Failure scenarios

Enterprise applications

NetApp
February 10, 2026

Table of Contents

Failure scenarios

Overview

Planning a complete SnapMirror active sync application architecture requires understanding how SM-as will respond in various planned and unplanned failover scenarios.

For the following examples, assume that site A is configured as the preferred site.

Loss of replication connectivity

If SM-as replication is interrupted, write IO cannot be completed because it would be impossible for a cluster to replicate changes to the opposite site.

Site A (Preferred site)

The result of replication link failure on the preferred site will be an approximate 15 second pause in write IO processing as ONTAP retries replicated write operations before it determines that the replication link is genuinely unreachable. After the 15 seconds elapses, the site A system resumes read and write IO processing. The SAN paths will not change, and the LUNs will remain online.

Site B

Since site B is not the SnapMirror active sync preferred site, its LUN paths will become unavailable after about 15 seconds.

Storage system failure

The result of a storage system failure is nearly identical to the result of losing the replication link. The surviving site should experience a roughly 15 second IO pause. Once that 15 second period elapses, IO will resume on that site as usual.

Loss of the mediator

The mediator service does not directly control storage operations. It functions as an alternate control path between clusters. It exists primarily to automate failover without the risk of a split-brain scenario. In normal operation, each cluster is replicating changes to its partner, and each cluster therefore can verify that the partner cluster is online and serving data. If the replication link failed, replication would cease.

The reason a mediator is required for safe automated failover is because it would otherwise be impossible for a storage cluster to be able to determine whether loss of bidirectional communication was the result of a network outage or actual storage failure.

The mediator provides an alternate path for each cluster to verify the health of its partner. The scenarios are as follows:

- If a cluster can contact its partner directly, replication services are operational. No action required.
- If a preferred site cannot contact its partner directly or via the mediator, it will assume the partner is either actually unavailable or was isolated and has taken its LUN paths offline. The preferred site will then proceed to release the RPO=0 state and continue processing both read and write IO.

- If a non-preferred site cannot contact its partner directly, but can contact it via the mediator, it will take its paths offline and await the return of the replication connection.
- If a non-preferred site cannot contact its partner directly or via an operational mediator, it will assume the partner is either actually unavailable or was isolated and has taken its LUN paths offline. The non-preferred site will then proceed to release the RPO=0 state and continue processing both read and write IO. It will assume the role of the replication source and will become the new preferred site.

If the mediator is wholly unavailable:

- Failure of replication services for any reason, including failure of the nonpreferred site or storage system, will result in the preferred site releasing the RPO=0 state and resuming read and write IO processing. The non-preferred site will take its paths offline.
- Failure of the preferred site will result in an outage because the non-preferred site will be unable to verify that the opposite site is truly offline and therefore it would not be safe for the nonpreferred site to resume services.

Restoring services

After a failure is resolved, such as restoring site-to-site connectivity or powering on a failed system, the SnapMirror active sync endpoints will automatically detect the presence of a faulty replication relationship and bring it back to an RPO=0 state. Once synchronous replication is reestablished, the failed paths will come online again.

In many cases, clustered applications will automatically detect the return of failed paths, and those applications will also come back online. In other cases, a host-level SAN scan may be required, or applications may need to be brought back online manually. It depends on the application and how it is configured, and in general such tasks can be easily automated. ONTAP itself is self-healing and should not require any user intervention to resume RPO=0 storage operations.

Manual failover

Changing the preferred site requires a simple operation. IO will pause for a second or two as authority over replication behavior switches between clusters, but IO is otherwise unaffected.
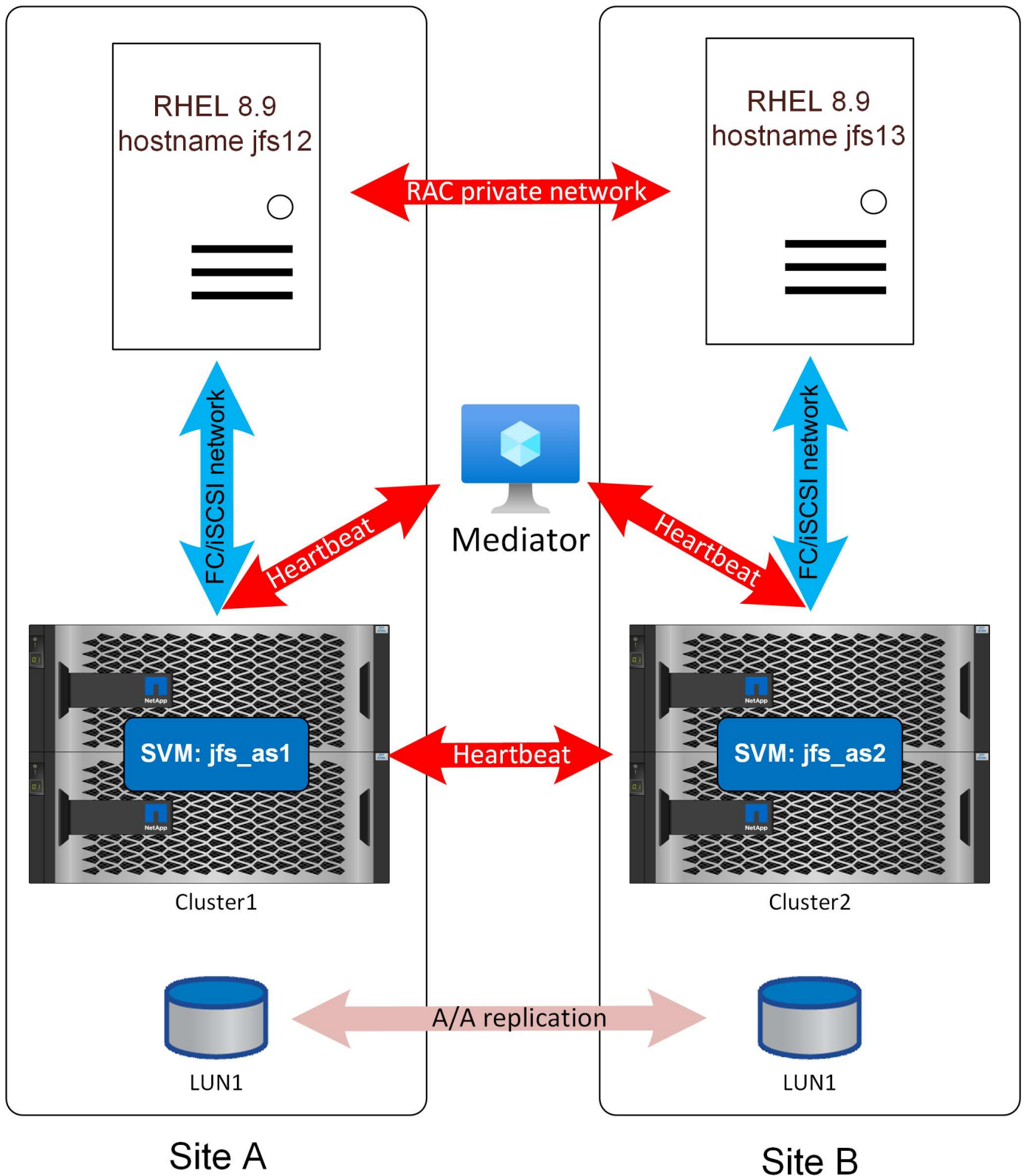
Sample architecture

The detailed failure examples shown in this sections are based on the architecture shown below.

> ⓘ This is only one of many options for Oracle databases on SnapMirror active sync. This design was chosen because it illustrates some of the more complicated scenarios.

In this design, assume that site A is set at the preferred site.

RAC interconnect failure

Loss of the Oracle RAC replication link will produce a similar result to loss of SnapMirror connectivity, except the timeouts will be shorter by default. Under default settings, an Oracle RAC node will wait 200 seconds after loss of storage connectivity before evicting,

but it will only wait 30 seconds after loss of the RAC network heartbeat.

The CRS messages are similar to those shown below. You can see the 30 second timeout lapse. Since css_critical was set on jfs12, located on site A, that will be the site to survive and jfs13 on site B will be evicted.

```
2024-09-12 10:56:44.047 [ONMD(3528)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 6.980 seconds
2024-09-12 10:56:48.048 [ONMD(3528)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.980 seconds
2024-09-12 10:56:51.031 [ONMD(3528)]CRS-1607: Node jfs13 is being evicted
in cluster incarnation 621599354; details at (:CSSNM00007:) in
/gridbase/diag/crs/jfs12/crs/trace/onmd.trc.
2024-09-12 10:56:52.390 [CRSD(6668)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:33194;', interface list of remote node 'jfs13' is
'192.168.30.2:33621;'.
2024-09-12 10:56:55.683 [ONMD(3528)]CRS-1601: CSSD Reconfiguration
complete. Active nodes are jfs12 .
2024-09-12 10:56:55.722 [CRSD(6668)]CRS-5504: Node down event reported for
node 'jfs13'.
2024-09-12 10:56:57.222 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'Generic'.
2024-09-12 10:56:57.224 [CRSD(6668)]CRS-2773: Server 'jfs13' has been
removed from pool 'ora.NTAP'.
```

SnapMirror communication failure

If the SnapMirror active sync replication link, write IO cannot be completed because it would be impossible for a cluster to replicate changes to the opposite site.

Site A

The result on site A of a replication link failure will be an approximately 15 second pause in write IO processing as ONTAP attempts to replicate writes before it determines that the replication link is genuinely inoperable. After the 15 seconds elapses, the ONTAP cluster on site A resumes read and write IO processing. The SAN paths will not change, and the LUNs will remain online.

Site B

Since site B is not the SnapMirror active sync preferred site, its LUN paths will become unavailable after about 15 seconds.

The replication link was cut at the timestamp 15:19:44. The first warning from Oracle RAC arrives 100 seconds later as the 200 second timeout (controlled by the Oracle RAC parameter disktimeout) approaches.

```
2024-09-10 15:21:24.702 [ONMD(2792)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99340 milliseconds.
2024-09-10 15:22:14.706 [ONMD(2792)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49330 milliseconds.
2024-09-10 15:22:44.708 [ONMD(2792)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19330 milliseconds.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-10 15:23:04.710 [ONMD(2792)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.716 [ONMD(2792)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-10 15:23:04.731 [OCSSD(2794)]CRS-1652: Starting clean up of CRSD
resources.
```

Once the 200 second voting disk timeout has been reached, this Oracle RAC node will evict itself from the cluster and reboot.

Total network interconnectivity failure

If the replication link between sites is completely lost, both SnapMirror active sync and Oracle RAC connectivity will be interrupted.

Oracle RAC split-brain detection has a dependency on the Oracle RAC storage heartbeat. If loss of site-to-site connectivity results in simultaneous loss of both the RAC network heartbeat and storage replication services, the result is the RAC sites will not be able to communicate cross-site via either the RAC interconnect or the RAC voting disks. The result in an even-numbed set of nodes may be eviction of both sites under default settings. The exact behavior will depend on the sequence of events and the timing of the RAC network and disk heartbeat polls.

The risk of a 2-site outage can be addressed in two ways. First, a tiebreaker configuration can be used.

If a 3rd site is not available, this risk can be addressed by adjusting the misscount parameter on the RAC cluster. Under the defaults, the RAC network heartbeat timeout is 30 seconds. This normally is used by RAC to identify failed RAC nodes and remove them from the cluster. It also has a connection to the voting disk heartbeat.

If, for example, the conduit carrying intersite traffic for both Oracle RAC and storage replication services is cut by a backhoe, the 30 second misscount countdown will begin. If the RAC preferred site node cannot reestablish contact with the opposite site within 30 seconds, and it also cannot use the voting disks to confirm

the opposite site is down within that same 30 second window, then the preferred site nodes will also evict. The result is a full database outage.

Depending on when the misscount polling occurs, 30 seconds may not be enough time for SnapMirror active sync to time out and allow storage on the preferred site to resume services before the 30 second window expires. This 30 second window can be increased.

```
[root@jfs12 ~]# /grid/bin/crsctl set css misscount 100
CRS-4684: Successful set of parameter misscount to 100 for Cluster
Synchronization Services.
```

This value allows the storage system on the preferred site to resume operations before the miscount timeout expires. The result will then be eviction only of the nodes at the site where the LUN paths were removed. Example below:

```
2024-09-12 09:50:59.352 [ONMD(681360)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 49.570 seconds
2024-09-12 09:51:10.082 [CRSD(682669)]CRS-7503: The Oracle Grid
Infrastructure process 'crsd' observed communication issues between node
'jfs12' and node 'jfs13', interface list of local node 'jfs12' is
'192.168.30.1:46039;', interface list of remote node 'jfs13' is
'192.168.30.2:42037;'.
2024-09-12 09:51:24.356 [ONMD(681360)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 24.560 seconds
2024-09-12 09:51:39.359 [ONMD(681360)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 9.560 seconds
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8011: reboot advisory message
from host: jfs13, component: cssagent, with time stamp: L-2024-09-12-
09:51:47.451
2024-09-12 09:51:47.527 [OHASD(680884)]CRS-8013: reboot advisory message
text: oracssdagent is about to reboot this node due to unknown reason as
it did not receive local heartbeats for 10470 ms amount of time
2024-09-12 09:51:48.925 [ONMD(681360)]CRS-1632: Node jfs13 is being
removed from the cluster in cluster incarnation 621596607
```

Oracle Support strongly discourages altering with the misscount or disktimeout parameters to solve configuration problems. Changing these parameters can, however, be warranted and unavoidable in many cases, including SAN booting, virtualized, and storage replication configurations. If, for example, you had stability problems with a SAN or IP network that was resulting in RAC evictions you should fix the underlying problem and not charge the values of the misscount or disktimeout. Changing timeouts to address configuration errors is masking a problem, not solving a problem. Changing these parameters to properly configure a RAC environment based on design aspects of the underlying infrastructure is different and is consistent with Oracle support statements. With SAN booting, it is common to adjust misscount all the way up to 200 to match disktimeout. See this link for additional information.

Site failure

The result of a storage system or site failure is nearly identical to the result of losing the replication link. The surviving site should experience a roughly 15 second IO pause on writes. Once that 15 second period elapses, IO will resume on that site as usual.

If only the storage system was affected, the Oracle RAC node on the failed site will lose storage services and enter the same 200 second disktimeout countdown before eviction and subsequent reboot.

```
2024-09-11 13:44:38.613 [ONMD(3629)]CRS-1615: No I/O has completed after
50% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 99750 milliseconds.
2024-09-11 13:44:51.202 [ORAAGENT(5437)]CRS-5011: Check of resource "NTAP"
failed: details at "(:CLSN00007:)" in
"/gridbase/diag/crs/jfs13/crs/trace/crsd_oraagent_oracle.trc"
2024-09-11 13:44:51.798 [ORAAGENT(75914)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 75914
2024-09-11 13:45:28.626 [ONMD(3629)]CRS-1614: No I/O has completed after
75% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 49730 milliseconds.
2024-09-11 13:45:33.339 [ORAAGENT(76328)]CRS-8500: Oracle Clusterware
ORAAGENT process is starting with operating system process ID 76328
2024-09-11 13:45:58.629 [ONMD(3629)]CRS-1613: No I/O has completed after
90% of the maximum interval. If this persists, voting file
/dev/mapper/grid2 will be considered not functional in 19730 milliseconds.
2024-09-11 13:46:18.630 [ONMD(3629)]CRS-1604: CSSD voting file is offline:
/dev/mapper/grid2; details at (:CSSNM00058:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc.
2024-09-11 13:46:18.631 [ONMD(3629)]CRS-1606: The number of voting files
available, 0, is less than the minimum number of voting files required, 1,
resulting in CSSD termination to ensure data integrity; details at
(:CSSNM00018:) in /gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.638 [ONMD(3629)]CRS-1699: The CSS daemon is
terminating due to a fatal error from thread:
clssnmvDiskPingMonitorThread; Details at (:CSSSC00012:) in
/gridbase/diag/crs/jfs13/crs/trace/onmd.trc
2024-09-11 13:46:18.651 [OCSSD(3631)]CRS-1652: Starting clean up of CRSD
resources.
```

The SAN path state on the RAC node that has lost storage services looks like this:

```
oradata7 (3600a0980383041334a3f55676c697347) dm-20 NETAPP,LUN C-Mode
size=128G features='3 queue_if_no_path pg_init_retries 50' hwhandler='1
alua' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
| `- 34:0:0:18 sdam 66:96  failed faulty running
`-+- policy='service-time 0' prio=0 status=enabled
  `- 33:0:0:18 sdaj 66:48  failed faulty running
```

The linux host detected the loss of the paths much quicker than 200 seconds, but from a database perspective the client connections to the host on the failed site will still be frozen for 200 seconds under the default Oracle RAC settings. Full database operations will only resume after the eviction is completed.

Meanwhile, the Oracle RAC node on the opposite site will record the loss of the other RAC node. It otherwise continues to operate as usual.

```
2024-09-11 13:46:34.152 [ONMD(3547)]CRS-1612: Network communication with
node jfs13 (2) has been missing for 50% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 14.020 seconds
2024-09-11 13:46:41.154 [ONMD(3547)]CRS-1611: Network communication with
node jfs13 (2) has been missing for 75% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 7.010 seconds
2024-09-11 13:46:46.155 [ONMD(3547)]CRS-1610: Network communication with
node jfs13 (2) has been missing for 90% of the timeout interval.  If this
persists, removal of this node from cluster will occur in 2.010 seconds
2024-09-11 13:46:46.470 [OHASD(1705)]CRS-8011: reboot advisory message
from host: jfs13, component: cssmonit, with time stamp: L-2024-09-11-
13:46:46.404
2024-09-11 13:46:46.471 [OHASD(1705)]CRS-8013: reboot advisory message
text: At this point node has lost voting file majority access and
oracssdmonitor is rebooting the node due to unknown reason as it did not
receive local heartbeats for 28180 ms amount of time
2024-09-11 13:46:48.173 [ONMD(3547)]CRS-1632: Node jfs13 is being removed
from the cluster in cluster incarnation 621516934
```

Mediator failure

The mediator service does not directly control storage operations. It functions as an alternate control path between clusters. It exists primarily to automate failover without the risk of a split-brain scenario.

In normal operation, each cluster is replicating changes to its partner, and each cluster therefore can verify that the partner cluster is online and serving data. If the replication link failed, replication would cease.

The reason a mediator is required for safe automated operations is because it would otherwise be impossible for a storage clusters to be able to determine whether loss of bidirectional communication was the result of a

network outage or actual storage failure.

The mediator provides an alternate path for each cluster to verify the health of its partner. The scenarios are as follows:

- If a cluster can contact its partner directly, replication services are operational. No action required.
- If a preferred site cannot contact its partner directly or via the mediator, it will assume the partner is either actually unavailable or was isolated and has taken its LUN paths offline. The preferred site will then proceed to release the RPO=0 state and continue processing both read and write IO.
- If a non-preferred site cannot contact its partner directly, but can contact it via the mediator, it will take its paths offline and await the return of the replication connection.
- If a non-preferred site cannot contact its partner directly or via an operational mediator, it will assume the partner is either actually unavailable or was isolated and has taken its LUN paths offline. The non-preferred site will then proceed to release the RPO=0 state and continue processing both read and write IO. It will assume the role of the replication source and will become the new preferred site.

If the mediator is wholly unavailable:

- Failure of replication services for any reason will result in the preferred site releasing the RPO=0 state and resuming read and write IO processing. The non-preferred site will take its paths offline.
- Failure of the preferred site will result in an outage because the non-preferred site will be unable to verify that the opposite site is truly offline and therefore it would not be safe for the nonpreferred site to resume services.

Service restoration

SnapMirror is self-healing. SnapMirror active sync will automatically detect the presence of a faulty replication relationship and bring it back to an RPO=0 state. Once synchronous replication is reestablished, the paths will come online again.

In many cases, clustered applications will automatically detect the return of failed paths, and those applications will also come back online. In other cases, a host-level SAN scan may be required, or applications may need to be brought back online manually.

It depends on the application and how it's configured, and in general such tasks can be easily automated. SnapMirror active sync itself is self-fixing and should not require any user intervention to resume RPO=0 storage operations once power and connectivity is restored.

Manual failover

The term "failover" does not refer to the direction of replication with SnapMirror active sync because it is a bidirectional replication technology. Instead, 'failover' refers to which storage system will be the preferred site in the event of failure.

For example, you may want to perform a failover to change the preferred site before you shut down a site for maintenance, or before performing a DR test.

Changing the preferred site requires a simple operation. IO will pause for a second or two as authority over replication behavior switches between clusters, but IO is otherwise unaffected.

GUI example:

Example of changing it back via the CLI:

```
Cluster2::> snapmirror failover start -destination-path jfs_as2:/cg/jfsAA
[Job 9575] Job is queued: SnapMirror failover for destination
"jfs_as2:/cg/jfsAA                 ".

Cluster2::> snapmirror failover show

Source      Destination                                        Error
Path        Path          Type      Status    start-time end-time  Reason
--------    -----------   --------  --------- ---------- ---------- ----------
jfs_as1:/cg/jfsAA
            jfs_as2:/cg/jfsAA
                          planned   completed 9/11/2024  9/11/2024
                                              09:29:22   09:29:32


The new destination path can be verified as follows:


Cluster1::> snapmirror show -destination-path jfs_as1:/cg/jfsAA

                         Source Path: jfs_as2:/cg/jfsAA
                    Destination Path: jfs_as1:/cg/jfsAA
                   Relationship Type: XDP
             Relationship Group Type: consistencygroup
               SnapMirror Policy Type: automated-failover-duplex
                    SnapMirror Policy: AutomatedFailOverDuplex
                          Tries Limit: -
                          Mirror State: Snapmirrored
                 Relationship Status: InSync
```