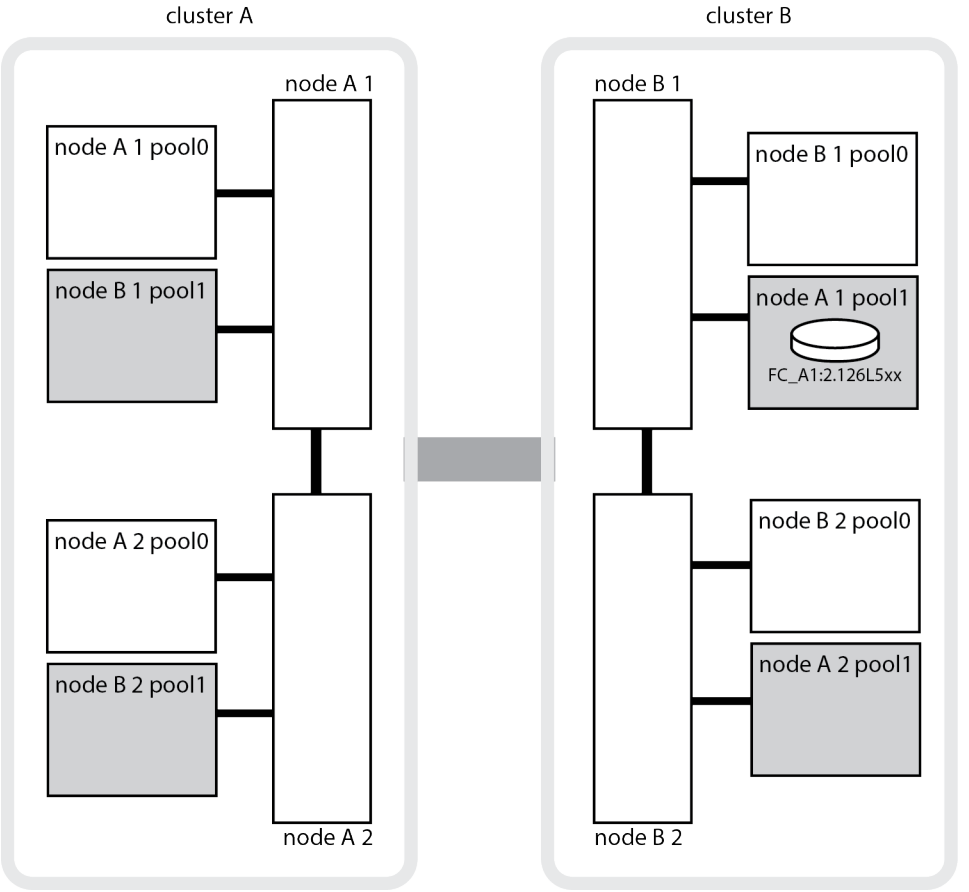
- `Owner`
- `Home owner`
- `DR Home owner`

In the MetroCluster configuration, in the event of a switchover, a node can take ownership of an aggregate originally owned by nodes in the partner cluster. Such aggregates are referred to as *cluster-foreign aggregates*. The distinguishing feature of a cluster-foreign aggregate is that it is an aggregate not currently known to the cluster, and so the `DR Home owner` field is used to show that it is owned by a node from the partner cluster. A traditional foreign aggregate within an HA pair is identified by `Owner` and `Home owner` values being different, but the `Owner` and `Home owner` values are the same for a cluster-foreign aggregate; thus, you can identify a cluster-foreign aggregate by the `DR Home owner` value.

As the state of the system changes, the values of the fields change, as shown in the following table:

	Value during...			
Field	**Normal operation**	**Local HA takeover**	**MetroCluster switchover**	**Takeover during switchover**
`Owner`	ID of the node that has access to the disk.	ID of the the HA partner, which temporarily has access to the disk.	ID of the DR partner, which temporarily has access to the disk.	ID of the DR auxiliary partner, which temporarily has access to the disk.
`Home owner`	ID of the original owner of the disk within the HA pair.	ID of the original owner of the disk within the HA pair.	ID of the DR partner, which is the `Home owner` in the HA pair during the switchover.	ID of the DR partner, which is the `Home owner` in the HA pair during the switchover.
`DR Home owner`	Empty	Empty	ID of the original owner of the disk within the MetroCluster configuration.	ID of the original owner of the disk within the MetroCluster configuration.

The following illustration and table provide an example of how ownership changes, for a disk in node_A_1's disk pool1, physically located in cluster_B.



MetroCluster state	Owner	Home owner	DR Home owner	Notes
Normal with all nodes fully operational.	node_A_1	node_A_1	not applicable	
Local HA takeover, node_A_2 has taken over disks belonging to its HA partner node_A_1.	node_A_2	node_A_1	not applicable	

MetroCluster state	Owner	Home owner	DR Home owner	Notes
DR switchover, node_B_1 has taken over disks belong to its DR partner, node_A_1.	node_B_1	node_B_1	node_A_1	The original home node ID is moved to the `DR Home owner` field. After aggregate switchback or healing, ownership goes back to node_A_1.
In DR switchover and local HA takeover (double failure), node_B_2 has taken over disks belonging to its HA node_B_1.	node_B_2	node_B_1	node_A_1	After giveback, ownership goes back to node_B_1. After switchback or healing, ownership goes back to node_A_1.
After HA giveback and DR switchback, all nodes fully operational.	node_A_1	node_A_1	not applicable	

What happens during healing

During healing, the resynchronization of mirrored aggregates occurs in a phased process that prepares the nodes at the repaired disaster site for switchback. It is a planned event, thereby giving you full control of each step to minimize downtime. Healing is a two-step process that occurs on the storage and controller components.

Data aggregate healing

After the problem at the disaster site is resolved, you start the storage healing phase:

1. Checks that all nodes are up and running for the surviving site HA pair.
2. Changes ownership of all the pool 0 disks at the disaster site, including root aggregates.

During this phase of healing, the RAID subsystem resynchronizes mirrored aggregates, and the WAFL subsystem replays both the `nvsave` files of unmirrored aggregates and the `nvsave` files of mirrored aggregates that had a failed pool 1 plex at the time of switchover.

If no errors are reported, the aggregates are successfully resynchronized. This process can sometimes take hours to complete.

If some source storage components failed, the command reports the errors at applicable levels: `Storage`, `Sanown`, or `RAID`.

If no errors are reported, the aggregates are successfully resynchronized. This process can sometimes take hours to complete.

Healing the data aggregates on page 75

Root aggregate healing

After the aggregates are synchronized, you start the controller healing phase by giving back the CFO aggregates and root aggregates to their respective DR partners.

Healing the root aggregates on page 76

What happens during a switchback

After the disaster site has recovered and aggregates have healed, the MetroCluster switchback process returns storage and client access from the disaster recovery site to the home cluster.

The `metrocluster switchback` command returns the primary site to full, normal MetroCluster operation. Any configuration changes are propagated to the original SVMs. Data server operation is then returned to the sync-source SVMs on the disaster site and the sync-dest SVMs that had been operating on the surviving site are deactivated.

If SVMs were deleted on the surviving site while the MetroCluster configuration was in switchover state, the switchback process does the following:

- Deletes the corresponding SVMs on the partner site (the former disaster site).
- Deletes any peering relationships of the deleted SVMs.

Related tasks

Reestablishing SVM peering relationships after switchover on page 26

Performing switchover for tests or maintenance

If you want to test the MetroCluster functionality or to perform planned maintenance, you can perform a negotiated switchover in which one cluster is cleanly switched over to the partner cluster. You can then heal and switch back the configuration.

Steps

1. Verifying that your system is ready for a switchover on page 23
2. Performing a negotiated switchover on page 24
3. Confirming that the DR partners have come online on page 25
4. Reestablishing SnapMirror or SnapVault SVM peering relationships on page 26
5. Healing the configuration on page 27
6. Performing a switchback on page 30
7. Verifying a successful switchback on page 31
8. Reestablishing SnapMirror or SnapVault SVM peering relationships on page 33

Verifying that your system is ready for a switchover

You can use the `-simulate` option to preview the results of a switchover. A verification check gives you a way to ensure that most of the preconditions for a successful run are met before you start the operation.

Steps

1. Simulated the switchover operation at the advanced privilege level:

   `metrocluster switchover -simulate`

   .

2. Review the output that is returned.

   The output shows whether any vetoes would prevent a switchover.

Example: Verification results

The following example shows errors encountered in a simulation of a switchover operation:

```
cluster4::*> metrocluster switchover -
simulate

[Job 126] Preparing the cluster for the switchover operation...
```

```
[Job 126] Job failed: Failed to prepare the cluster for the switchover
operation. Use the "metrocluster operation show" command to view detailed error
information. Resolve the errors, then try the command again.
```

Performing a negotiated switchover

A negotiated switchover cleanly shuts down processes on the partner site and then switches over operations from the partner site. Negotiated switchover can be used to perform maintenance on a MetroCluster site or test the switchover functionality.

Before you begin

- Any nodes that were previously down must be booted and in cluster quorum. The cluster peering network must be available from both sites.

About this task

While preparing and executing the negotiated switchover, do not make configuration changes to either cluster.

After the switchover operation finishes, all SnapMirror and SnapVault relationships using a disaster site node as a destination must be reconfigured.

Clustered Data ONTAP 8.3 Data Protection Guide

Steps

1. Use the `metrocluster check run`, `metrocluster check show` and `metrocluster check config-replication show` to make sure no configuration updates are in progress or pending.

   Checking the MetroCluster configuration on page 89

2. Enter the following command to implement the switchover:

   `metrocluster switchover`

   The operation can take several minutes to complete.

3. Monitor the completion of the switchover:

   `metrocluster operation show`

   Example

```
mcc1A::*> metrocluster operation show
  Operation: Switchover
 Start time: 10/4/2012 19:04:13
      State: in-progress
```

```
   End time: -
     Errors:

mcc1A::*> metrocluster operation show
  Operation: Switchover
 Start time: 10/4/2012 19:04:13
      State: successful
   End time: 10/4/2012 19:04:22
     Errors: -
```

4. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Confirming that the DR partners have come online

After the switchover is complete, you should verify that the DR partners have taken ownership of the disks and the partner SVMs have come online.

Steps

1. Confirm that the aggregate disks have switched over to the disaster site:

   `storage disk show -fields owner,dr-home`

   #### Example

   In this example, the output shows that the switched over disks have the `dr-home` field set:

```
mcc1A::> storage disk show -fields owner,dr-home
disk                              owner   dr-home
--------------------------------- ------- -------
mcc1-a1:mcc-sw1A-fab1:1-7.126L1   mcc1-a1 -
mcc1-a1:mcc-sw1A-fab1:1-7.126L24  mcc1-a1 mcc1-b2
mcc1-a1:mcc-sw1A-fab1:1-7.126L36  mcc1-a1 mcc1-b2
mcc1-a1:mcc-sw1A-fab1:1-7.126L38  mcc1-a1 -
....
mcc1-a1:mcc-sw1A-fab1:1-7.126L48  mcc1-a1 -
mcc1-a1:mcc-sw1A-fab1:1-7.126L49  mcc1-a1 -
mcc1-a1:mcc-sw1A-fab1:1-8.126L6   mcc1-a1 -
mcc1-a1:mcc-sw1A-fab1:1-8.126L13  mcc1-a1 mcc1-b2
mcc1-a1:mcc-sw1A-fab1:1-8.126L23  mcc1-a1 mcc1-b2
mcc1-a1:mcc-sw1A-fab1:1-8.126L24  mcc1-a1 mcc1-b2
```

2. Check that the aggregates were switched over by using the `storage aggregate show` command.

   #### Example

   In this example, the aggregates were switched over. The root aggregate (aggr0_b2) is in a degraded state. The data aggregate (b2_aggr2) is in a mirrored, normal state:

```
mcc1A::*> storage aggregate show

.
.
.
mcc1-b Switched Over Aggregates:
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
aggr0_b2    227.1GB   45.1GB    80% online       0 mcc1-a1          raid_dp,
                                                                    mirror
                                                                    degraded
b2_aggr1    227.1GB   200.3GB    20% online       0 mcc1-a1          raid_dp,
                                                                    mirrored
                                                                    normal
```

3. Confirm that the secondary SVMs have come online by using the `vserver show` command.

Example

In this example, the previously dormant sync-destination SVMs on the secondary site have been activated and have an `Admin State` of `running`:

```
mcc1A::*> vserver show
                                     Admin      Root                 Name    Name
Vserver         Type  Subtype        State      Volume    Aggregate  Service Mapping
-----------     ----- ----------     ---------- --------- ---------- ------- -------
...
mcc1B-vs1b-mc  data   sync-destination  running    vs1b_vol   aggr_b1   file    file
```

Reestablishing SnapMirror or SnapVault SVM peering relationships

After a switchover or switchback operation, you must manually reestablish any SVM peering to clusters outside of the MetroCluster configuration if the destination of the relationship is in the MetroCluster configuration.

About this task

This procedure is done at the surviving site after a switchover has occurred or at the disaster site after a switchback has occurred.

Steps

1. Check whether the SVM peering relationships have been reestablished by using the `metrocluster vserver show` command.
2. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Healing the configuration

Following a switchover, you must perform a healing operation in a specific order to restore MetroCluster functionality.

Before you begin

- ISLs must be up and operating.
- Switchover must have been performed and the surviving site must be serving data.
- Nodes in the surviving site must not be in HA failover state (all nodes must be up and running for each HA pair).
- Nodes on the disaster site must be halted or remain powered off.
  They must not be fully booted during the healing process.
- Storage at the disaster site must be accessible (shelves are powered up, functional, and accessible).

About this task

The healing operation must first be performed on the data aggregates, and then on the root aggregates.

Steps

1. Healing the data aggregates after negotiated switchover on page 27
2. Healing the root aggregates after negotiated switchover on page 29

Healing the data aggregates after negotiated switchover

You must heal the data aggregates after completing any maintenance or testing. This process resynchronizes the data aggregates and prepares the disaster site for normal operation. You must heal the data aggregates prior to healing the root aggregates.

About this task

All configuration updates in the remote cluster successfully replicate to the local cluster. You power up the storage on the disaster site as part of this procedure, but you do not and must not power up the controller modules on the disaster site.

Steps

1. Ensure that switchover has been completed by running the `metrocluster operation show` command.

Example

```
controller_A_1::> metrocluster operation show
  Operation: switchover
      State: successful
 Start Time: 7/25/2014 20:01:48
   End Time: 7/25/2014 20:02:14
     Errors: -
```

2. Resynchronize the data aggregates by running the `metrocluster heal -phase aggregates` command from the surviving cluster.

Example

```
controller_A_1::> metrocluster heal -phase aggregates
[Job 130] Job succeeded: Heal Aggregates is successful.
```

If the healing is vetoed, you have the option of reissuing the `metrocluster heal` command with the `–override-vetoes` parameter. If you use this optional parameter, the system overrides any soft vetoes that prevent the healing operation.

3. Verify that the operation has been completed by running the `metrocluster operation show` command.

Example

```
controller_A_1::> metrocluster operation show
    Operation: heal-aggregates
      State: successful
Start Time: 7/25/2014 18:45:55
   End Time: 7/25/2014 18:45:56
     Errors: -
```

4. Check the state of the aggregates by running the `storage aggregate show` command.

Example

```
controller_A_1::> storage aggregate show
Aggregate     Size Available Used% State   #Vols  Nodes
RAID Status
--------- -------- --------- ----- ------- ------ ----------------
------------
...
aggr_b2    227.1GB   227.1GB    0% online       0 mcc1-a2
raid_dp, mirrored, normal...
```

5. If storage has been replaced at the disaster site, you might need to remirror the aggregates.

Clustered Data ONTAP 8.3 Data Protection Guide

Healing the root aggregates after negotiated switchover

After the data aggregates have been healed, you must heal the root aggregates in preparation for the switchback operation.

Before you begin

The data aggregates phase of the MetroCluster healing process must have been completed successfully.

Steps

1. Switch back the mirrored aggregates by running the `metrocluster heal -phase root-aggregates` command.

   Example

   ```
   mcc1A::> metrocluster heal -phase root-aggregates
   [Job 137] Job succeeded: Heal Root Aggregates is successful
   ```

   If the healing is vetoed, you have the option of reissuing the `metrocluster heal` command with the `–override-vetoes` parameter. If you use this optional parameter, the system overrides any soft vetoes that prevent the healing operation.

2. Confirm the heal operation is complete by running the `metrocluster operation show` command on the destination cluster:

   Example

   ```
   mcc1A::> metrocluster operation show
     Operation: heal-root-aggregates
         State: successful
    Start Time: 7/29/2014 20:54:41
      End Time: 7/29/2014 20:54:42
        Errors: -
   ```

3. Power up each controller module on the disaster site.

4. After nodes are booted, verify that the root aggregates are mirrored.

   If both plexes are present, resynchronization will start automatically. If one plex has failed, that plex must be destroyed and the mirror must be recreated using the `storage aggregate mirror -aggregate` *`aggregate-name`* command to reestablish the mirror relationship.

Clustered Data ONTAP 8.3 Data Protection Guide

Performing a switchback

After you heal the MetroCluster configuration you can perform the MetroCluster switchback operation. The MetroCluster switchback operation returns the configuration to its normal operating state, with the sync-source SVMs on the disaster site active and serving data from the local disk pools.

Before you begin

- The disaster cluster must have successfully switched over to the surviving cluster.
- Healing must have been performed on the data and root aggregates.
- The surviving cluster nodes must not be in HA failover state (all nodes must be up and running for each HA pair).
- The disaster site controller modules must be completely booted and not in HA takeover mode.
- The root aggregate must be mirrored.
- The Inter-Switch Links (ISLs) must be online.

Steps

1. Confirm that all nodes are in the enabled state:

   `metrocluster node show`

   Example

```
cluster_B::>  metrocluster node show
DR                              Configuration  DR
Group Cluster Node              State          Mirroring Mode
----- ------- ----------------- -------------- --------- ----
1     sti65-vsim-ucs258f8e_siteB
              sti65-vsim-ucs258e configured     enabled   normal
              sti65-vsim-ucs258f configured     enabled   normal
      sti65-vsim-ucs258g8h_siteA
              sti65-vsim-ucs258g configured     enabled   normal
              sti65-vsim-ucs258h configured     enabled   normal
4 entries were displayed.
```

2. Confirm that resynchronization is complete on all SVMs:

   `metrocluster vserver show`

3. Verify that any automatic LIF migrations being performed by the healing operations have been successfully completed: `metrocluster check lif show`
4. From any node in the MetroCluster configuration, simulate the switchback to ensure that switchback can succeed by running the `metrocluster switchback -simulate` command at the advanced privilege level.

   If the command indicates any vetoes that would prevent switchback, resolve those issues before proceeding.
5. Perform the switchback by running the `metrocluster switchback` command from any node in the surviving cluster.
6. Check the progress of the switchback operation:

   `metrocluster show`

   Example

   ```
   cluster_B::> metrocluster show
   Cluster              Configuration State    Mode
   --------------------    ------------------     ---------
    Local: cluster_B configured          switchover
   Remote: cluster_A configured          waiting-for-switchback
   ```

   The switchback operation is complete when the output displays `normal`:

   ```
   cluster_B::> metrocluster show
   Cluster              Configuration State    Mode
   --------------------    ------------------     ---------
    Local: cluster_B configured             normal
   Remote: cluster_A configured             normal
   ```

   If a switchback takes a long time to complete, you can check on the status of in-progress baselines by using the `metrocluster config-replication resync-status show` command.
7. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Verifying a successful switchback

You want to confirm that all aggregates and SVMs are switched back and online.

Steps

1. Switched over data aggregates (in the following example, aggr_b2 on node B2) are switched back:

   `aggr show`

Example

```
controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2    227.1GB   227.1GB    0% online       0 controller_B_2   raid_dp,
                                                                   mirrored,
                                                                   normal

controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2          -         -     - unknown      - controller_A_1
```

2. All sync-destination SVMs on the surviving cluster are dormant (showing an `Admin State` of `stopped`) and the sync-source SVMs on the disaster cluster are up and running.

   `vserver show -subtype sync-source`

Example

```
controller_B_1::> vserver show -subtype sync-source
                              Admin      Root                        Name    Name
Vserver     Type    Subtype   State      Volume     Aggregate        Service Mapping
----------- ------- ---------- ---------- ---------- ----------       ------- -------
...
vs1a-mc        data    sync-source
                              running    vs1a_vol   controller_B_2  file    file
                                                              aggr_b2

controller_A_1::> vserver show -subtype sync-destination
                              Admin      Root                  Name    Name
Vserver     Type    Subtype   State      Volume     Aggregate  Service Mapping
----------- ------- ---------- ---------- ---------- ---------- ------- -------
...
mcc1A-vs1a-mc  data    sync-destination
                              stopped    vs1a_vol   sosb_      file    file
                                                    aggr_b2
```

3. Confirm that the switchback operations succeeded by using the `metrocluster operation show` command.

If the command output shows...	Then...
That the switchback operation state is successful.	The switchback process is complete and you can proceed with operation of the system.
That the switchback operation or switchback-continuation-agent operation is partially successful.	Perform the suggested fix provided in the output of the `metrocluster operation show` command.

Reestablishing SnapMirror or SnapVault SVM peering relationships

After a switchover or switchback operation, you must manually reestablish any SVM peering to clusters outside of the MetroCluster configuration if the destination of the relationship is in the MetroCluster configuration.

About this task

This procedure is done at the surviving site after a switchover has occurred or at the disaster site after a switchback has occurred.

Steps

1. Check whether the SVM peering relationships have been reestablished by using the `metrocluster vserver show` command.
2. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Performing a forced switchover after a disaster

An administrator, or the MetroCluster Tiebreaker software if it is configured, must determine that a disaster has occurred and perform the MetroCluster switchover. In either case, there are steps you must perform on both the disaster cluster and the surviving cluster after the switchover to ensure safe and continued data service.

Steps

1. Fencing off the disaster site on page 34
2. Performing a forced switchover on page 35
3. Reestablishing SnapMirror or SnapVault SVM peering relationships on page 37
4. Accessing volumes in NVFAIL state after a switchover on page 37

Fencing off the disaster site

If the disaster site is not accessible or its power has been lost, you should manually fence off the disaster site because you cannot physically turn off power to the disaster site node. The fencing process restricts storage access to the cluster, preventing events that could impact the surviving site, such as RAID reconstructions if disks power up in a staggered manner.

About this task

You should only use this procedure if the disaster site components cannot be powered off. It is important to do fencing if there is any chance that the disaster site components can unexpectedly come online.

Step

1. Disable all ISLs on the switches at the surviving site.

If you are disabling ports on a...	Then...

Brocade switch	**a.** Persistently disable the first ISL port:
	`portpersistentdisable` ***`port-number`***
	b. Repeat substep a on the other ISL ports.
	c. Repeat substeps a and b for the second FC switch at the surviving site.
	In the following example, ports 14 and 15 are persistently disabled:
	`FC_switch_A_1:admin> portpersistentdisable 14`
	`FC_switch_A_1:admin> portpersistentdisable 15`
	`FC_switch_A_1:admin>`

Cisco switch	**a.** Enter configuration mode for the interface, and then disable them with the `shut` command.
	In the following example, port fc1/36 is disabled:
	`FC_switch_A_1# conf t`
	`FC_switch_A_1(config)# interface fc1/36`
	`FC_switch_A_1(config)# shut`
	`FC_switch_A_1(config-if)# end`
	`FC_switch_A_1# copy running-config startup-config`
	b. Verify that the switch port is disabled:
	`show interface brief`
	c. Repeat substeps a and b on the other ISL ports.
	d. Repeat substeps a, b, and c for the second FC switch at the surviving site.

Performing a forced switchover

The switchover process, in addition to providing nondisruptive operations during testing and maintenance, enables you to recover from a site failure with a single command.

Before you begin

Ensure that at least one of the surviving site nodes is up and running before you run the switchover.

Note: SnapMirror and SnapVault configurations will be deleted automatically.

Steps

1. Implement the switchover by running the `metrocluster switchover -forced-on-disaster true` command.

   The operation can take a period of minutes to complete.

2. Answer **y** when prompted by the system to continue with the switchover.

3. Monitor the completion of the switchover by running the `metrocluster operation show` command.

   #### Example

   ```
   mcc1A::> metrocluster operation show
     Operation: switchover
    Start time: 10/4/2012 19:04:13
         State: in-progress
     End time: -
       Errors:

   mcc1A::> metrocluster operation show
     Operation: switchover
    Start time: 10/4/2012 19:04:13
         State: successful
     End time: 10/4/2012 19:04:22
       Errors: -
   ```

   If the switchover is vetoed, you have the option of reissuing the `metrocluster switchover -forced-on-disaster true` command with the `–override-vetoes` option. If you use this optional parameter, the system overrides any soft vetoes that prevent the switchover.

After you finish

SnapMirror relationships need to be reestablished after switchover.

Clustered Data ONTAP 8.3 Data Protection Guide

Reestablishing SnapMirror or SnapVault SVM peering relationships

After a switchover or switchback operation, you must manually reestablish any SVM peering to clusters outside of the MetroCluster configuration if the destination of the relationship is in the MetroCluster configuration.

About this task

This procedure is done at the surviving site after a switchover has occurred or at the disaster site after a switchback has occurred.

Steps

1. Check whether the SVM peering relationships have been reestablished by using the `metrocluster vserver show` command.
2. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Accessing volumes in NVFAIL state after a switchover

After a switchover, you must clear the NVFAIL state by resetting the `-in-nvfailed-state` parameter of the `volume modify` command to remove the restriction of clients to access data.

Before you begin

The database or file system must not be running or trying to access the affected volume.

About this task

Setting `-in-nvfailed-state` parameter requires advanced-level privilege.

Step

1. Recover the volume by using the `volume modify` command with the `-in-nvfailed-state` parameter set to **`false`**.

After you finish

For instructions about examining database file validity, see the documentation for your specific database software.

If your database uses LUNs, review the steps to make the LUNs accessible to the host after an NVRAM failure.

Related concepts

Monitoring and protecting database validity by using NVFAIL on page 93

Recovering from the disaster

To restore the original configuration, you perform a series of steps on the MetroCluster components at the disaster site. The set of procedures you use depends on whether the system controllers failed.

Choices

- Recovering from a disaster when both controllers failed on page 39
- Recovering from a site failure when no controllers were replaced on page 73

Recovering from a disaster when both controllers failed

If the controller modules must be replaced, to recover from the disaster, you must replace the equipment and reassign ownership of disks.

Before you begin

- The disaster site must be fenced off, as described in *Fencing off the disaster site* on page 34.
- Switchover must have been performed, as described in *Performing a forced switchover* on page 35.
- Disks and the controller modules must be new and must not have been assigned ownership previously.

Replacing hardware at the disaster site

If hardware components need to be replaced, you must replace them using their individual hardware replacement and install guides.

Before you begin

The equipment must be powered off or remain halted (showing the LOADER prompt).

Step

1. Replace the components as necessary.

   Note: In this step, you replace and cable the components exactly as they were cabled prior to the disaster. You must not power up the components.

If you are replacing...	Perform these steps...	Using these guides...
FC switches	**a.** Install the new switches. **b.** Cable the ISL links. Do not power on the FC switches at this time.	*MetroCluster Service Guide*
Disk shelves	**a.** Install the disk shelves and disks. • Disk shelf stacks should be the same configuration as at the surviving site. • Disks can be the same size or larger, but must be of the same type (SAS or SATA). **b.** Cable the disk shelves to adjacent shelves within the stack and to the FC-to-SAS bridge. Do not power on the disk shelves at this time.	*SAS Disk Shelves Installation and Service Guide for DS4243, DS2246, DS4486, and DS4246*
FC-to-SAS bridges	**a.** Install the FC-to-SAS bridges. **b.** Cable the FC-to-SAS bridges to the FC switches. Do not power on the FC-to-SAS bridges at this time.	*Clustered Data ONTAP 8.3 MetroCluster Installation and Configuration Guide*

If you are replacing...	Perform these steps...	Using these guides...
Controller modules	**a.** Install the new controller modules: • The controller modules must be the same model as those being replaced. For example, FAS8080 controller modules must be replaced with FAS8080 controller modules. • The controller modules must not have previously been part of either cluster within the MetroCluster configurations or have any previously existing cluster configuration. • Ensure that all network interface cards (such as Ethernet or FC) are in the same slots used on the old controller modules. **b.** Cable the new controller modules exactly the same as the old ones. The ports connecting the controller module to the FC switches should be the same as those used prior to the disaster. Do not power on the controller modules at this time.	*Installation and Setup Instructions FAS8020 systems* *Installation and Setup Instructions FAS8040/FAS8060 Systems* *Installation and setup Instructions FAS80xx Systems with I/O Expansion Modules* *Installation and Setup Instructions 62xx Systems* *Installation and Setup Instructions 32xx Systems*

Determining the system IDs of the old controller modules

After you have replaced all hardware at the disaster site, you must determine the system IDs of the replaced storage controllers. You will need the system IDs when you reassign disks to the new controller modules.

Before you begin

All equipment at the disaster site must be powered off.

About this task

The examples in this procedure assume that:

- Site A is the disaster site.
- node_A_1 has failed and is being completely replaced.
- node_A_2 has failed and is being completely replaced.
- Site B is the surviving site.
- node_B_1 is healthy.
- node_B_2 is healthy.

The controller modules have the following original system IDs.

Node	Original system ID
node_A_1	4068741258
node_A_2	4068741260
node_B_1	4068741254
node_B_2	4068741256

Step

1. From the surviving site, display the system IDs of the nodes in the MetroCluster configuration:

   `metrocluster node show -fields node-systemid,ha-partner-systemid,dr-partner-systemid,dr-auxiliary-systemid`

   #### Example

   In this example, the old system IDs are retrieved: Node_A_1: 4068741258 and Node_A_2: 4068741260. These are the system IDs of the old controller modules. Disks owned by the old controller modules are still owned these system IDs.

```
metrocluster node show -fields node-systemid,ha-partner-systemid,dr-partner-systemid,dr-
auxiliary-systemid

dr-group-id cluster      node       node-systemid ha-partner-systemid dr-partner-systemid
dr-auxiliary-systemid
----------- ------------ ---------  ------------ ------ ------------ ------ ------------
------ --------------------
1           Cluster_A    Node_A_1   4068741258    4068741260          4068741254
4068741256
1           Cluster_A    Node_A_2   4068741260    4068741258          4068741256
4068741254
1           Cluster_B    Node_B_1   -             -                   -
-
1           Cluster_B    Node_B_2   -             -                   -
-
4 entries were displayed.
```

Netbooting the new controllers

If the new controllers have a different version of Data ONTAP than that on the surviving controller modules, you must netboot the new controller modules.

Before you begin

- You must have access to an HTTP server.
- You must have access to *NetApp Support*.

  This enables you to download the necessary system files for your platform and version of Data ONTAP that is running on it.

Steps

1. Download and extract the file used for performing the netboot of your system:

   a. Download the appropriate `netboot.tgz` file for your platform from the NetApp Support Site to a web-accessible directory.

   b. Change to the web-accessible directory.

   c. Extract the contents of the `netboot.tgz` file to the target directory by entering the following command:

      `tar -zxvf netboot.tgz`

   Your directory listing should contain the following directory:
   `netboot/`

2. Download the `image.tgz` file from the NetApp Support Site to the web-accessible directory.

   Your directory listing should contain the following file and directory:
   `image.tgz netboot/`

3. Align the end of the controller module with the opening in the chassis, and then gently push the controller module into the system.

4. Depending on your network configuration, enter one of the following commands at the LOADER prompt:

If you...	Then...
Have DHCP enabled	Enter the following command: **`ifconfig e0M -auto`**

If you...	Then...
Do not have DHCP enabled	Enter the following command: **ifconfig e0M -addr=*filer_addr* -mask=*netmask* -gw=*gateway* -dns=*dns_addr* -domain=*dns_domain*** *filer_addr* is the IP address of the storage system. *netmask* is the network mask of the storage system. *gateway* is the gateway for the storage system. *dns_addr* is the IP address of a name server on your network. *dns_domain* is the Domain Name System (DNS) domain name. If you use this optional parameter, you do not need a fully qualified domain name in the netboot server URL; you need only the server's host name. **Note:** To netboot the node when your system is running in 7-Mode Data ONTAP, use an IP address that is not the management IP address for the target. If your system is running clustered Data ONTAP, you can use the management IP address. **Note:** Other parameters might be necessary for your interface. For details, use the `help ifconfig` command at the LOADER prompt.

5. At the LOADER prompt, enter the following command:

   netboot http://*path_to_the_web-accessible_directory*/netboot/kernel

   The system begins to boot, but stops at the Boot menu.

6. Select the **Install new software first** option from the displayed menu.

   This menu option downloads and installs the new Data ONTAP image to the boot device. If you are prompted to continue the procedure, enter **y** when prompted.

7. Complete the following substeps:

   a. Enter **n** to skip the backup recovery when you see the following prompt:

```
*********************************************************
*             Restore Backup Configuration              *
*  This procedure only applies to storage controllers that   *
*  are configured as an HA pair.                          *
*                                                         *
*  Choose Yes to restore the 'varfs' backup configuration    *
*  from a TFTP server.  Refer to the Boot Device Replacement *
*  guide for more details.                                *
*  Choose No to skip the back up recovery and return to the  *
*  boot menu.                                             *
```

```
*************************************************************

Do you want to restore the backup configuration
now? {y|n} n
```

b. Reboot the node by entering **y** when you see the following prompt:

```
The node must be rebooted to start using the newly installed
software. Do you want to reboot now? {y/n} y
```

c. Boot Data ONTAP by entering the following command at the boot environment prompt:

boot_ontap

After reboot, if prompted to update firmware and BIOS, enter **y** to accept.

The controller module displays the Boot menu because the boot device was reformatted and the configuration data needs to be restored.

8. From the Boot menu, select **option 5** to enter Maintenance mode.
9. Repeat this procedure on the other controller module.

Determining the system IDs of the replacement controller modules

After you have replaced all hardware at the disaster site, you must determine the system ID of the newly installed storage controllers.

About this task

This procedure is performed with the replacement controller modules in Maintenance mode.

The examples in this procedure are written with the following assumptions:

- Site A is the disaster site.
- node_A_1 has failed and is being completely replaced.
- node_A_2 has failed and is being completely replaced.
- Site B is the surviving site.
- node_B_1 is healthy.
- node_B_2 is healthy.

The examples in this procedure use controllers with the following system IDs:

Node	Original system ID	New system ID
node_A_1	4068741258	1574774970

Node	Original system ID	New system ID
node_A_2	4068741260	1574774991
node_B_1	4068741254	unchanged
node_B_2	4068741256	unchanged

Steps

1. With the node in Maintenance mode, display the local system ID of the node by issuing the following command on each node:

   `disk show`

   #### Example

   In the following example, the new system ID is 1574774970:

   ```
   *> disk show
    Local System ID: 1574774970
    ...
   ```

2. On the second node, repeat the previous step.

   #### Example

   In the following example, the new system ID is 1574774991:

   ```
   *> disk show
    Local System ID: 1574774991
    ...
   ```

Verifying that the HA state of components is mcc

In a MetroCluster configuration, the HA state of the controller and chassis components must be set to `mcc` so they boot up properly.

Before you begin

The system must be in Maintenance mode.

About this task

This task must be performed on each node that has been replaced.

Steps

1. In Maintenance mode, display the HA state of the controller module and chassis:

ha-config show

The HA state should be `mcc` for all components.

2. If necessary, set the HA state for the controller module or chassis to `mcc` by using the applicable command:

   ha-config modify controller mcc

   ha-config modify chassis mcc

3. Repeat these steps on the other replacement node.

Verifying port configuration and setting environmental variables

You must set the environmental variables on the node and then power it off to prepare it for MetroCluster configuration.

About this task

This procedure is performed with the replacement controller modules in Maintenance mode.

The steps to check configuration of ports is needed only on systems in which FC or CNA ports are used in initiator mode.

Steps

1. Set the system to default values:

   set-defaults

2. To prevent possible problems due to the content of the mailbox disks not matching the new configuration, set the system to ignore mailbox issues: **setenv bootargs.mbx.ignore.uncertain true**

```
LOADER-A> setenv bootargs.mbx.ignore.uncertain true
LOADER-A> bye
```

3. In Maintenance mode, enter the following command to restore the FC port configuration:

   ucadmin modify -mode fc -type *initator* *adapter_name*

   If you only want to use one of a port pair in the initiator configuration, enter a precise *adapter_name*.

4. Take one of the following actions, depending on your configuration:

If the FC port configuration is...	Then...
The same for both ports	Answer **y** when prompted by the system since modifying one port in a port pair modifies the other port as well.

If the FC port configuration is...	Then...
Different	**a.** Answer **n** when prompted by the system. **b.** Enter the following command to restore the FC port configuration: **ucadmin modify -mode fc -type *initator|target* *adapter_name***

5. Exit Maintenance mode by entering the following command:

   halt

   After you issue the command, wait until the system stops at the LOADER prompt.

6. Boot the node back into Maintenance mode for the configuration changes to take effect:

   boot_ontap maint

7. Verify the values of the variables by entering the following command:

   ucadmin show

8. Exit Maintenance mode and display the LOADER prompt:

   halt

Configuring the FC-to-SAS bridges

If you replaced the FC-to-SAS bridges, you must configure them when restoring the MetroCluster configuration. The procedure is identical to the initial configuration of an FC-to-SAS bridge.

Steps

1. Power on the FC-to-SAS bridges.
2. Set the IP address on the Ethernet ports by using the `set IPAddress` *`port ipaddress`* command.

   `port` can be either **`MP1`** or **`MP2`**.

   `ipaddress` can be an IP address in the format `xxx.xxx.xxx.xxx`.

Example

In the following example, the IP address is 10.10.10.55 on Ethernet port 1:

```
Ready.
set IPAddress MP1 10.10.10.55

Ready. *
```

3. Set the IP subnet mask on the Ethernet ports by using the `set IPSubnetMask` *`port mask`* command.

   `port` can be **`MP1`** or **`MP2`**.

   `mask` can be a subnet mask in the format `xxx.xxx.xxx.xxx`.

   Example

   In the following example, the IP subnet mask is 255.255.255.0 on Ethernet port 1:

```
Ready.
set IPSubnetMask MP1 255.255.255.0

Ready. *
```

4. Set the speed on the Ethernet ports by using the `set EthernetSpeed` *`port speed`* command.

   `port` can be **`MP1`** or **`MP2`**.

   `speed` can be **`100`**, **`1000`**, or **`auto`**.

   Example

   In the following example, the Ethernet speed is set to 1000 on Ethernet port 1.

```
Ready.
set EthernetSpeed MP1 1000

Ready. *
```

5. Save the configuration by using the `saveConfiguration` command, and restart the bridge when prompted to do so.

   Saving the configuration after configuring the Ethernet ports enables you to proceed with the bridge configuration using Telnet and enables you to access the bridge using FTP to perform firmware updates.

   Example

   The following example shows the `saveConfiguration` command and the prompt to restart the bridge.

```
Ready.
SaveConfiguration
  Restart is necessary....
  Do you wish to restart (y/n) ?
Confirm with 'y'. The bridge will save and restart with the new
settings.
```

6. After the FC-to-SAS bridge reboots, log in again.
7. Set the speed on the FC ports by using the `set FCConnMode` *`port speed`* command.

   `port` can be **`1`** or **`2`**.

   `speed` can be **`2 Gb`**, **`4 Gb`**, **`8 Gb`**, or **`auto`**.

   Example

   In the following example, the port FC1 speed is set to 8 Gb.

```
Ready.
set fcdatarate 1 8Gb

Ready. *
```

8. Set the topology on the FC ports by using the `set FCConnMode` *`port mode`* command.

   `port` can be **`1`** or **`2`**.

   `mode` can be **`ptp`**, **`loop`**, **`ptp-loop`**, or **`auto`**.

   Example

   In the following example, the port FC1 topology is set to ptp.

```
Ready.
set FCConnMode 1 ptp

Ready. *
```

9. Save the configuration by using the `saveConfiguration` command, and restart the bridge when prompted to do so.

   Example

   The following example shows the `saveConfiguration` command and the prompt to restart the bridge.

```
 Ready.
 SaveConfiguration
    Restart is necessary....
    Do you wish to restart (y/n) ?
 Confirm with 'y'. The bridge will save and restart with the new
settings.
```

10. After the FC-to-SAS bridge reboots, log in again.
11. If the FC-to-SAS bridge is running firmware 1.60 or later, enable SNMP.

    Example

```
Ready.
set snmp enabled

Ready. *
saveconfiguration

Restart is necessary....
Do you wish to restart (y/n) ?

Verify with 'y' to restart the FibreBridge.
```

12. Power off the FC-to-SAS bridges.

Configuring the FC switches

If you have replaced the FC switches in the disaster site, you must configure them using the vendor-specific procedures. You must configure one switch, verify that storage access on the surviving site is not impacted, and then configure the second switch.

Choices

- Configuring a Brocade FC switch after site disaster on page 51
- Configuring a Cisco FC switch after site disaster on page 54

Configuring a Brocade FC switch after site disaster

You must use this Brocade-specific procedure to configure the replacement switch and enable the ISL ports.

About this task

The examples in this procedure are based on the following assumptions:

- Site A is the disaster site.

- FC_switch_A_1 has been replaced.
- FC_switch_A_2 has been replaced.
- Site B is the surviving site.
- FC_switch_B_1 is healthy.
- FC_switch_B_2 is healthy.

The following table shows the switch port usage:

Role	Ports
FC-VI connections	0, 3
HBA connections	1, 2, 4, 5
FC-to-SAS bridge connections	6, 7
ISL connections	10, 11

The examples show two FC-to-SAS bridges. If you have more, you must disable and subsequently enable the additional ports.

Steps

1. Boot and pre-configure the new switch:
   a. Power up the new switch and let it boot up.
   b. Check the firmware version on the switch to confirm it matches the version of the other FC switches:

      `firmwareShow`

   c. Configure the new switch as described in the *MetroCluster Installation and Configuration Guide*, skipping the steps for configuring zoning on the switch.

      Clustered Data ONTAP 8.3 MetroCluster Installation and Configuration Guide

   d. Disable the FC-VI, HBA and storage ports on the new switch, and the ports connected to the FC-SAS bridges.

      #### Example

```
FC_switch_A_1:admin> portcfgpersistentdisable 0
FC_switch_A_1:admin> portcfgpersistentdisable 1
FC_switch_A_1:admin> portcfgpersistentdisable 2
FC_switch_A_1:admin> portcfgpersistentdisable 3
FC_switch_A_1:admin> portcfgpersistentdisable 4
```

```
FC_switch_A_1:admin> portcfgpersistentdisable 5

FC_switch_A_1:admin> portcfgpersistentdisable 6
FC_switch_A_1:admin> portcfgpersistentdisable 7
```

2. Complete configuration of the new switch:

a. Enable the ISLs on the surviving site:

portcfgpersistentenable *port-number*

Example

```
FC_switch_B_1:admin> portcfgpersistentenable 10
FC_switch_B_1:admin> portcfgpersistentenable 11
```

b. Enable the ISLs on the replacement switches:

portcfgpersistentenable *port-number*

Example

```
FC_switch_A_1:admin> portcfgpersistentenable 10
FC_switch_A_1:admin> portcfgpersistentenable 11
```

c. On the replacement switch (FC_switch_A_1 in our example) verify that the ISL's are online:

switchshow

Example

```
FC_switch_A_1:admin> switchshow
switchName: FC_switch_A_1
switchType: 71.2
switchState:Online
switchMode: Native
switchRole: Principal
switchDomain:       4
switchId:   fffc03
switchWwn:  10:00:00:05:33:8c:2e:9a
zoning:             OFF
switchBeacon:       OFF

Index Port Address Media Speed State  Proto
==============================================
...
10   10    030A00 id   16G     Online  FC E-Port
10:00:00:05:33:86:89:cb "FC_switch_A_1"
11   11    030B00 id   16G     Online  FC E-Port
10:00:00:05:33:86:89:cb "FC_switch_A_1" (downstream)
...
```

3. Verify that the ports are online:

   `switchshow`

Configuring a Cisco FC switch after site disaster

You must use this Cisco-specific procedure to configure the replacement switch and enable the ISL ports.

About this task

The examples in this procedure are based on the following assumptions:

- Site A is the disaster site.
- FC_switch_A_1 has been replaced.
- FC_switch_A_2 has been replaced.
- Site B is the surviving site.
- FC_switch_B_1 is healthy.
- FC_switch_B_2 is healthy.

Steps

1. Configure the switch:

   a. Download the *Clustered Data ONTAP 8.3 MetroCluster Installation and Configuration Guide*.

   b. Follow the steps for configuring the switch in the section "Configuring the Cisco FC switches" *except* for the section "Configuring zoning on a Cisco FC switch."

   Zoning is configured later in this procedure.

2. On the healthy switch (in this example, FC_switch_B_1), enable the ISL ports.

   #### Example

   The following example shows the commands to enable the ports:

   ```
   FC_switch_B_1# conf t
   FC_switch_B_1(config)# int fc1/14-15
   FC_switch_B_1(config)# no shut
   FC_switch_B_1(config)# end
   FC_switch_B_1# copy running-config startup-config
   FC_switch_B_1#
   ```

3. Verify that the ISL ports are up by issuing the `show interface brief` command.

4. Retrieve the zoning information from the fabric.

Example

The following example shows the commands to distribute the zoning configuration:

```
FC_switch_B_1(config-zone)# zoneset distribute full vsan 10
FC_switch_B_1(config-zone)# zoneset distribute full vsan 20
FC_switch_B_1(config-zone)# end
```

FC_switch_B_1 is distributed to all other switches in the fabric for vsan 10 and vsan 20, and the zoning information is retrieved from FC_switch_A_1.

5. On the healthy switch, verify that the zoning information is properly retrieved from the partner switch:

   `show zone`

   Example

```
FC_switch_B_1# show zone
zone name FC-VI_Zone_1_10 vsan 10
  interface fc1/1 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/2 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/1 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/2 swwn 20:00:54:7f:ee:b8:24:c0

zone name STOR_Zone_1_20_25A vsan 20
  interface fc1/5 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/8 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/9 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/10 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/11 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/8 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/9 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/10 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/11 swwn 20:00:54:7f:ee:b8:24:c0

zone name STOR_Zone_1_20_25B vsan 20
  interface fc1/8 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/9 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/10 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/11 swwn 20:00:54:7f:ee:e3:86:50
  interface fc1/5 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/8 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/9 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/10 swwn 20:00:54:7f:ee:b8:24:c0
  interface fc1/11 swwn 20:00:54:7f:ee:b8:24:c0
FC_switch_B_1#
```

6. Determine the worldwide names (WWNs) of the switches in the switch fabric.

   Example

   In this example, the two switch WWNs are as follows:

- FC_switch_A_1: 20:00:54:7f:ee:b8:24:c0
- FC_switch_B_1: 20:00:54:7f:ee:c6:80:78

```
FC_switch_B_1# show wwn switch
Switch WWN is 20:00:54:7f:ee:c6:80:78
FC_switch_B_1#

FC_switch_A_1# show wwn switch
Switch WWN is 20:00:54:7f:ee:b8:24:c0
FC_switch_A_1#
```

7. Enter configuration mode for the zone and remove zone members that do not belong to the switch WWNs of the two switches:

   no member interface *interface-ide* swwn *wwn*

   Example

   In this example, the following members are not associated with the WWN of either of the switches in the fabric and must be removed:

   - zone name FC-VI_Zone_1_10 vsan 10
     - interface fc1/1 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/2 swwn 20:00:54:7f:ee:e3:86:50
   - zone name STOR_Zone_1_20_25A vsan 20
     - interface fc1/5 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/8 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/9 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/10 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/11 swwn 20:00:54:7f:ee:e3:86:50
   - zone name STOR_Zone_1_20_25B vsan 20
     - interface fc1/8 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/9 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/10 swwn 20:00:54:7f:ee:e3:86:50
     - interface fc1/11 swwn 20:00:54:7f:ee:e3:86:50

   The following example shows the removal of these interfaces:

```
 FC_switch_B_1# conf t
 FC_switch_B_1(config)# zone name FC-VI_Zone_1_10 vsan 10
 FC_switch_B_1(config-zone)# no member interface fc1/1 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/2 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# zone name STOR_Zone_1_20_25A vsan 20
 FC_switch_B_1(config-zone)# no member interface fc1/5 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/8 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/9 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/10 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/11 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# zone name STOR_Zone_1_20_25B vsan 20
 FC_switch_B_1(config-zone)# no member interface fc1/8 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/9 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/10 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# no member interface fc1/11 swwn
20:00:54:7f:ee:e3:86:50
 FC_switch_B_1(config-zone)# save running-config startup-config
 FC_switch_B_1(config-zone)# zoneset distribute full 10
 FC_switch_B_1(config-zone)# zoneset distribute full 20
 FC_switch_B_1(config-zone)# end
 FC_switch_B_1# copy running-config startup-config
```

8. Add the ports of the new switch to the zones.

Example

The following example assumes that the cabling on the replacement switch is the same as on the old switch:

```
 FC_switch_B_1# conf t
 FC_switch_B_1(config)# zone name FC-VI_Zone_1_10 vsan 10
 FC_switch_B_1(config-zone)# member interface fc1/1 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/2 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# zone name STOR_Zone_1_20_25A vsan 20
 FC_switch_B_1(config-zone)# member interface fc1/5 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/8 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/9 swwn
20:00:54:7f:ee:c6:80:78
```

```
 FC_switch_B_1(config-zone)# member interface fc1/10 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/11 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# zone name STOR_Zone_1_20_25B vsan 20
 FC_switch_B_1(config-zone)# member interface fc1/8 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/9 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/10 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# member interface fc1/11 swwn
20:00:54:7f:ee:c6:80:78
 FC_switch_B_1(config-zone)# save running-config startup-config
 FC_switch_B_1(config-zone)# zoneset distribute full 10
 FC_switch_B_1(config-zone)# zoneset distribute full 20
 FC_switch_B_1(config-zone)# end
 FC_switch_B_1# copy running-config startup-config
```

9. Verify that the zoning is properly configured:

   show zone

   Example

   The following example output shows the three zones:

```
FC_switch_B_1# show zone
  zone name FC-VI_Zone_1_10 vsan 10
    interface fc1/1 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/2 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/1 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/2 swwn 20:00:54:7f:ee:b8:24:c0

  zone name STOR_Zone_1_20_25A vsan 20
    interface fc1/5 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/8 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/9 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/10 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/11 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/8 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/9 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/10 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/11 swwn 20:00:54:7f:ee:b8:24:c0

  zone name STOR_Zone_1_20_25B vsan 20
    interface fc1/8 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/9 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/10 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/11 swwn 20:00:54:7f:ee:c6:80:78
    interface fc1/5 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/8 swwn 20:00:54:7f:ee:b8:24:c0
```

```
    interface fc1/9 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/10 swwn 20:00:54:7f:ee:b8:24:c0
    interface fc1/11 swwn 20:00:54:7f:ee:b8:24:c0
FC_switch_B_1#
```

Powering on the equipment and enabling non-ISL ports

You must power on the MetroCluster components at the disaster site when you are ready to prepare for switchback.

Before you begin

You must have already replaced and cabled the MetroCluster components exactly as the old ones.

Clustered Data ONTAP 8.3 MetroCluster Installation and Configuration Guide

About this task

The examples in this procedure assume the following:

- Site A is the disaster site.
- FC_switch_A_1 has been replaced.
- FC_switch_A_2 has been replaced.
- Site B is the surviving site.
- FC_switch_B_1 is healthy.
- FC_switch_B_2 is healthy.

Steps

1. On the surviving site, disable disk autoassignment:

   `storage disk option modify -autoassign off *`

   Example

   ```
   cluster_B::> storage disk option modify -autoassign off *
   2 entries were modified.
   ```

2. On the surviving site, confirm that disk autoassignment is off:

   `storage disk option show`

Example

```
 cluster_B::> storage disk option show
 Node     BKg. FW. Upd.  Auto Copy   Auto Assign  Auto Assign Policy
--------  -------------  -----------  -----------  ------------------
node_B_1       on            on          off             default
node_B_2       on            on          off             default
2 entries were displayed.

 cluster_B::>
```

3. Turn on the disk shelves at the disaster site and make sure that all disks are running.
4. Turn on all FC-to-SAS bridges at the disaster site.
5. If any disks were replaced, leave the controllers powered off or at the LOADER prompt.
6. Enable the non-ISL ports on the FC switches.

If the switch vendor is...	Then use these steps to enable the ports...

Brocade	**a.** Persistently enable the ports connected to the FC-to-SAS bridges:
	`portpersistentenable` ***`port-number`***
	In the following example, ports 6 and 7 are enabled:
	`FC_switch_A_1:admin> portpersistentenable 6`
	`FC_switch_A_1:admin> portpersistentenable 7`
	`FC_switch_A_1:admin>`
	b. Persistently enable the ports connected to the HBAs and FC-VI adapters:
	`portpersistentenable` ***`port-number`***
	In the following example, ports 6 and 7 are enabled:
	`FC_switch_A_1:admin> portpersistentenable 1`
	`FC_switch_A_1:admin> portpersistentenable 2`
	`FC_switch_A_1:admin> portpersistentenable 4`
	`FC_switch_A_1:admin> portpersistentenable 5`
	`FC_switch_A_1:admin>`
	c. Repeat substeps a and b for the second FC switch at the surviving site.

If the switch vendor is...	Then use these steps to enable the ports...

Cisco	**a.** Enter configuration mode for the interface, and then enable them with the `no shut` command.
	In the following example, port fc1/36 is disabled:
	`FC_switch_A_1# conf t`
	`FC_switch_A_1(config)# interface fc1/36`
	`FC_switch_A_1(config)# no shut`
	`FC_switch_A_1(config-if)# end`
	`FC_switch_A_1# copy running-config startup-config`
	b. Verify that the switch port is enabled:
	`show interface brief`
	c. Repeat substeps a and b on the other ports connected to the FC-to-SAS bridges, HBAs, and FC-VI adapters.
	d. Repeat substeps a, b, and c for the second FC switch at the surviving site.

Verifying the storage configuration

You must check that all storage is visible from the surviving nodes.

Steps

1. Confirm that all storage components at the disaster site are the same at the surviving site in quantity and type.

   The surviving site and disaster site should have the same number of FC-to-SAS bridges, disks shelf stacks and disk shelves, and disks.

2. Confirm that any disks that have been replaced at the disaster site are unowned: `run local disk show -n`

   Disks should appear as being `unowned`.

3. If no disks were replaced, confirm that all disks are present: `disk show`

Assigning ownership for replaced disks

If you replaced disks when restoring hardware at the disaster site or you had to zero disks or remove ownership, you must assign ownership to the affected disks.

Before you begin

The disaster site must have at least as many available disks as it did prior to the disaster.

About this task

These steps are performed on the cluster at the disaster site.

This procedure shows the reassignment of all disks at the disaster site.

The examples in this procedure assume that:

- Site A is the disaster site.
- node_A_1 has been replaced.
- node_A_2 has been replaced.
- Site B is the surviving site.
- node_B_1 is healthy.
- node_B_2 is healthy.

The controller modules have the following original system IDs:

Node	Original system ID
node_A_1	4068741258
node_A_2	4068741260
node_B_1	4068741254
node_B_2	4068741256

Steps

1. Assign the new, unowned disks to the appropriate disk pools using the following series of commands:

   storage disk assign -sysid *sysid* -count *disk-count* -pool *pool-number*

   a. Systematically assign the replaced disks for each node to their respective disk pools:

      disk assign -s *sysid* -n *old-count-of-disks* -p *pool*

      From the surviving site, you issue a `disk assign` command for each node:

      ```
      cluster_B::> disk assign -s node_B_1-sysid -n old-count-of-disks -
      p 0 (remote pool of surviving site)
      cluster_B::> disk assign -s node_B_2-sysid -n old-count-of-disks -
      p 0 (remote pool of surviving site)
      cluster_B::> disk assign -s node_A_1-old-sysid -n old-count-of-
      disks -p 1 (local pool of surviving site)
      cluster_B::> disk assign -s node_A_2-old-sysid -n old-count-of-
      disks -p 1 (local pool of surviving site)
      ```

Example

The following example shows the commands with the system IDs:

```
cluster_B::> disk assign -s 4068741254 -n 24 -p 0
cluster_B::> disk assign -s 4068741256 -n 24 -p 0
cluster_B::> disk assign -s 4068741258 -n 24 -p 1
cluster_B::> disk assign -s 4068741260 -n 24 -p 1
```

old-count-of-disks indicates the number of disks that will be assigned to the disk pool. This number must be at least the same number of disks for each node that were present before the disaster. If a lower number of disks is specified or present the healing operations may not complete due to insufficient space.

2. Confirm the ownership of the disks:

   storage disk show -fields owner, pool

   Example

```
storage disk show -fields owner, pool
cluster_A::> storage disk show -fields owner, pool
disk     owner          pool
-------- ------------- -----
0c.00.1  node_A_1      Pool0
0c.00.2  node_A_1      Pool0
.
.
.
0c.00.8  node_A_1      Pool1
0c.00.9  node_A_1      Pool1
.
.
.
0c.00.15 node_A_2      Pool0
0c.00.16 node_A_2      Pool0
.
.
.
0c.00.22 node_A_2      Pool1
0c.00.23 node_A_2      Pool1
.
.
.
```

3. On the surviving site, turn disk autoassignment back on:

   storage disk option modify -autoassign on *

Example

```
cluster_B::> storage disk option modify -autoassign on *
2 entries were modified.
```

4. On the surviving site, confirm that disk autoassignment is on:

   storage disk option show

   Example

```
 cluster_B::> storage disk option show
 Node     BKg. FW. Upd.  Auto Copy   Auto Assign  Auto Assign Policy
--------  -------------  -----------  -----------  ------------------
node_B_1       on            on          on             default
node_B_2       on            on          on             default
2 entries were displayed.

 cluster_B::>
```

Related concepts

Aggregate mirroring with SyncMirror on page 12

Related information

Clustered Data ONTAP 8.3 Physical Storage Management Guide

Performing aggregate healing and restoring mirrors

After replacing hardware and assigning disks, you can perform the MetroCluster healing operations. You must then confirm that aggregates are mirrored and, if necessary, restart mirroring.

Steps

1. Perform the two phases of healing (aggregate healing and root healing) on the disaster site:

   Example

```
cluster_B::> metrocluster heal -phase aggregates

cluster_B::> metrocluster heal -phase root aggregates
```

2. Monitor the healing and verify that the aggregates are in either the `resyncing` or `mirrored` state:

   storage aggregate show -node local

If the aggregate shows this state	Then...
`resyncing`	No action is required. Let the aggregate complete resyncing.
`mirror degraded`	Proceed to *step 3* on page 65.
`mirrored, normal`	No action is required.

Example

```
cluster_B::> storage aggregate show -node local

 Aggregate     Size Available Used% State   #Vols  Nodes      RAID Status
 --------- -------- --------- ----- ------- ------ ---------- ------------
 node_B_1_aggr1
           227.1GB   11.00GB   95% online       1 node_B_1   raid_dp,
                                                             resyncing
 NodeA_1_aggr2
           430.3GB   28.02GB   93% online       2 node_B_1   raid_dp,
                                                             mirror
                                                             degraded
 node_B_1_aggr3
           812.8GB   85.37GB   89% online       5 node_B_1   raid_dp,
                                                             mirrored,
                                                             normal
 3 entries were displayed.

cluster_B::>
```

In the following examples, the three aggregates are each in a different state:

Node	State
node_B_1_aggr1	`resyncing`
node_B_1_aggr2	`mirror degraded`
node_B_1_aggr3	`mirrored, normal`

3. If one or more plexes remain offline, additional steps are required to rebuild the mirror.

   In the preceding table, the mirror for node_B_1_aggr2 must be rebuilt.

   a. View details of the aggregate to identify any failed plexes:

      `storage aggregate show -r -aggregate node_B_1_aggr2`

      Example

      In the following example, plex /node_B_1_aggr2/plex0 is in a failed state:

```
cluster_B::> storage aggregate show -r -aggregate node_B_1_aggr2

 Owner Node: node_B_1
  Aggregate: node_B_1_aggr2 (online, raid_dp, mirror degraded) (block checksums)
    Plex: /node_B_1_aggr2/plex0 (offline, failed, inactive, pool0)
```

```
    RAID Group /node_B_1_aggr2/plex0/rg0 (partial)
                                                                    Usable Physical
      Position Disk                     Pool Type     RPM     Size     Size Status
      -------- ------------------------ ---- ----- ------ -------- -------- ----------

   Plex: /node_B_1_aggr2/plex1 (online, normal, active, pool1)
    RAID Group /node_B_1_aggr2/plex1/rg0 (normal, block checksums)
                                                                    Usable Physical
      Position Disk                     Pool Type     RPM     Size     Size Status
      -------- ------------------------ ---- ----- ------ -------- -------- ----------
      dparity  1.44.8                    1   SAS    15000  265.6GB  273.5GB (normal)
      parity   1.41.11                   1   SAS    15000  265.6GB  273.5GB (normal)
      data     1.42.8                    1   SAS    15000  265.6GB  273.5GB (normal)
      data     1.43.11                   1   SAS    15000  265.6GB  273.5GB (normal)
      data     1.44.9                    1   SAS    15000  265.6GB  273.5GB (normal)
      data     1.43.18                   1   SAS    15000  265.6GB  273.5GB (normal)
6 entries were displayed.

cluster_B::>
```

b. Delete the failed plex:

   storage aggregate plex delete -aggregate *aggregate-name* -plex *plex*

c. Re-establish the mirror:

   storage aggregate mirror -aggregate *aggregate-name*

d. Monitor the resynchronization and mirroring status of the plex until all mirrors are reestablished and all aggregates show `mirrored, normal` status:

   storage aggregate show

Reassigning disk ownership for root aggregates to replacement controller modules

If one or both of the controller modules or NVRAM cards were replaced at the disaster site, the system ID has changed and you must reassign disks belonging to the root aggregates to the replacement controller modules.

About this task

Because the nodes are in switchover mode and healing has been done, only the disks containing the root aggregates of the disaster site will be reassigned in this section.

The examples in this procedure were written with the following assumptions:

- Site A is the disaster site.
- node_A_1 has been replaced.
- node_A_2 has been replaced.
- Site B is the surviving site.
- node_B_1 is healthy.
- node_B_2 is healthy.

The old and new system IDs were identified in *Acquiring the new System ID* on page 41.

The examples in this procedure use controllers with the following system IDs:

Node	Original system ID	New system ID
node_A_1	4068741258	1574774970
node_A_2	4068741260	1574774991
node_B_1	4068741254	unchanged
node_B_2	4068741256	unchanged

Steps

1. With the replacement node in Maintenance mode, reassign disks:

   disk reassign -s *old-system-ID* -d *new-system-ID*

   #### Example

   ```
   *> disk reassign -s 4068741258 -d 1574774970
   ```

2. View the disks to confirm the ownership change:

   disk show

   #### Example

   The output might show more or fewer disks, depending on how many disks are in the root aggregate and whether any of these disks failed and were replaced. If the disks were replaced, then Pool0 disks will not appear in the output.

   ```
   *> disk show
   Local System ID: 1574774970

     DISK                 OWNER                    POOL   SERIAL NUMBER
   HOME                    DR HOME
   ------------           -------------            -----  -------------
   -------------           -------------
   FC_switch_A_1:6.126L19 node_A_1(1574774970)    Pool0  6SJ2M4Z20000N149N2UT
   node_A_1(1574774970)
   FC_switch_A_1:6.126L3  node_A_1(1574774970)    Pool0  6SJ2M3D00000N1500H9D
   node_A_1(1574774970)
   FC_switch_A_1:6.126L7  node_A_1(1574774970)    Pool0  6SJ2M3JA0000N1500G4F
   node_A_1(1574774970)
   FC_switch_B_1:6.126L8  node_A_1(1574774970)    Pool1  6SJ2LW1N0000N15169U9
   node_A_1(1574774970)
   FC_switch_B_1:6.126L24 node_A_1(1574774970)    Pool1  6SJ2LW270000N151906L
   node_A_1(1574774970)
   FC_switch_B_1:6.126L2  node_A_1(1574774970)    Pool1  6SJ2LVYC0000N1516BSW
   node_A_1(1574774970)

   *> aggr status
            Aggr State           Status
   ```

```
 node_A_1_root online          raid_dp, aggr
                               mirror degraded
                               64-bit
*>
```

3. View the aggregate status:

   `aggr status`

   Example

   The output might show more or fewer disks, depending on how many disks are in the root aggregate and whether any of these disks failed and were replaced. If disks were replaced, then Pool0 disks will not appear in output.

```
*> aggr status
          Aggr State           Status
  node_A_1_root online          raid_dp, aggr
                                mirror degraded
                                64-bit
*>
```

4. Delete the contents of the mailbox disks:

   `mailbox destroy local`

5. Destroy the failed plex:

If the aggregate state is...	Then...
`mirrored`	Proceed to the next step.
`mirror degraded` (or any other non-mirrored state)	**a.** Determine which plex failed: **`aggr status -r`** **b.** Delete the failed plex: **`aggr delete`** ***`failedPlexName`***

6. If the aggregate is not online, bring it online:

   `aggr online` ***`aggr_name`***

7. Halt the node to display the LOADER prompt:

   `halt`

Booting the new controller modules

After aggregate healing has been completed for both the data and root aggregates, you must boot the nodes at the disaster site.

About this task

This task begins with the nodes showing the LOADER prompt.

Steps

1. Display the boot menu:

   `boot_ontap menu`

2. From the boot menu, select option 6, **Update flash from backup config**.

   You can confirm the selection if prompted.

   Wait until the node has fully booted before proceeding.

3. Refresh the MetroCluster configuration:

   a. Enter advanced privilege mode:

      `set -privilege advanced`

   b. Refresh the configuration:

      `metrocluster configure -refresh true`

   c. Return to admin privilege mode:

      `set -privilege admin`

4. Repeat the previous steps on the other node at the disaster site.

Performing a switchback

After you heal the MetroCluster configuration you can perform the MetroCluster switchback operation. The MetroCluster switchback operation returns the configuration to its normal operating state, with the sync-source SVMs on the disaster site active and serving data from the local disk pools.

Before you begin

- The disaster cluster must have successfully switched over to the surviving cluster.
- Healing must have been performed on the data and root aggregates.
- The surviving cluster nodes must not be in HA failover state (all nodes must be up and running for each HA pair).

- The disaster site controller modules must be completely booted and not in HA takeover mode.
- The root aggregate must be mirrored.
- The Inter-Switch Links (ISLs) must be online.

Steps

1. Confirm that all nodes are in the enabled state:

   `metrocluster node show`

   Example

   ```
   cluster_B::>  metrocluster node show
   DR                              Configuration  DR
   Group Cluster Node              State          Mirroring Mode
   ----- ------- ----------------- -------------- --------- ----
   1     sti65-vsim-ucs258f8e_siteB
                 sti65-vsim-ucs258e configured     enabled   normal
                 sti65-vsim-ucs258f configured     enabled   normal
         sti65-vsim-ucs258g8h_siteA
                 sti65-vsim-ucs258g configured     enabled   normal
                 sti65-vsim-ucs258h configured     enabled   normal
   4 entries were displayed.
   ```

2. Confirm that resynchronization is complete on all SVMs:

   `metrocluster vserver show`

3. Verify that any automatic LIF migrations being performed by the healing operations have been successfully completed: `metrocluster check lif show`

4. From any node in the MetroCluster configuration, simulate the switchback to ensure that switchback can succeed by running the `metrocluster switchback -simulate` command at the advanced privilege level.

   If the command indicates any vetoes that would prevent switchback, resolve those issues before proceeding.

5. Perform the switchback by running the `metrocluster switchback` command from any node in the surviving cluster.

6. Check the progress of the switchback operation:

   `metrocluster show`

Example

```
cluster_B::> metrocluster show
Cluster              Configuration State    Mode
-------------------  -------------------    ---------
 Local: cluster_B configured           switchover
Remote: cluster_A configured           waiting-for-switchback
```

The switchback operation is complete when the output displays `normal`:

```
cluster_B::> metrocluster show
Cluster              Configuration State    Mode
-------------------  -------------------    ---------
 Local: cluster_B configured              normal
Remote: cluster_A configured              normal
```

If a switchback takes a long time to complete, you can check on the status of in-progress baselines by using the `metrocluster config-replication resync-status show` command.

7. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Verifying a successful switchback

You want to confirm that all aggregates and SVMs are switched back and online.

Steps

1. Switched over data aggregates (in the following example, aggr_b2 on node B2) are switched back:

   `aggr show`

   Example

```
controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2    227.1GB   227.1GB    0% online       0 controller_B_2   raid_dp,
                                                                   mirrored,
                                                                   normal

controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2          -         -     - unknown      - controller_A_1
```

2. All sync-destination SVMs on the surviving cluster are dormant (showing an `Admin State` of `stopped`) and the sync-source SVMs on the disaster cluster are up and running.

   `vserver show -subtype sync-source`

Example

```
controller_B_1::> vserver show -subtype sync-source
                                Admin      Root                      Name    Name
Vserver     Type    Subtype     State      Volume     Aggregate      Service Mapping
----------- ------- ---------- ---------- ---------- ----------     ------- -------
...
vs1a-mc        data    sync-source
                                running    vs1a_vol   controller_B_2  file    file
                                                      aggr_b2

controller_A_1::> vserver show -subtype sync-destination
                                Admin      Root                   Name    Name
Vserver     Type    Subtype     State      Volume     Aggregate  Service Mapping
----------- ------- ---------- ---------- ---------- ---------- ------- -------
...
mcc1A-vs1a-mc  data    sync-destination
                                stopped    vs1a_vol   sosb_      file    file
                                                      aggr_b2
```

3. Confirm that the switchback operations succeeded by using the `metrocluster operation show` command.

If the command output shows...	Then...
That the switchback operation state is successful.	The switchback process is complete and you can proceed with operation of the system.
That the switchback operation or switchback-continuation-agent operation is partially successful.	Perform the suggested fix provided in the output of the `metrocluster operation show` command.

Verifying the health of the MetroCluster configuration

You should check the health of the MetroCluster to ensure proper operation.

Steps

1. Check that the MetroCluster is configured and in normal mode on each cluster:

   `metrocluster show`

   Example

```
cluster_A::> metrocluster show
Cluster                       Configuration State    Mode
----------------------------- ---------------------- ------------------------
 Local: cluster_A             configured             normal
Remote: cluster_B             configured             normal
```

2. Check that mirroring is enabled on each node:

   `metrocluster node show`

Example

```
cluster_A::> metrocluster node show
DR                           Configuration  DR
Group Cluster Node           State          Mirroring Mode
----- ------- -------------- -------------- --------- --------------------
1     cluster_A
              node_A_1       configured     enabled   normal
              node_A_2       configured     enabled   normal
      cluster_B
              node_B_1       configured     enabled   normal
              node_B_2       configured     enabled   normal
4 entries were displayed.
```

3. Check that the MetroCluster components are healthy:

   `metrocluster check run`

   Example

```
cluster_A::> metrocluster check run

Last Checked On: 10/1/2014 16:03:37

Component           Result
------------------- ---------
nodes               ok
lifs                ok
config-replication  ok
aggregates          ok
4 entries were displayed.

Command completed. Use the "metrocluster check show -instance"
command or sub-commands in "metrocluster check" directory for
detailed results.
To check if the nodes are ready to do a switchover or switchback
operation, run "metrocluster switchover -simulate" or "metrocluster
switchback -simulate", respectively.
```

4. Check that there are no health alerts:

   `system health alert show`

5. Confirm that a switchover is possible:

   `metrocluster switchover -simulate`

   The `-simulate` option is executed at the advanced privilege level.

Recovering from a site failure when no controllers were replaced

After the equipment at the disaster site has undergone any required maintenance or replacement, but neither of the controllers were replaced, you can begin the process of returning the MetroCluster

configuration to a fully redundant state. This includes healing the configuration (first the data aggregates and then the root aggregates) and performing the switchback operation.

Before you begin

- All MetroCluster hardware in the disaster cluster must be functional.
- The overall MetroCluster configuration must be in switchover.
- The ISL must be up and operating between the MetroCluster sites.

About this task

Steps

1. Healing the configuration on page 74
2. Verifying that your system is ready for a switchback on page 77
3. Performing a switchback on page 78
4. Verifying a successful switchback on page 80
5. Deleting stale aggregate listings after switchback on page 81

Healing the configuration

Following a switchover, you must perform a healing operation in a specific order to restore MetroCluster functionality.

Before you begin

- ISLs must be up and operating.
- Switchover must have been performed and the surviving site must be serving data.
- Nodes in the surviving site must not be in HA failover state (all nodes must be up and running for each HA pair).
- Nodes on the disaster site must be halted or remain powered off.
  They must not be fully booted during the healing process.
- Storage at the disaster site must be accessible (shelves are powered up, functional, and accessible).

About this task

The healing operation must first be performed on the data aggregates, and then on the root aggregates.

Steps

1. Healing the data aggregates on page 75
2. Healing the root aggregates after a disaster on page 76

Healing the data aggregates

You must heal the data aggregates after repairing and replacing any hardware on the disaster site. This process resynchronizes the data aggregates and prepares the (now repaired) disaster site for normal operation. You must heal the data aggregates prior to healing the root aggregates.

About this task

The following example shows a forced switchover, where you bring the switched-over aggregate online. All configuration updates in the remote cluster successfully replicate to the local cluster. You power up the storage on the disaster site as part of this procedure, but you do not and must not power up the controller modules on the disaster site.

Steps

1. Ensure that switchover has completed by running the `metrocluster operation show` command.

   Example

   ```
   controller_A_1::> metrocluster operation show
     Operation: switchover
         State: successful
    Start Time: 7/25/2014 20:01:48
      End Time: 7/25/2014 20:02:14
        Errors: -
   ```

2. Resynchronize the data aggregates by running the `metrocluster heal -phase aggregates` command from the surviving cluster.

   Example

   ```
   controller_A_1::> metrocluster heal -phase aggregates
   [Job 130] Job succeeded: Heal Aggregates is successful.
   ```

   If the healing is vetoed, you have the option of reissuing the `metrocluster heal` command with the `–override-vetoes` parameter. If you use this optional parameter, the system overrides any soft vetoes that prevent the healing operation.

3. Verify that the operation has been completed by running the `metrocluster operation show` command.

Example

```
controller_A_1::> metrocluster operation show
    Operation: heal-aggregates
      State: successful
Start Time: 7/25/2014 18:45:55
   End Time: 7/25/2014 18:45:56
     Errors: -
```

4. Check the state of the aggregates by running the `storage aggregate show` command.

Example

```
controller_A_1::> storage aggregate show
Aggregate     Size Available Used% State   #Vols  Nodes
RAID Status
--------- -------- --------- ----- ------- ------ ----------------
------------
...
aggr_b2    227.1GB   227.1GB    0% online       0 mcc1-a2
raid_dp, mirrored, normal...
```

5. If storage has been replaced at the disaster site, you might need to remirror the aggregates.

   Clustered Data ONTAP 8.3 Data Protection Guide

Healing the root aggregates after a disaster

After the data aggregates have been healed, you must heal the root aggregates in preparation for the switchback operation.

Before you begin

The data aggregates phase of the MetroCluster healing process must have been completed successfully.

Steps

1. Switch back the mirrored aggregates by running the `metrocluster heal -phase root-aggregates` command.

Example

```
mcc1A::> metrocluster heal -phase root-aggregates
[Job 137] Job succeeded: Heal Root Aggregates is successful
```

If the healing is vetoed, you have the option of reissuing the `metrocluster heal` command with the `–override-vetoes` parameter. If you use this optional parameter, the system overrides any soft vetoes that prevent the healing operation.

2. Ensure that the heal operation is complete by running the `metrocluster operation show` command on the destination cluster:

Example

```
mcc1A::> metrocluster operation show
  Operation: heal-root-aggregates
      State: successful
 Start Time: 7/29/2014 20:54:41
   End Time: 7/29/2014 20:54:42
     Errors: -
```

3. Power up each controller module on the disaster site.
4. After nodes are booted, verify that the root aggregates are mirrored.

   If both plexes are present, any resynchronization will start automatically. If one plex has failed, that plex must be destroyed and the mirror recreated using the `storage aggregate mirror -aggregate` *`aggregate-name`* command to reestablish the mirror relationship.

   Clustered Data ONTAP 8.3 Data Protection Guide

Verifying that your system is ready for a switchback

If your system is already in the switchover state, you can use the `-simulate` option to preview the results of a switchback operation.

Steps

1. Simulate the switchback operation:
   a. From either surviving node's prompt, change to the advanced privilege level:

      `set -privilege advanced`

      You need to respond with **`y`** when prompted to continue into advanced mode and see the advanced mode prompt (`*>`).

   b. Perform the switchback operation with the `-simulate` parameter:

      `metrocluster switchback -simulate`

   c. Return to the admin privilege level:

      `set -privilege admin`

2. Review the output that is returned.

   The output shows whether the switchback operation would run into errors.

Example: Verification results

The following example shows a successful verification of a switchback operation:

```
cluster4::*> metrocluster switchback -simulate
  (metrocluster switchback)
[Job 130] Setting up the nodes and cluster components for the
switchback operation...DBG:backup_api.c:
327:backup_nso_sb_vetocheck : MCC Switch Back
[Job 130] Job succeeded: Switchback simulation is
successful.

cluster4::*> metrocluster op show
  (metrocluster operation show)
  Operation: switchback-simulate
      State: successful
 Start Time: 5/15/2014 16:14:34
   End Time: 5/15/2014 16:15:04
     Errors: -

cluster4::*> job show -name Me*
                            Owning
Job ID Name                 Vserver    Node           State
------ -------------------- ---------- -------------- ----------
130    MetroCluster Switchback
                            cluster4
                                       cluster4-01
                                                      Success
       Description: MetroCluster Switchback Job - Simulation
```

Performing a switchback

After you heal the MetroCluster configuration you can perform the MetroCluster switchback operation. The MetroCluster switchback operation returns the configuration to its normal operating state, with the sync-source SVMs on the disaster site active and serving data from the local disk pools.

Before you begin

- The disaster cluster must have successfully switched over to the surviving cluster.
- Healing must have been performed on the data and root aggregates.
- The surviving cluster nodes must not be in HA failover state (all nodes must be up and running for each HA pair).
- The disaster site controller modules must be completely booted and not in HA takeover mode.

- The root aggregate must be mirrored.
- The Inter-Switch Links (ISLs) must be online.

Steps

1. Confirm that all nodes are in the enabled state:

   `metrocluster node show`

   #### Example

   ```
   cluster_B::>  metrocluster node show
   DR                               Configuration  DR
   Group Cluster Node               State          Mirroring Mode
   ----- ------- ------------------ -------------- --------- ----
   1     sti65-vsim-ucs258f8e_siteB
                 sti65-vsim-ucs258e configured     enabled   normal
                 sti65-vsim-ucs258f configured     enabled   normal
         sti65-vsim-ucs258g8h_siteA
                 sti65-vsim-ucs258g configured     enabled   normal
                 sti65-vsim-ucs258h configured     enabled   normal
   4 entries were displayed.
   ```

2. Confirm that resynchronization is complete on all SVMs:

   `metrocluster vserver show`

3. Verify that any automatic LIF migrations being performed by the healing operations have been successfully completed: `metrocluster check lif show`

4. From any node in the MetroCluster configuration, simulate the switchback to ensure that switchback can succeed by running the `metrocluster switchback -simulate` command at the advanced privilege level.

   If the command indicates any vetoes that would prevent switchback, resolve those issues before proceeding.

5. Perform the switchback by running the `metrocluster switchback` command from any node in the surviving cluster.

6. Check the progress of the switchback operation:

   `metrocluster show`

Example

```
cluster_B::> metrocluster show
Cluster              Configuration State    Mode
-------------------  -------------------    ---------
 Local: cluster_B configured           switchover
Remote: cluster_A configured           waiting-for-switchback
```

The switchback operation is complete when the output displays `normal`:

```
cluster_B::> metrocluster show
Cluster              Configuration State    Mode
-------------------  -------------------    ---------
 Local: cluster_B configured              normal
Remote: cluster_A configured              normal
```

If a switchback takes a long time to complete, you can check on the status of in-progress baselines by using the `metrocluster config-replication resync-status show` command.

7. Reestablish any SnapMirror or SnapVault configurations.

   Clustered Data ONTAP 8.3 Data Protection Guide

Verifying a successful switchback

You want to confirm that all aggregates and SVMs are switched back and online.

Steps

1. Switched over data aggregates (in the following example, aggr_b2 on node B2) are switched back:

   `aggr show`

   Example

```
controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2    227.1GB   227.1GB    0% online       0 controller_B_2   raid_dp,
                                                                   mirrored,
                                                                   normal

controller_B_1::> aggr show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
...
aggr_b2          -         -     - unknown      - controller_A_1
```

2. All sync-destination SVMs on the surviving cluster are dormant (showing an `Admin State` of `stopped`) and the sync-source SVMs on the disaster cluster are up and running.

   `vserver show -subtype sync-source`

Example

```
controller_B_1::> vserver show -subtype sync-source
                                 Admin      Root                      Name    Name
Vserver     Type    Subtype      State      Volume     Aggregate      Service Mapping
----------- ------- ---------- ---------- ---------- ----------      ------- -------
...
vs1a-mc        data    sync-source
                                 running    vs1a_vol   controller_B_2  file    file
                                                                    aggr_b2

controller_A_1::> vserver show -subtype sync-destination
                                 Admin      Root                   Name    Name
Vserver     Type    Subtype      State      Volume     Aggregate   Service Mapping
----------- ------- ---------- ---------- ---------- ----------  ------- -------
...
mcc1A-vs1a-mc  data    sync-destination
                                 stopped    vs1a_vol   sosb_       file    file
                                                                aggr_b2
```

3. Confirm that the switchback operations succeeded by using the `metrocluster operation show` command.

If the command output shows...	Then...
That the switchback operation state is successful.	The switchback process is complete and you can proceed with operation of the system.
That the switchback operation or switchback-continuation-agent operation is partially successful.	Perform the suggested fix provided in the output of the `metrocluster operation show` command.

Deleting stale aggregate listings after switchback

In some circumstances after switchback, you might notice the presence of *stale* aggregates . Stale aggregates are aggregates that have been removed from Data ONTAP, but whose information remains recorded on disk. Stale aggregates are displayed in the nodeshell `aggr status -r` command but not in the `storage aggregate show` command. You can delete these records so that they no longer appear.

About this task

Stale aggregates can occur if you relocated aggregates while the MetroCluster configuration was in switchover. For example:

1. Site A switches over to Site B.
2. You delete the mirroring for an aggregate and relocate the aggregate from node_B_1 to node_B_2 for load balancing.
3. You perform aggregate healing.

At this point a stale aggregate appears on node_B_1, even though the actual aggregate has been deleted from that node. This aggregate appears in the output from the nodeshell `aggr status -r` command. It does not appear in the output of the `storage aggregate show` command.

Steps

1. Compare the output of the output of the `storage aggregate show` command and the nodeshell `aggr status -r` command:

   `storage aggregate show`

   `run local aggr status -r`

   Stale aggregates appear in the `run local aggr status -r` output but not in the `storage aggregate show` output. For example, the following aggregate might appear in the `run local aggr status -r` output:

```
Aggregate aggr05 (failed, raid_dp, partial) (block checksums)
  Plex /aggr05/plex0 (offline, failed, inactive)
    RAID group /myaggr/plex0/rg0 (partial, block checksums)

      RAID Disk Device  HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/
blks)    Phys (MB/blks)
      --------- ------  ------------ ---- ---- ---- -----
--------------    --------------
      dparity   FAILED          N/A                        82/ -
      parity    0b.5    0b    -   -   SA:A   0 VMDISK   N/A
82/169472         88/182040
      data      FAILED          N/A                        82/ -
      data      FAILED          N/A                        82/ -
      data      FAILED          N/A                        82/ -
      data      FAILED          N/A                        82/ -
      data      FAILED          N/A                        82/ -
      data      FAILED          N/A                        82/ -
      Raid group is missing 7 disks.
```

2. Remove the stale aggregate:

   a. From either node's prompt, change to the advanced privilege level:

      `set -privilege advanced`

      You need to respond with **y** when prompted to continue into advanced mode and see the advanced mode prompt (`*>`).

   b. Remove the stale aggregate:

      `aggregate remove-stale-record -aggregate` ***`aggregate_name`***

   c. Return to the admin privilege level:

      `set -privilege admin`

3. Confirm that the stale aggregate record was removed:

```
run local aggr status -r
```

Commands for switchover, healing, and switchback

There are specific Data ONTAP commands for performing the MetroCluster disaster recovery processes.

If you want to...	Use this command...
Verify that switchover can be performed without errors or vetos.	`metrocluster switchover -simulate` at the advanced privilege level *Clustered Data ONTAP 8.3 man page: metrocluster switchover - Switch over storage and client access*
Verify that switchback can be performed without errors or vetos.	`metrocluster switchback -simulate` at the advanced privilege level *Clustered Data ONTAP 8.3 man page: metrocluster switchback - Switch back storage and client access*
Switch over to the partner nodes (negotiated switchover).	`metrocluster switchover` *Clustered Data ONTAP 8.3 man page: metrocluster switchover - Switch over storage and client access*
Switch over to the partner nodes (forced switchover).	`metrocluster switchover -forced-on-disaster` **`true`** *Clustered Data ONTAP 8.3 man page: metrocluster switchover - Switch over storage and client access*
Perform data aggregate healing.	`metrocluster heal -phase` **`aggregates`** *Clustered Data ONTAP 8.3 man page: metrocluster heal - Heal DR data aggregates and DR root aggregates*
Perform root aggregate healing.	`metrocluster heal -phase` **`root-aggregates`** *Clustered Data ONTAP 8.3 man page: metrocluster heal - Heal DR data aggregates and DR root aggregates*

If you want to...	Use this command...
Switch back to the home nodes.	`metrocluster switchback` *Clustered Data ONTAP 8.3 man page: metrocluster switchback - Switch back storage and client access*

Monitoring the MetroCluster configuration

You can use Data ONTAP MetroCluster commands, OnCommand Unified Manager, and OnCommand Performance Manager to monitor the health of a variety of software components and the state of MetroCluster operations.

Configuring MetroCluster components for health monitoring

You must perform some special configuration steps before monitoring the components in the MetroCluster configuration.

Configuring the MetroCluster FC switches for health monitoring

You must perform some special configuration steps to monitor the FC switches in the MetroCluster configuration.

Steps

1. Issue the following command on each MetroCluster node:

   storage switch add -switch-ipaddress *ipaddress*

   This command must be repeated on all four switches in the MetroCluster configuration.

   #### Example

   The following example shows the command to add a switch with IP 10.10.10.10:

   ```
   controller_A_1::> storage switch add -switch-ipaddress 10.10.10.10
   ```

2. Verify that all switches are properly configured:

   storage switch show

   It may take up to 15 minutes to reflect all data due to the 15-minute polling interval.

   #### Example

   The following example shows the command given to verify the MetroCluster's FC switches are configured:

   ```
   controller_A_1::> storage switch show
   Fabric           Switch Name     Vendor  Model        Switch WWN       Status
   ---------------- --------------- ------- ------------ ---------------- ------
   1000000533a9e7a6 brcd6505-fcs40  Brocade Brocade6505  1000000533a9e7a6 OK
   1000000533a9e7a6 brcd6505-fcs42  Brocade Brocade6505  1000000533d3660a OK
   1000000533ed94d1 brcd6510-fcs44  Brocade Brocade6510  1000000533eda031 OK
   ```

```
1000000533ed94d1 brcd6510-fcs45  Brocade Brocade6510  1000000533ed94d1 OK
4 entries were displayed.

controller_A_1::>
```

If the switch's worldwide name (WWN) is shown, the Data ONTAP health monitor is able to contact and monitor the FC switch.

Configuring FC-to-SAS bridges for health monitoring

You must perform some special configuration steps to monitor the FC-to-SAS bridges in the MetroCluster configuration.

Steps

1. Configure the FC-to-SAS bridges for monitoring by issuing the following command for each bridge on each storage controller:

   storage bridge add -address *ipaddress*

   This command must be repeated for all FC-to-SAS bridges in the MetroCluster configuration.

   #### Example

   The following example shows the command you must use to add an FC-to-SAS bridge with an IP address of 10.10.20.10:

   ```
   controller_A_1::> storage bridge add -address 10.10.20.10
   ```

2. Verify that all FC-to-SAS bridges are properly configured:

   storage bridge show

   It might take as long as 15 minutes to reflect all data due to the polling interval.

   #### Example

   The following example shows that the FC-to-SAS bridges are configured:

   ```
   controller_A_1::> storage bridge show

   Bridge  Symbolic Name Vendor Model             Bridge WWN  Monitored Status
   ------- ------------- ------ ---------         ----------- --------- -------
   ATTO_1  atto6500n-1   Atto   FibreBridge 6500N wwn         true      ok
   ATTO_2  atto6500n-2   Atto   FibreBridge 6500N wwn         true      ok
   ATTO_3  atto6500n-3   Atto   FibreBridge 6500N wwn         true      ok
   ATTO_4  atto6500n-4   Atto   FibreBridge 6500N wwn         true      ok

    controller_A_1::>
   ```

   If the FC-to-SAS bridge's worldwide name (WWN) is shown, the Data ONTAP health monitor is able to contact and monitor the bridge.

Detecting failures with NetApp MetroCluster Tiebreaker software

The Tiebreaker software resides on a Linux host. You need the Tiebreaker software only if you want to monitor two clusters and the connectivity status between them from a third site. Doing so enables each partner in a cluster to distinguish between an ISL failure, when inter-site links are down, from a site failure.

After you install the Tiebreaker software on a Linux host, you can configure the clusters in a MetroCluster configuration to monitor for disaster conditions.

How the Tiebreaker software detects ISL failures

In addition to monitoring the nodes in a MetroCluster configuration, the HA pair, and the cluster, the NetApp MetroCluster Tiebreaker software monitors the network paths between the MetroCluster clusters. A loss of connection between a node and DR partner triggers a health check of the intercluster peering network to detect an ISL failure.

Types of network paths

There are two types of network paths between the two clusters in a MetroCluster configuration:

FC network

This type of network is composed of two, redundant FC switch fabrics. Each switch fabric has two FC switches, and one switch of each switch fabric is co-located with a cluster. Each cluster has two FC switches, one from each switch fabric. Each switch fabric includes interswitch links (ISLs) of up to 200 km that connect the clusters. All the nodes have FC (FC-VI and FCP initiator) connectivity to each of the co-located FC switches. Data is replicated from cluster to cluster over the ISL.

Intercluster peering network

This type of network is composed of a redundant IP network path between the two clusters. Each node has redundant FC-VI connectivity with its nodes on the DR partner cluster.

Monitoring the intercluster peering network

The Tiebreaker software regularly pings the intercluster peering network to monitor its health. When the FC-VI connection is lost between the node and its DR partner cluster and the intercluster peering network does not respond to pings, the Tiebreaker software assumes that the network is isolated. Isolation occurs when the nodes are reachable but the inter-site links are down.

How the Tiebreaker software detects site failures

NetApp MetroCluster Tiebreaker software looks at reachability of all four nodes in a MetroCluster configuration, the HA pair, and the cluster to determine whether a site failure has occurred.

Monitoring the nodes

The Tiebreaker software monitors each controller in the MetroCluster nodes by establishing redundant SSH sessions through multiple paths to a node management LIF (hosted on the IP network) and communicating with the nodes.

Upon loss of all the SSH connections and a subsequent failure to reconnect, the Tiebreaker software checks the status of the HA partner. The HA partner can be in one of the following states: *taken over* because it has failed, panicked, rebooted, or shut down; *normal*, or *unreachable* over the node management LIF.

The Tiebreaker software can detect an SSH session failure in three to five seconds by setting the TCP keep-alive timer to one second and the retry interval to three seconds.

Monitoring the HA pairs

The Tiebreaker software monitors the status of the HA pairs by determining whether the nodes in the HA pairs are reachable. Upon loss of connectivity to a node, the Tiebreaker software checks the status of the HA pair. The HA pair can be in one of the following states: *taken over*, *normal*, or *unreachable*.

Monitoring the clusters

The Tiebreaker software monitors each controller in the MetroCluster nodes by establishing redundant SSH sessions through multiple paths to a cluster LIF (hosted on the IP network). If the Tiebreaker software loses the SSH connection to the cluster, it checks the status of the HA pairs, which in turn report on the aggregate status of the nodes in the cluster.

The Tiebreaker software can indirectly monitor the cluster by checking the status of its DR partner cluster. If the DR partner cluster reports that all node links are down, there is a strong possibility of a site failure at the cluster. The DR partner cluster might also report that the nodes are reachable and are considered to be healthy, but the nodes are isolated from the Tiebreaker software from all the paths (also called *ISL failures*).

The Tiebreaker software can detect an SSH session failure in three to five seconds by setting the TCP keep-alive timer to one second and the retry interval to three seconds.

Checking the MetroCluster configuration

You can check that the components and relationships in the MetroCluster configuration are working correctly. You should do a check after initial configuration and after making any changes to the

MetroCluster configuration. After you run the `metrocluster check run` command, you then display the results of the check with the `metrocluster check show` command.

About this task

If the `metrocluster check run` command is issued twice within a short time, on either or both clusters, a conflict can occur and the command might not collect all data. Subsequent `metrocluster check show` commands will not show the expected output.

Steps

1. Check the configuration:

   `metrocluster check run`

   #### Example

   The following example shows the output for a healthy MetroCluster configuration:

   ```
   controller_A_1::> metrocluster check run

   Last Checked On: 9/24/2014 17:10:33

   Component           Result
   ------------------- ---------
   nodes               ok
   lifs                ok
   config-replication  ok
   aggregates          ok
   4 entries were displayed.

   Command completed. Use the "metrocluster check show -instance"
   command or sub-commands in "metrocluster check" directory for
   detailed results.
   To check if the nodes are ready to do a switchover or switchback
   operation, run "metrocluster switchover -simulate" or "metrocluster
   switchback
   -simulate", respectively.
   ```

2. Display more detailed results from the most recent `metrocluster check run` command:

   `metrocluster check aggregate show`

   `metrocluster check config-replication show`

   `metrocluster check lif show`

   `metrocluster check node show`

   The `metrocluster check show` commands show the results of the most recent `metrocluster check run` command. You should always run the `metrocluster check`

`run` command prior to using the `metrocluster check show` commands to ensure that the information displayed is current.

Example

The following example shows the `metrocluster check aggregate show` output for a healthy MetroCluster configuration:

```
controller_A_1::> metrocluster check aggregate show

Last Checked On: 8/5/2014 00:42:58

Node                  Aggregate             Check                 Result
--------------------- --------------------- --------------------- ---------
controller_A_1     aggr0_controller_A_1_0
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
                      controller_A_1_aggr1
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
controller_A_2     aggr0_controller_A_2
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
                      controller_A_2_aggr1
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
controller_B_1     aggr0_controller_B_1_0
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
                      controller_B_1_aggr1
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
controller_B_2     aggr0_controller_B_2
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
                   controller_B_2_aggr1
                                             mirroring-status      ok
                                             disk-pool-allocation  ok
16 entries were displayed.
```

Commands for checking and monitoring the MetroCluster configuration

There are specific Data ONTAP commands for monitoring the MetroCluster configuration and checking MetroCluster operations.

Commands for checking MetroCluster operations

If you want to...	Use this command...
Perform a check of the MetroCluster operations. **Note:** This command should not be used as the only command for pre-DR operation system validation.	*Clustered Data ONTAP 8.3 man page: metrocluster check run - Check the MetroCluster setup*

If you want to...	Use this command...
View the results of the last check on MetroCluster operations.	*Clustered Data ONTAP 8.3 man page: metrocluster show - Display MetroCluster configuration information*
View results of check on configuration replication between the sites.	*Clustered Data ONTAP 8.3 man page: metrocluster check config-replication show - Display MetroCluster config-replication status information* *Clustered Data ONTAP 8.3 man page: metrocluster check config-replication show-aggregate-eligibility -*
View results of check on node configuration.	*Clustered Data ONTAP 8.3 man page: metrocluster check node show - Show results of MetroCluster check for nodes*
View results of check on aggregate configuration.	*Clustered Data ONTAP 8.3 man page: metrocluster check aggregate show - Show results of MetroCluster check for aggregates*
View the LIF placement failures in the MetroCluster configuration.	*Clustered Data ONTAP 8.3 man page: metrocluster check lif show - Show LIF placement failures in MetroCluster configuration*

Commands for monitoring the MetroCluster interconnect

If you want to...	Use this command...
Display the HA and DR mirroring status and information for the MetroCluster nodes in the cluster.	*Clustered Data ONTAP 8.3 man page: metrocluster interconnect mirror show - Display MetroCluster interconnect mirror information*

Commands for monitoring MetroCluster SVMs

If you want to...	Use this command...
View all SVMs in both sites in the MetroCluster configuration.	*Clustered Data ONTAP 8.3 man page: metrocluster vserver show - Display MetroCluster Vserver relationships*

Monitoring and protecting database validity by using NVFAIL

The `-nvfail` parameter of the `volume modify` command enables Data ONTAP to detect nonvolatile RAM (NVRAM) inconsistencies when the system is booting or after a switchover operation. It also warns you and protects the system against data access and modification until the volume can be manually recovered.

If Data ONTAP finds any problems, database or file system instances stop responding or shut down, and Data ONTAP sends error messages to the console to alert you to check the state of the database or file system. You can enable NVFAIL to warn database administrators of NVRAM inconsistencies among clustered nodes that can compromise database validity. After a system crash or switchover operation, NFS clients cannot access data from any of the nodes until the NVFAIL state is cleared. CIFS clients are unaffected.

How NVFAIL protects database files

The NVFAIL state is set in two cases, either when Data ONTAP detects NVRAM errors when booting up or when a MetroCluster switchover operation occurs. If no errors are detected at startup, the file service is started normally. However, if NVRAM errors are detected or the `force-fail` option was set and then there was a switchover, Data ONTAP stops database instances from responding.

When you enable the `NVFAIL` option, one of the following processes takes place during bootup.

If...	Then...
Data ONTAP detects no NVRAM errors	File service starts normally.
Data ONTAP detects NVRAM errors	• Data ONTAP returns a stale file handle (`ESTALE`) error to NFS clients trying to access the database, causing the application to stop responding, crash, or shut down. Data ONTAP then sends an error message to the system console and log file. • When the application restarts, files are available to CIFS clients, even if you have not verified that they are valid. For NFS clients, files remain inaccessible until you reset the `in-nvfailed-state` option on the affected volume.

If...	Then...
Data ONTAP detects NVRAM errors on a volume that contains LUNs	LUNs in that volume are brought offline. Then the `in-nvfailed-state` option on the volume must be cleared, and the `NVFAIL` attribute on the LUNs must be cleared by bringing each LUN in the affected volume online. You can perform the steps to check the integrity of the LUNs and recover the LUN from Snapshot copy or backup as necessary. After all the LUNs in the volume are recovered, the `in-nvfailed-state` option on the affected volume is cleared.

Commands for monitoring data loss events

If you enable the `NVFAIL` option, you receive notification when a system crash caused by NVRAM inconsistencies or a MetroCluster switchover occurs.

By default, the NVFAIL parameter is not enabled.

If you want to...	Use this command...
Create a new volume with NVFAIL enabled	`volume create -nvfail` **on**
Enable NVFAIL on an existing volume	`volume modify` **Note:** You set the `-nvfail` option to **on** to enable NVFAIL on the created volume.
Display whether NVFAIL is currently enabled for a specified volume	`volume show` **Note:** You set the `-fields` parameter to **`nvfail`** to display the NVFAIL attribute for a specified volume.

See the man page for each command for more information.

Accessing volumes in NVFAIL state after a switchover

After a switchover, you must clear the NVFAIL state by resetting the `-in-nvfailed-state` parameter of the `volume modify` command to remove the restriction of clients to access data.

Before you begin

The database or file system must not be running or trying to access the affected volume.

About this task

Setting `-in-nvfailed-state` parameter requires advanced-level privilege.

Step

1. Recover the volume by using the `volume modify` command with the `-in-nvfailed-state` parameter set to **`false`**.

After you finish

For instructions about examining database file validity, see the documentation for your specific database software.

If your database uses LUNs, review the steps to make the LUNs accessible to the host after an NVRAM failure.

Related concepts

Monitoring and protecting database validity by using NVFAIL on page 93

Recovering LUNs in NVFAIL states after switchover

After a switchover, the host no longer has access to data on the LUNs that are in NVFAIL states. You must perform a number of actions before the database has access to the LUNs.

Before you begin

The database must not be running.

Steps

1. Clear the NVFAIL state on the affect volume that hosts the LUNs by resetting the `-in-nvfailed-state` parameter of the `volume modify` command.
2. Bring the affected LUNs online.
3. Examine the LUNs for any data inconsistencies and resolve them.

   This might involve host-based recovery or recovery done on the storage controller using SnapRestore.
4. Bring the database application online after recovering the LUNs.

Glossary of MetroCluster terms

aggregate

A grouping of physical storage resources (disks or array LUNs) that provides storage to volumes associated with the aggregate. Aggregates provide the ability to control the RAID configuration for all associated volumes.

data SVM

Formerly known as data Vserver. In clustered Data ONTAP, a Storage Virtual Machine (SVM) that facilitates data access from the cluster; the hardware and storage resources of the cluster are dynamically shared by data SVMs within a cluster.

admin SVM

Formerly known as admin Vserver. In clustered Data ONTAP, a Storage Virtual Machine (SVM) that has overall administrative access to all objects in the cluster, including all objects owned by other SVMs, but does not provide data access to clients or hosts.

inter-switch link (ISL)

A connection between two switches using the E-port.

destination

The storage to which source data is backed up, mirrored, or migrated.

disaster recovery (DR) group

The four nodes in a MetroCluster configuration that synchronously replicate each others' configuration and data.

disaster recovery (DR) partner

A node's partner at the remote MetroCluster site. The node mirrors its DR partner's NVRAM or NVMEM partition.

disaster recovery auxiliary (DR auxiliary) partner

The HA partner of a node's DR partner. The DR auxiliary partner mirrors a node's NVRAM or NVMEM partition in the event of an HA takeover after a MetroCluster switchover operation.

HA pair

- In Data ONTAP 8.x, a pair of nodes whose controllers are configured to serve data for each other if one of the two nodes stops functioning.
  Depending on the system model, both controllers can be in a single chassis, or one controller can be in one chassis and the other controller can be in a separate chassis.
- In the Data ONTAP 7.3 and 7.2 release families, this functionality is referred to as an *active/active configuration*.

HA partner

A node's partner within the local HA pair. The node mirrors its HA partner's NVRAM or NVMEM cache.

high availability (HA)

In Data ONTAP 8.x, the recovery capability provided by a pair of nodes (storage systems), called an *HA pair*, that are configured to serve data for each other if one of the two nodes stops functioning. In the Data ONTAP 7.3 and 7.2 release families, this functionality is referred to as an *active/active configuration*.

healing

The two required MetroCluster operations that prepare the storage located at the DR site for switchback. The first heal operation resynchronizes the mirrored plexes. The second heal operation returns ownership of root aggregates to the DR nodes.

LIF (logical interface)

A logical network interface, representing a network access point to a node. LIFs currently correspond to IP addresses, but could be implemented by any interconnect. A LIF is generally bound to a physical network port; that is, an Ethernet port. LIFs can fail over to other physical ports (potentially on other nodes) based on policies interpreted by the LIF manager.

NVRAM

nonvolatile random-access memory.

NVRAM cache

Nonvolatile RAM in a storage system, used for logging incoming write data and NFS requests. Improves system performance and prevents loss of data in case of a storage system or power failure.

NVRAM mirror

A synchronously updated copy of the contents of the storage system NVRAM (nonvolatile random access memory) contents kept on the partner storage system.

node

- In Data ONTAP, one of the systems in a cluster or an HA pair.
  To distinguish between the two nodes in an HA pair, one node is sometimes called the *local node* and the other node is sometimes called the *partner node* or *remote node*.

- In Protection Manager and Provisioning Manager, the set of storage containers (storage systems, aggregates, volumes, or qtrees) that are assigned to a dataset and designated either primary data (primary node), secondary data (secondary node), or tertiary data (tertiary node).
  A *dataset node* refers to any of the nodes configured for a dataset.
  A *backup node* refers to either a secondary or tertiary node that is the destination of a backup or mirror operation.
  A *disaster recovery node* refers to the dataset node that is the destination of a failover operation.

remote storage

The storage that is accessible to the local node, but is at the location of the remote node.

root volume

A special volume on each Data ONTAP system. The root volume contains system files and configuration information, and can also contain data. It is required for the system to be able to boot and to function properly. Core dump files, which are important for troubleshooting, are written to the root volume if there is enough space.

switchback

The MetroCluster operation that restores service back to one of the MetroCluster sites.

switchover

The MetroCluster operation that transfers service from one of the MetroCluster sites.

- A *negotiated* switchover is planned in advance and cleanly shuts down components of the target MetroCluster site.
- A *forced* switchover immediately transfers service; the shut down of the target site might not be clean.

Copyright information

Trademark information

How to send comments about documentation and receive update notification

You can help us to improve the quality of our documentation by sending us your feedback. You can receive automatic notification when production-level (GA/FCS) documentation is initially released or important changes are made to existing production-level documents.

If you have suggestions for improving this document, send us your comments by email to *doccomments@netapp.com*. To help us direct your comments to the correct division, include in the subject line the product name, version, and operating system.

If you want to be notified automatically when production-level documentation is released or important changes are made to existing production-level documents, follow Twitter account @NetAppDoc.

You can also contact us in the following ways:

- NetApp, Inc., 495 East Java Drive, Sunnyvale, CA 94089 U.S.
- Telephone: +1 (408) 822-6000
- Fax: +1 (408) 822-4501
- Support telephone: +1 (888) 463-8277

Index

E

F

G

H

I

L

M

N

O

P

R

S

T

U

V