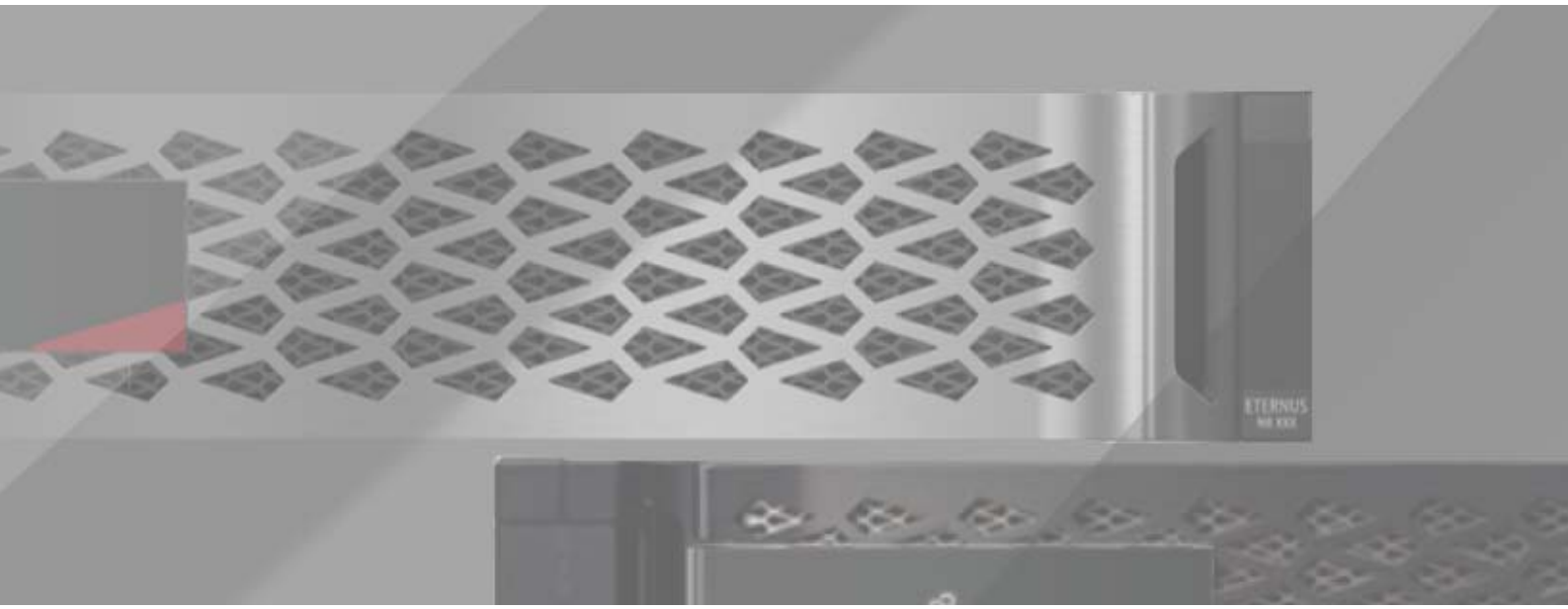
C140-0040-01ENZ3

Fujitsu Storage
ETERNUS AX series All-Flash Arrays

SnapMirror Business Continuity (SM-BC)
ONTAP 9.12.1



FUJITSU

Table of Contents

List of Figures

Preface

SnapMirror Business Continuity (SM-BC) is a continuously available storage solution with application-level granularity, available for ONTAP running on ETERNUS AX series or ETERNUS AX series All SAN Array (ASA) storage systems, to meet the RPO 0 and RTO 0 needs of the most critical business applications.

First Edition
March 2023

Trademarks

Third-party trademark information related to this product is available at:
https://www.fujitsu.com/global/products/computing/storage/eternus/trademarks.html

Trademark symbols such as ™ and ® are omitted in this document.

About This Manual

Intended Audience

This manual is intended for system administrators who configure and manage operations of the ETERNUS AX, or field engineers who perform maintenance. Refer to this manual as required.

Related Information and Documents

The latest information for the ETERNUS AX is available at:
https://www.fujitsu.com/global/support/products/computing/storage/manuals-list.html

Document Conventions

Notice Symbols

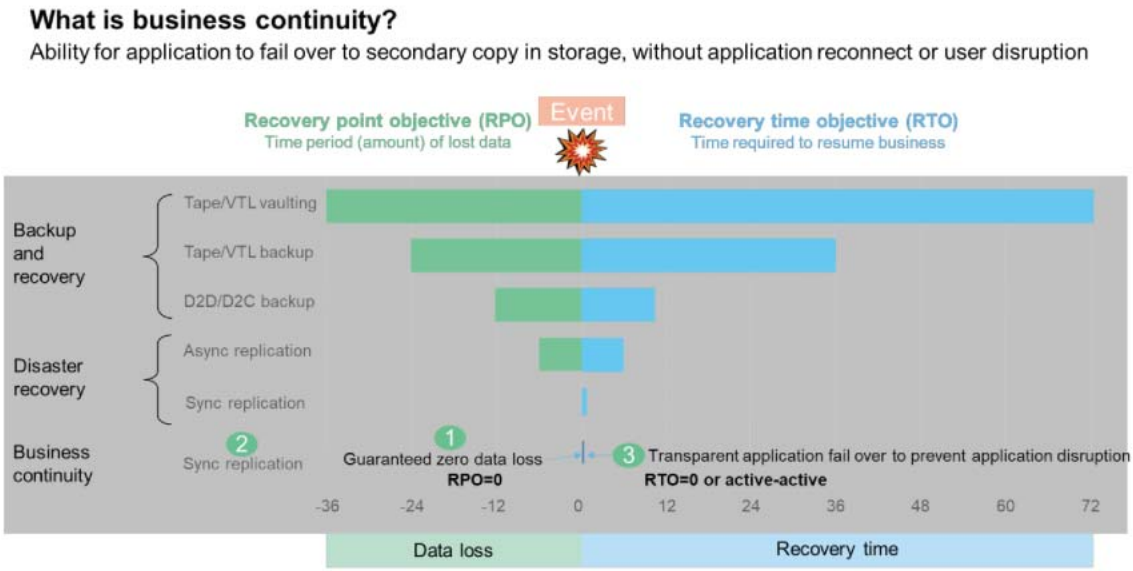
The following notice symbols are used in this manual:

Caution	Indicates information that you need to observe when using the ETERNUS AX. Make sure to read the information.
Note	Indicates information and suggestions that supplement the descriptions included in this manual.

1. Executive Summary

In today's constantly connected global business environment, organizations want rapid recovery of business-critical application data with zero data loss, following a disruption such as a cyberattack, power outage, equipment failure to natural disaster. Especially, financial organizations have zero tolerance to data loss or application unavailability and must adhere to General Data Protection Regulation (GDPR) and other regulatory mandates. Figure 1 provides a high-level description of business continuity.

Figure 1 What is business continuity?



Organizations can devise an effective business continuity and disaster recovery (BCDR) plan with the following requirements as priority:

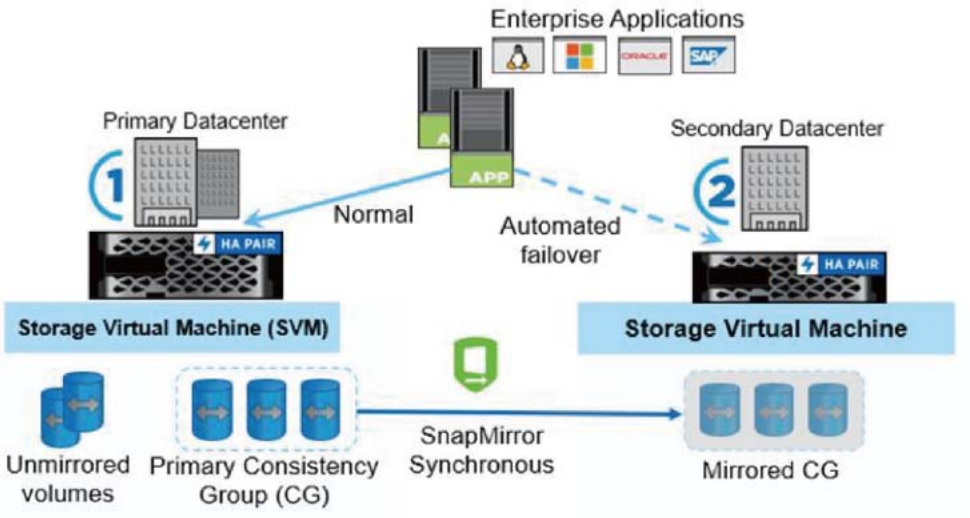
- Zero recovery point objective (RPO), to achieve zero data loss enabled by synchronous replication
- Zero recovery time objective (RTO) through Transparent Application Failover (TAF), to prevent disruption of business-critical applications in case of disaster

Introduced in ONTAP 9.9.1, SM-BC is a new business continuity solution for zero RPO and near zero RTO. SM-BC gives you flexibility with easy-to-use application-level granularity and automatic failover. SM-BC uses proven SnapMirror Synchronous (SM-S) replication over IP network to replicate data at high speeds over LAN or WAN, to achieve high data availability and fast data replication for your business-critical applications such as Oracle, Microsoft SQL Server, and so on, in both virtual and physical environments. SM-BC enables mission-critical business services to continue operating even through a complete site failure, with TAF to the secondary copy. No manual intervention or no additional scripting are required to trigger this failover.

2. Solution Overview

SM-BC allows you to protect your data LUNs, which enables applications to fail over transparently for the purpose of business continuity in the event of a disaster. Figure 2 is an illustration of the SM-BC solution.

Figure 2 SM-BC solution



The key features of SM-BC include:

- Business continuity for SAN applications (iSCSI or FC) with protection across two separate geographic locations
- TAF for enterprise applications such as Oracle, Microsoft SQL Server, VMware vSphere Metro Storage Cluster (vMSC) solution, and so on, without manual intervention (no reconnect to storage or disruption to application users)
- Consistency group (CG) maintains dependent write order consistency for a collection of volumes for application data
- Tight integration with ONTAP to leverage robust Fujitsu technologies to create a highly scalable, enterprise-level data protection
- Simplified data management for storage provisioning, host connections, and creation of Snapshot copies and clones for both sites

Benefits

SM-BC provides the following benefits:

- Application granularity for business continuity
- Automated failover with the ability to assess failover for each application
- LUN identity remains the same, so the application sees them as a shared virtual device
- Ability to reuse secondary with flexibility to create instantaneous clones for application usage for dev-test, UAT or reporting purposes, without impacting application performance or availability
- Simplified application management using consistency groups to maintain dependent write-order consistency

3. Key Terminology and Concepts

As you begin to explore the ONTAP SM-BC solution and plan a deployment, it is important to become familiar with the key terminology and concepts.

SM-BC
Acronym for the SnapMirror Business Continuity (SM-BC) solution available with ONTAP 9.9.1 and later.

Consistency group (CG)
A CG is a container that holds several volumes so that you can perform snapshot image operations such as creation, scheduling and rolling back, and so on. For example, if the host has application data spread across multiple volumes (such as a virtual machine [VM] with multiple virtual drives, or a database server with isolation of data, logs, and other files), then it becomes critical to ensure snapshot consistency for the protected and replicated data. The CG is a collection of FlexVol volumes that provide a consistency guarantee for the application workload that must be protected for business continuity. The purpose of a CG is to take simultaneous snapshot images of multiple volumes, thus ensuring crash-consistent copies of a collection of volumes at a point-in-time (PiT). A CG ensures all volumes of a dataset are quiesced and then snapped at precisely the same PiT. This provides a data-consistent restore point across volumes supporting the dataset. A CG thereby maintains dependent write-order consistency. If you decide to protect applications for business continuity, the group of volumes corresponding to this application must be added to a CG and then the data protection relationship is established between a source CG and a destination CG. The source and destination CGs must contain the same number and type of volumes.

Constituent
The individual FlexVol volumes that are part of a CG.

Zero RPO
Zero recovery point objective; recover from a disaster or outage with no data loss.

Zero RTO
Zero recovery time objective or TAF for nondisruptive access to the storage, achieved by using host multipath I/O (MPIO).

Transparent Application Failover (TAF)
A feature that makes storage outage on site completely agnostic to application and does not require any reconfiguration post a site disaster or storage outage. In other words, host access to storage is nondisruptive in the event of a site-disaster or storage outage of site. For SAN, this requires host MPIO to make a storage failover transparent to an application.

Out of sync
The application I/O is not replicating to the secondary storage system. The destination volume is not synchronized with the source volume because SnapMirror replication is not taking place, indicating transfer failure or failure due to an unsupported operation.

ONTAP Mediator
ONTAP Mediator is a tool that the ONTAP clusters/nodes report their heartbeat information to periodically, to determine if its peer is up and running or not. ONTAP Mediator provides the following health information:

- Peer ONTAP cluster
- Peer ONTAP cluster nodes
- CG (which is the failover unit); for each CG, the following information is provided:
  - Replication state: Uninitialized, In Sync, or Out of Sync
  - Which cluster hosts the primary copy
  - Operation context (used for planned failover)

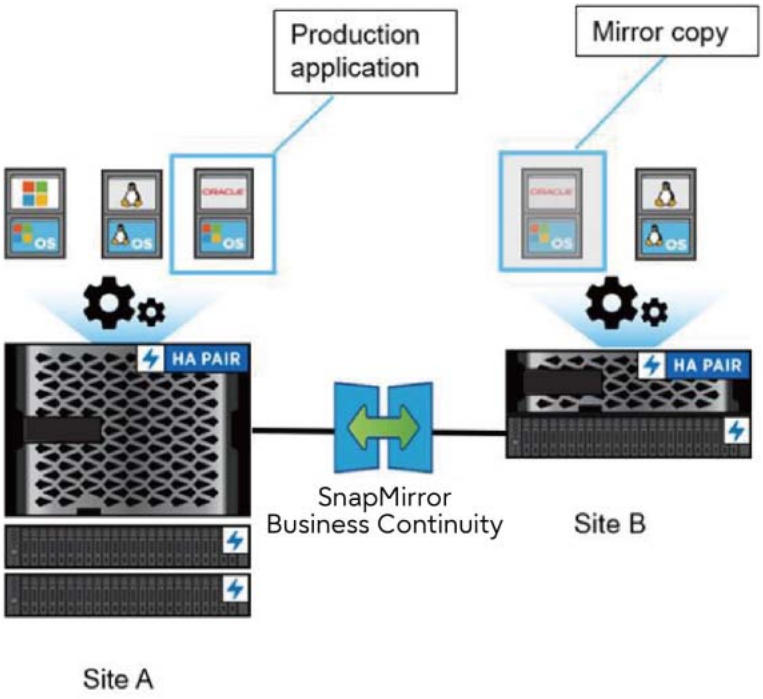
With this ONTAP Mediator health information, clusters can differentiate between distinct types of failures and determine whether to perform an automated failover. ONTAP Mediator is one of the three parties in the SM-BC quorum along with both ONTAP clusters (primary and secondary). To reach consensus, at least two parties in the quorum must agree to a certain operation.

Architecture

Figure 3 illustrates the SM-BC architecture for active workloads on both clusters, where primary workloads can be served simultaneously from both clusters. Regulations for financial institutions in some countries require businesses to be periodically serviceable from their secondary data centers as well, called "Tick-Tock" deployments, which SM-BC enables.

The data protection relationship to protect for business continuity is created between the source storage system and destination storage system, by adding the application specific LUNs from different volumes within a storage virtual machine (SVM) to the consistency group. Under normal operations, the enterprise application writes to the primary consistency group, which synchronously replicates this I/O to the mirror consistency group.

Figure 3 SM-BC architecture



Even though two separate copies exist in the data protection relationship, because SM-BC maintains the same LUN identity, the application host sees this as a shared virtual device with multiple paths while only one LUN copy is being written to at a time. When a failure renders the primary storage system offline, ONTAP Mediator detects this failure and enables seamless application failover to the mirror consistency group. This process results in failing over only a specific application without the need for the manual intervention or scripting which was previously required for the purpose of failover.

Other points to consider:

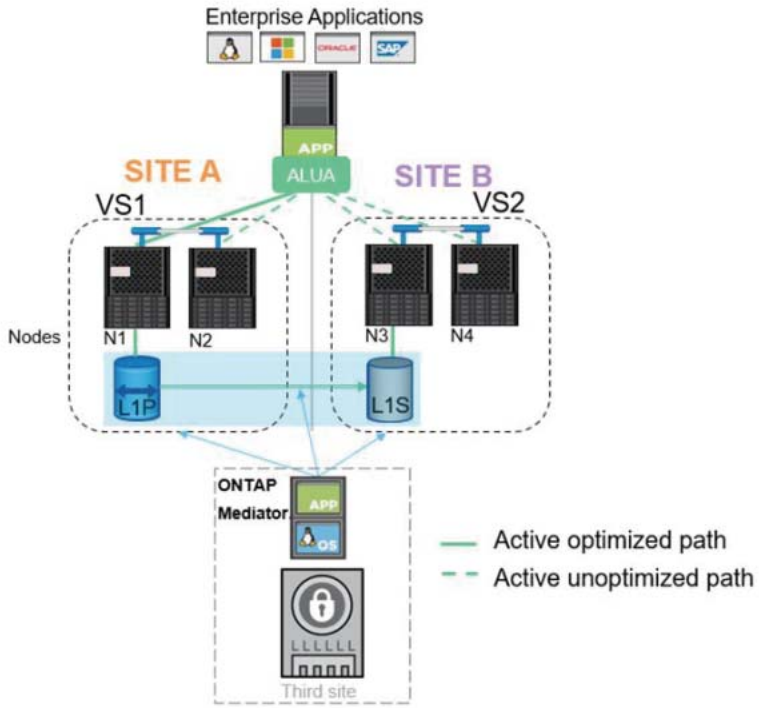
- Unmirrored volumes which exist outside of protection for business continuity are supported.
- Only one other SnapMirror asynchronous relationship is supported for volumes being protected for business continuity.
- Cascade topologies are not supported with protection for business continuity.

Use Cases

Application Deployment for Zero RTO or Transparent Application Failover (TAF)

TAF is based on host MPIO software-based path failover to achieve nondisruptive access to the storage. Both LUN copies, for example, primary (L1P) and mirror copy (L1S), have the same identity (serial number) and are reported as read-writable to the host. However, reads and writes are serviced only by the primary volume. I/Os issued to the mirror copy are proxied to the primary copy. The host's preferred path to L1 is VS1:N1 based on asymmetric logical unit access (ALUA) access state Active Optimized (A/O). ONTAP Mediator is required as part of the deployment, primarily to perform failover (planned or unplanned) in the event of a storage outage on the primary.

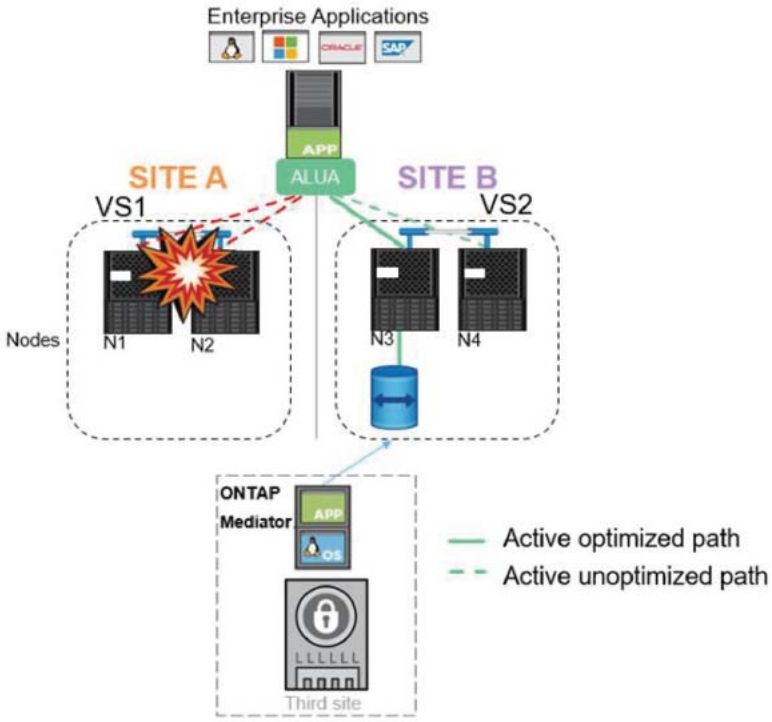
Figure 4 SM-BC uses ALUA paths



Disaster Scenario

When the site hosting the primary cluster experiences a disaster, the host multipathing software marks all paths through the cluster as down and uses paths from the secondary cluster. The result is a nondisruptive failover enabled by ONTAP Mediator, to the mirror copy for LUN L1 with L1S now converted from a mirror copy to an active copy with the hosts preferred path through VS2:N3.

Figure 5 Changed ALUA paths upon disaster



4. Configuring Applications

Application Host

The supported host operating systems are stand-alone Microsoft Windows Server, stand-alone Red Hat Enterprise Linux (RHEL), and VMware vMSC.

Protocols

Using iSCSI and FC protocols, SM-BC protects LUNs (logical units) in SAN environments which appear as disk devices to physical or ESXi hosts.

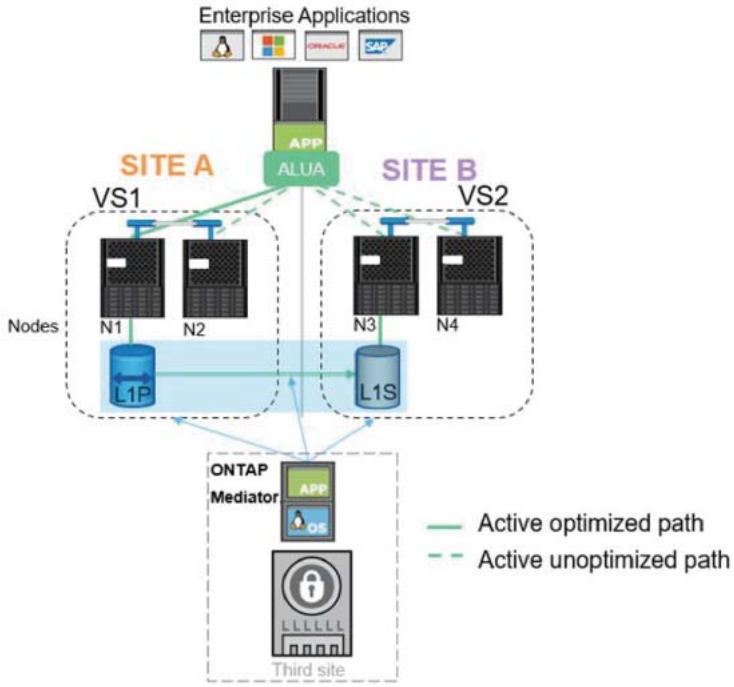
Host Connectivity

The host-to-storage network connectivity can use either FC or Ethernet (iSCSI) connectivity.

Multipathing

As shown in Figure 6, SM-BC uses ALUA, a standard SCSI mechanism that allows an application host multipathing software with paths advertised with priorities and access availability for the application host communication with the storage array. ALUA marks active optimized paths to the controllers owning the LUN as primary paths and others as nonoptimized paths. These nonoptimized paths are used only if the primary path fails.

Figure 6 SM-BC using ALUA

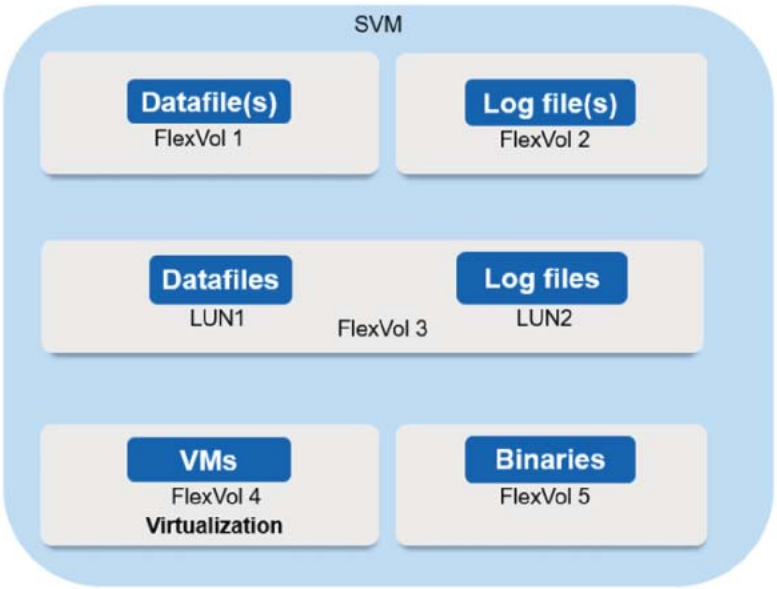


Data Layout

It is important that your data protection strategy clearly has the workloads identified, which need to be protected for business continuity. The most critical step in your data protection strategy is to have clarity in your enterprise application data layout so that you can still decide how you are distributing the volumes and protecting business continuity. Because the seamless failover is at the CG on a per-application basis, make sure to add the necessary data volumes to the CG.

Figure 7 shows the data layout within the SVM for enterprise application.

Figure 7 Data layout within the SVM for enterprise application



- Data volumes:
  - Random read workloads are isolated from sequential writes; therefore, depending on the database size, the data and log files are typically placed on separate volumes.
    - For large critical databases, the single data file is on FlexVol 1 and its corresponding log file is on FlexVol 2.
    - For better consolidation, small-to-medium-size noncritical databases are grouped such that all the data files are on FlexVol 1 and their corresponding log files are on FlexVol 2. However, you will lose application-level granularity through this grouping.
  - Another variant is to have all the files within the same FlexVol 3, with data files in LUN1 and its log files in LUN 2.
- If your environment is virtualized, you would have all the VMs for various enterprise applications shared in a datastore. Typically, the VMs and application binaries are asynchronously replicated using SnapMirror.

5. Configuring Protection for Business Continuity

Prerequisites

Hardware

SM-BC supports only two-node HA clusters of the ETERNUS AX series or ETERNUS AX series ASA models. Both primary and secondary clusters must be the same model of ETERNUS AX series or ETERNUS AX series ASA. Protection for business continuity involving ETERNUS HX series models is not supported.

Caution

The purpose of business continuity is to protect from failures that can render a site inoperable, such as disasters, and to allow your business operations to continue without any disruption. Therefore, you cannot protect business continuity within the same cluster. The source and destination clusters must be separate.

License

You are entitled to use SM-BC if you have the data protection on both the source and destination storage clusters.

Software

- Your storage system should be ONTAP 9.9.1 or later.
- All nodes on the source and destination clusters should be installed or upgraded to ONTAP 9.9.1 or later.
- ONTAP Mediator 1.5 or later will need to be installed on a Linux server or VM running RHEL (7.6, 7.7, 7.8, 7.9, 8.1, 8.2, 8.3, 8.4, 8.5) or CentOS (7.6, 7.7, 7.8, 7.9).

Network

Storage array-based replication transport is over TCP/IP network, with a maximum round trip time (RTT) latency of less than 10 ms between the source and destination storage systems.

ONTAP Cluster Configuration

Ensure source and destination clusters are configured properly.

ONTAP Mediator

ONTAP Mediator is installed in a third failure domain, separate from the two ONTAP clusters, and is present to establish consensus across a three-party quorum for:

- Primary ONTAP cluster hosting the SM-BC primary CG
- Secondary ONTAP cluster hosting the mirror CG
- ONTAP Mediator

ONTAP periodically sends a heartbeat of the node and controller through the node management LIF and cluster management LIF respectively, along with the status of the replication between storage systems to ONTAP Mediator. Redundant connections through multiple paths are established to distinguish between site failure and Inter-Switch Link (ISL) failure (when inter-site links are down). With the Mediator's health information, clusters can differentiate between intercluster LIF failure, site failure, and so on. With the Mediator-provided information and the intercluster LIF health check information, ONTAP can decide whether to perform the automated failover or not. When there is a loss of connection between the ONTAP Mediator software and all the nodes in the cluster and to the cluster itself (due to an event), that cluster is declared as not reachable. ONTAP Mediator attempts this every three seconds and tries three times to detect a failure (connection, site, and so on), after which the surviving cluster (which is still reachable) indicates that all the links to the partner cluster are severed. ONTAP Mediator then triggers an alert and enables automated failover to the mirror CG in the secondary site such that there is no I/O disruption to the client. If a network glitch or a network event (for example, a link going down) is not rectified manually or automatically within this nine-second window, it can result in heartbeat failure and the relationship goes out of sync, unless there is a redundant path available (for example, LIF failing over to another port) that can sustain the heartbeat.

ONTAP Mediator also helps protect from a split-brain scenario where every node, due to a network or disk heartbeat failure with other nodes in the cluster, assumes that it is the sole surviving member of the cluster, thereby proclaiming itself to be the primary.

To summarize, ONTAP Mediator is used for the following purposes:

- Establish a quorum
- Avoid a split-brain scenario
- Enable the automated failover when a failure is detected

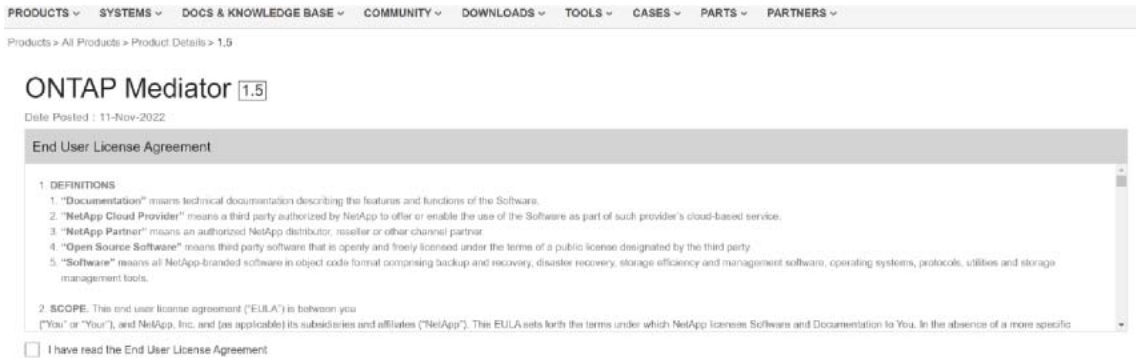
Note

- ONTAP Mediator 1.5 can manage ten cluster pairs for the purpose of business continuity.
- ONTAP Mediator 1.5 can also manage MetroCluster IP (MC IP) configurations, in addition to SM-BC configurations.
- When ONTAP Mediator is not available, you cannot perform planned or automated failovers. Although the application data will continue to be synchronously replicated without any interruption to for zero data loss.

Install ONTAP Mediator

You must install ONTAP Mediator, which includes accepting the licensing agreement and the Caution/Must Read statements shown in Figure 8, before you can configure and use the SM-BC solution.

Figure 8 Caution/MustRead statement



Note

When installing the mediator, you should replace the self-signed certificate with a valid certificate signed by a mainstream reliable CA.

Configure ONTAP Mediator

You must initialize ONTAP Mediator before you can start protecting for business continuity, which can be done either using the graphical user interface with ONTAP System Manager or ONTAP CLI.

In ONTAP System Manager, complete the following steps:

Procedure ▶▶▶

1. Navigate to Protection > Overview > Mediator > Configure.
2. Click Add, and enter the following Mediator server information:
   - IPv4 Address
   - User Name
   - Password

   Use the procedure to initialize ONTAP Mediator using CLI.

◀◀◀

During normal operation, the ONTAP Mediator state should be connected. If it is in any other state, this might indicate an error condition which can be viewed in Event Management System (EMS) messages and take appropriate corrective actions.

Protection for Business Continuity

Protection for business continuity involves creating a data protection relationship between two ONTAP storage systems and adding the LUNs specific to the application to the consistency group, known as a protection group.

Note

LUNs must reside within the same SVM.

In ONTAP System Manager, complete the following steps:

Procedure

1. Protection > Overview > Protect for Business Continuity > Protect LUNs.
2. Select one or more LUNs to protect on the source cluster.
3. Select the destination cluster and SVM.
4. Initialize relationship is selected by default. Click Save to begin protection.
5. Use ONTAP System Manager on the destination cluster. To verify that the protection for business continuity relationship is "In sync," select Protection > Relationships.

Note

- Volumes associated with an AppDM (such as for use as an Oracle, Microsoft SQL Server, Virtual Server, Virtual desktop application, or a generic SAN or NAS container), created from the ONTAP System Manager > Applications tab, is not supported.
- In ONTAP 9.12.1, no more than fifty consistency groups are supported.
- You can additionally create only one more asynchronous data protection relationship for the source volumes protected for business continuity.

Protect for business continuity from the ONTAP CLI using procedure Creating a consistency group relationship and Initializing a consistency group.

Note

After setting up the protection relationship, replicated LUNs in the secondary cluster must be mapped to the host and the I/O paths to the LUNs from both the primary and secondary cluster must be discovered at the time of host configuration.

In ONTAP System Manager, complete the following steps:

Procedure

1. Storage > LUNs > Ensure the destination LUNs are visible.
2. Edit each LUN to map to correct host mapping > initiator group.
3. ONTAP CLI, refer to Mapping LUNs to the application hosts.
4. On the application host, you need to rescan the disks and confirm that the paths to the destination LUNs (from secondary storage system) are discovered and ALUA advertises these paths correctly.

Storage LIF

If the service policy of the SAN LIF is changed as per the scenarios shown below, these changes are not replicated to the peer cluster accurately. This requires a `snapmirror abort` followed by `resync` to ensure these changes are replicated.

Scenario 1

Change the service policy of SAN LIF from `default-data-blocks` to `default-data-files` and back.

```
C1_sti96-vsim-ucs540a_cluster::*> net int show -vserver vs0 -fields data-protocol
  (network interface show)
vserver lif  data-protocol
------- ---- -------------
vs0     lif1 iscsi

C1_sti96-vsim-ucs540a_cluster::*> iscsi interface show -fields relative-port-id
vserver lif  relative-port-id
------- ---- ----------------
vs0     lif1 2

C1_sti96-vsim-ucs540a_cluster::*> net int modify -vserver vs0 -lif lif1 -service-policy default-
data-files
  (network interface modify)

Warning: Assigning service policy "default-data-files" to LIF "lif1" on Vserver
         "vs0" will impact the data services supported by this LIF, which
         requires the LIF to be temporarily brought offline. Data service on
         this LIF will be briefly interrupted while this change is applied, and
         any existing network connections will be reset.
Do you want to continue? {y|n}: y

C1_sti96-vsim-ucs540a_cluster::*> net int show -vserver vs0 -fields data-protocol
(network interface show)
vserver lif  data-protocol
------- ---- ---------------
vs0     lif1 nfs,cifs,fcache

C1_sti96-vsim-ucs540a_cluster::*> iscsi interface show -fields relative-port-id
There are no entries matching your query.

C1_sti96-vsim-ucs540a_cluster::*> net int modify -vserver vs0 -lif lif1 -service-policy default-
data-blocks
  (network interface modify)

Warning: Assigning service policy "default-data-blocks" to LIF "lif1" on
         Vserver "vs0" will impact the data services supported by this LIF,
         which requires the LIF to be temporarily brought offline. Data service
         on this LIF will be briefly interrupted while this change is applied,
         and any existing network connections will be reset.
Do you want to continue? {y|n}: y

C1_sti96-vsim-ucs540a_cluster::*> iscsi interface show -fields relative-port-id
vserver lif  relative-port-id
------- ---- ----------------
vs0     lif1 3

C1_sti96-vsim-ucs540a_cluster::*> snapmirror abort  *
Operation is queued: snapmirror abort for the relationship with destination
"vs1:/cg/smbc_dst_hard1".

Warning: It is recommended to quiesce the relationship using the "snapmirror
         quiesce" command instead of aborting the SnapMirror Synchronous
         transfer. For relationships with a policy of type "strict-sync-mirror"
         this will fail client I/O on source volume if the status is InSync.
Do you want to continue? {y|n}: yes
Operation is queued: snapmirror abort for the relationship with destination "vs1:sync_dst_1".
1 entries were acted on.

C1_sti96-vsim-ucs540a_cluster::*> snapmirror resync *
Operation is queued: snapmirror resync to destination "vs1:/cg/smbc_dst_hard1".
```

Scenario 2

Change the service policy of intercluster LIF to node-mgmt.

```
C1_sti96-vsim-ucs540a_cluster::*> net int show -role intercluster -fields role,service-
policy,services
  (network interface show)
vserver                       lif service-policy       services                           role
----------------------------- --- -------------------- ---------------------------------- -------
-----
C1_sti96-vsim-ucs540a_cluster ic  default-intercluster intercluster-core,management-https
intercluster
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540a_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540b_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
3 entries were displayed.

SMBC_A::*> net int modify -vserver C1_sti96-vsim-ucs540a_cluster -lif ic -service-policy default-
management
  (network interface modify)

C1_sti96-vsim-ucs540a_cluster::*> net int show -role intercluster -fields role,service-
policy,services
  (network interface show)
vserver                       lif                                    service-policy
services                           role
----------------------------- -------------------------------------- -------------------- -------
--------------------------- ------------
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540a_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540b_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
2 entries were displayed.
```

Scenario 3

Change the service policy from management to intercluster.

```
C1_sti96-vsim-ucs540a_cluster::*> net int show -role node-mgmt -fields role,service-
policy,services
  (network interface show)
vserver                       lif service-
role
----------------------------- --- ------------------ ----------------------------------------
------------------------- ---------
C1_sti96-vsim-ucs540a_cluster ic  default-management management-core,management-
autosupport,management-ssh,management-https node-mgmt
C1_sti96-vsim-ucs540a_cluster ic2 default-management management-core,management-
autosupport,management-ssh,management-https node-mgmt
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540a_mgmt1 default-management management-
core,management-autosupport,management-ssh,management-https node-mgmt
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540b_mgmt1 default-management management-
core,management-autosupport,management-ssh,management-https node-mgmt
4 entries were displayed.

C1_sti96-vsim-ucs540a_cluster::*> net int show -role intercluster -fields role,service-
policy,services
  (network interface show)
vserver                       lif                                    service-policy
services                           role
----------------------------- -------------------------------------- -------------------- -------
--------------------------- ------------
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540a_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
C1_sti96-vsim-ucs540a_cluster sti96-vsim-ucs540b_inet4_intercluster1 default-intercluster
intercluster-core,management-https intercluster
2 entries were displayed.

C1_sti96-vsim-ucs540a_cluster::*> net int modify -vserver C1_sti96-vsim-ucs540a_cluster -lif ic2
-service-policy default-intercluster
  (network interface modify)
```

SM-BC Availability

You can check the availability of the SM-BC relationship using a series of commands, either on the primary cluster, the secondary cluster, or both. For example:

```
SMBC_A::*> snapmirror mediator show
Mediator Address Peer Cluster     Connection Status Quorum Status
---------------- ---------------- ----------------- -------------
10.236.172.86    SMBC_B           connected         true

SMBC_B::*> snapmirror mediator show
Mediator Address Peer Cluster     Connection Status Quorum Status
---------------- ---------------- ----------------- -------------
10.236.172.86    SMBC_A           connected         true

SMBC_B::*> snapmirror show -expand
                                                                    Progress
Source            Destination Mirror  Relationship   Total        Last
Path        Type  Path        State   Status         Progress  Healthy Updated
----------- ---- ------------ ------- -------------- --------- ------- --------
vs0:/cg/cg1 XDP  vs1:/cg/cg1_dp Snapmirrored InSync  -         true    -
vs0:vol1    XDP  vs1:vol1_dp  Snapmirrored InSync    -         true    -
2 entries were displayed.

SMBC_A::*> volume show -fields is-smbc-master,smbc-consensus,is-smbc-failover-capable -volume
vol1
vserver volume is-smbc-master is-smbc-failover-capable smbc-consensus
------- ------ -------------- ------------------------ --------------
vs0     vol1   true           false                    Consensus

SMBC_B::*> volume show -fields is-smbc-master,smbc-consensus,is-smbc-failover-capable -volume
vol1_dp
vserver volume  is-smbc-master is-smbc-failover-capable smbc-consensus
------- ------- -------------- ------------------------ --------------
vs1     vol1_dp false          true                      No-consensus
```

For a scenario in which the port hosting the intercluster LIFs is brought down (either admin down or link down), the SM-BC relationship goes "out of sync," causing zero RPO and zero RTO outage.

6. Issue

By default, switches use the hardware proxy mode for resolving L2 unknown unicast addresses where if a destination MAC address is unknown, the packet is forwarded to the spine proxy. And if the spine proxy does not have the destination MAC address information in its database, the packet is dropped. This causes L2 reachability issues for ONTAP behind the ACI switches.

For example, if you have two 2-node ONTAP clusters (primary and secondary) forming a cluster peer relationship using intercluster network interfaces (LIFs), and the intercluster LIFs of all the nodes are connected to Cisco Nexus switches running ACI code. Node 1 of the primary cluster has Ports e0c and e0d hosting IC LIFs IC1 and IC2, respectively. Ports e0c and e0d are in the IC LIF failover group.

If Port e0c is brought down administratively, then the LIF IC1 fails over to Port e0d. ONTAP running on Node 1 notifies the ACI switch that LIF IC1 is now hosted by Port e0d with relevant MAC address updates. However, if this ACI switch is driven by a Cisco Application Policy Infrastructure Controller (APIC) that is configured to use the hardware proxy mode for resolving L2 unknown unicast addresses, then the MAC address update from Node 1 might not get propagated further in the fabric, which causes LIF IC1 as unreachable from the secondary cluster.

7. Resolution

Change the L2 unknown unicast policy on the Cisco APIC from the default hardware proxy mode to using the flooding algorithm for the affected VLAN.

8. Failover Procedure

Failover and failback operations are crucial to the success of a BCDR plan. When disaster strikes, failover is the process of shifting mission-critical workloads from the primary storage system to the secondary storage system in a disaster recovery site. Sometimes, planned failover can be used to assess your disaster recovery configuration or to perform maintenance on the primary cluster. Or "Tick-Tock" deployments, where regulation for financial institutions in some countries requires businesses to periodically be serviceable from their secondary datacenters as well.

Planned Failover

You can perform a planned failover to assess your disaster recovery configuration or to perform maintenance on the primary cluster, initiated by the administrator from the secondary cluster. The operation requires switching the primary and secondary roles so that the secondary cluster takes over from the primary cluster. The new primary cluster can then begin processing input and output requests locally without disrupting client operations. ALUA will also update its paths accordingly.

In ONTAP System Manager, complete the following steps:

Procedure ▶▶▶

1. Select Protection > Overview > Relationships.
2. Hover on the relationship, select ellipse and click on Failover.

◀◀◀

You can monitor status and progress of a planned failover operation using the `snapmirror failover show` command. After the failover operation is complete, you can monitor the synchronous replication protection status from the new destination cluster using the `snapmirror show` command.

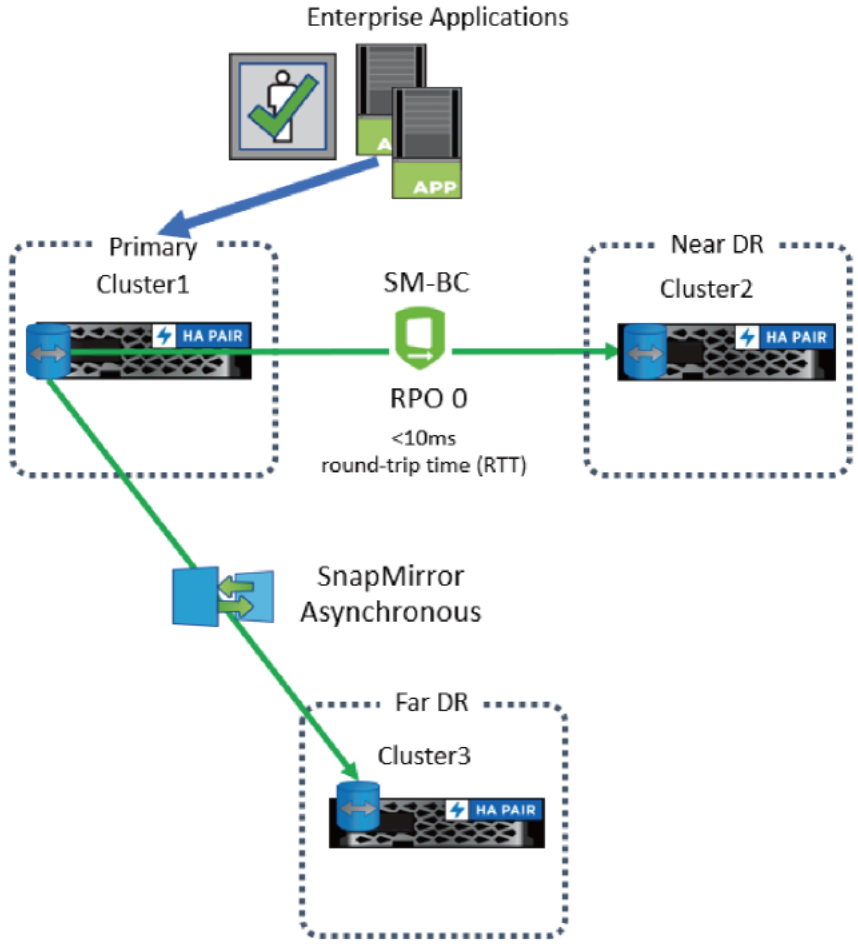
Note

Planned failover cannot begin when a nondisruptive operation including volume move, aggregate relocation, and storage failover is in progress.

Three-Way Topology

Figure 9 illustrates a three-way, fan-out topology.

Figure 9 Fan-out topology



Upon planned failover to Near DR site, creating a new Snapmirror relationship from Near DR site to Far DR site will result in error "Last Transfer Error: No common Snapshot copy found between SVM2:smbc_dest and SVM3:async," which implies the need to perform a rebaseline. To overcome this issue and avoid this rebaseline, run the following command in `diag privilege` (`set diag`) on the Cluster1 before running a planned or negotiated failover:

```
Cluster1::> set diag
Cluster1*::> run -node * setflag skip_cg_css_post_init_resync 1
Cluster1*::> run -node * sm_disable_cg_css_sched 1
```

It is important to disable the above commands in the new secondary storage (after failover), to ensure common snapshots continue to be taken on Cluster1.

```
Cluster1::> set diag
Cluster1*::> run -node * setflag skip_cg_css_post_init_resync 0
Cluster1*::> run -node * sm_disable_cg_css_sched 0
```

Automatic Unplanned Failover

An automatic unplanned failover (AUFO) operation is performed only with assistance from ONTAP Mediator and occurs when the primary cluster is down or isolated and the secondary cluster is converted to the primary and begins serving clients.

Note

After an automatic unplanned failover or an out-of-sync event that exceeds 80 seconds, it is important to rescan the host LUN I/O path to ensure that there is no I/O path loss. For more information, see the respective host OS vendor's documentation on rescan of LUN I/O paths.

During an automatic unplanned failover, you can monitor the status of the operation using the `snapmirror failover show` command.

```
ClusterB::> snapmirror failover show -instance

          Start Time: 9/23/2020 22:03:29
         Source Path: vs1:/cg/scg3
    Destination Path: vs3:/cg/dcg3
     Failover Status: completed
        Error Reason:
            End Time: 9/23/2020 22:03:30
Primary Data Cluster: cluster-2
Last Progress Update: -
       Failover Type: unplanned
  Error Reason codes: -
```

Note

- If the intersite links go down and cause primary isolation where the ONTAP cluster hosting the primary CG is unable to reach its peer cluster and mediator, then any subsequent storage failover will result in disruption. In this situation, Fujitsu recommends that you delete existing LUN maps from the isolated primary and remap again when the disaster recovery is reestablished.
- If a volume move was in progress and AUFO is triggered, you might need to abort the ongoing volume move in case this job is stuck and wait in cutover deferred state forever. The workaround is to abort the volume move instance and restart the volume move job again.

```
# Abort the stuck volume move job
Cluster::> volume move abort -volume <volume name>
# Restart the volume move job
Cluster::> volume move start -volume <volume name> -destination-aggregate <destination aggregate
name>
```

9. Add and Remove Volumes to a Consistency Group (CG)

Changing the composition of the CG is not supported in ONTAP 9.12.1. Therefore, the only way to change the composition is by deleting the original relationship and then creating a new SM-BC relationship with the new composition for the CG.

Note

The new volume you add to expand the CG must have a pair of common Snapshot copies between the source and destination volumes.

10. Single File SnapRestore

Single File Snap Restore is supported starting with ONTAP 9.11.1. This feature provides functionality to do a single file/lun restore from application created Snapshot copies that have been replicated between the SM-BC source (volume) and the destination (Snapshot copy) volumes. Because volumes can contain one or more LUNs, the granularity to restore a single LUN without disrupting other LUNs within the volume is needed. Single File SnapRestore has two options: in-place and out-of-place. The in-place option performs full restorations and overwrites the original file or LUN. The out-of-place option performs full restorations and writes the data to a new file and creates a new LUN.

11. Partial Single File SnapRestore

Partial Single File SnapRestore is supported starting with ONTAP 9.12.1. This feature provides functionality to restore a range of data from application-created Snapshot copies that have been replicated between the SM-BC source (volume) and the destination (Snapshot copy) volumes. A potential use case is to restore a database on a host that stores multiple databases on the same LUN. Using this functionality requires you to know the starting byte offset of the data and byte count.

12. Convert Existing SnapMirror Synchronous Relationship to Protect for Business Continuity

You can convert an existing SM-S relationship to SM-BC, provided all LUNs on the destination volume are unmapped from the application host. Otherwise, creation of the SM-BC protection relationship will fail with error.

13. Upgrading and Reverting ONTAP Versions with SM-BC Relationships

Under certain circumstances, an ONTAP installation with SM-BC relationships can be reverted to a previous version. However, careful consideration must be taken before proceeding. In most cases, existing SM-BC relationships will need to be deleted before reverting.

Fujitsu Storage
ETERNUS AX series All-Flash Arrays
SnapMirror Business Continuity (SM-BC)
ONTAP 9.12.1

C140-0040-01ENZ3

Date of issuance: March 2023
Issuance responsibility: Fujitsu Limited


FUJITSU