HITACHI
Inspire the Next

Unified Compute Platform RS with VMware Cloud Foundation Supports Metro Storage Cluster Datastores from Virtual Storage Platform

Lab Validation Report

MK-SL-239-00
March 2022

Feedback

Hitachi Vantara welcomes your feedback. Please share your thoughts by sending an email message to SolutionLab@HitachiVantara.com. To assist the routing of this message, use the paper number in the subject and the title of this white paper in the text.

Revision history

Changes	Date
Initial release	March 9, 2022

Lab Validation Report

A VMware vSphere Metro Storage Cluster environment on Hitachi Virtual Storage Platform (VSP) family provides an ideal solution for maximizing availability and uptime. It clusters physical data centers within metro distances. This Metro Storage Cluster solution from Hitachi Vantara consists of storage systems presenting replicated storage as a single LUN from different geographically distributed sites. This design enables high availability of services by allowing virtual machine failover between sites with no downtime.

Unified Compute Platform Rack-sale (UCP RS) leveraging VMware Cloud Foundation is the industry's most advanced hybrid cloud platform. It provides a complete set of software-defined services for compute, storage, networking, security, and cloud management to run enterprise apps — traditional or containerized — in private or public environments.

See The Easier Path to the Hybrid Cloud - Using Hitachi Virtual Storage Platform with VMware Cloud Foundation and VMware Virtual Volumes for reference Architecture documentation for UCP RS with VCF and VMFS/vVols.

This report demonstrates that UCP RS with VCF in a stretched cluster configuration supports stretched based datastore volumes using global-active device (GAD) technology from Hitachi Vantara on VSP storage and associated native multipathing.

The following test cases are covered:

- vSAN as primary storage:
  - Validate VMFS on FC as secondary storage with GAD optimized and non-optimized paths.
  - Validate non-optimized path failover when optimized path failure is induced.
- VMFS on Fibre Channel as primary storage:
  - Validate correct deployment of VCF Workload Domain with VMFS on Fibre Channel as primary storage.
  - Validate non-optimized path failover when optimized path failure is induced.

Featured Products

Metro Storage Cluster using GAD volumes in an active/active configuration was tested using the following products.

Hitachi Virtual Storage Platform

The Hitachi Storage Platform 5000 Series was used for both the primary and secondary site volumes.

- Virtual Storage Platform 5600H Primary Site
- Virtual Storage Platform 5600H Secondary Site

See the Interoperability Guide for a list of supported storage systems.

Hitachi UCP RS

The software-defined data center solution from Hitachi, powered by VMware Cloud Foundation was used for the testing environment.

Key hardware components

This section describes the hardware components used in this solution.

Table 1 Primary site

Hardware	Description	Version	Quantity
Hitachi Virtual Storage Platform 5600H	▪ 2 × 32 Gbps Fibre Channel ports (Replication: 6C, 6D) ▪ 1 × 32 Gbps Fibre Channel port (Host: 5D)	90-08-01-00/00	1
Advanced Server DS220	Each with the following configuration: ▪ 2 × Intel Xeon Platinum 8168 24 cores ▪ 256 GB Memory	All components updated to versions supplied with media kit 4.6	3

Hardware	Description	Version	Quantity
	▪ Emulex LPe36000 Fibre Channel Adapter 4-port ▪ Mellanox MT27710 Connectx-4-LX NIC 4-port		
Brocade G620	▪ SAN Switch with 48 × 32 Gbps Fibre Channel ports	FOS v8.2.3	1

Table 2 Secondary site

Hardware	Description	Version	Quantity
Hitachi Virtual Storage Platform 5600H	▪ 2 × 32 Gbps Fibre Channel ports (Replication: 6C, 6D) ▪ 3 × 32 Gbps Fibre Channel port (Host: 5D, 7C, 8C)	90-08-01-00/00	1
Advanced Server DS120	Each with the following configuration: ▪ 2 × Intel Xeon Silver 4210 10 cores ▪ 256 GB Memory ▪ Emulex LPe32000 Fibre Channel Adapter 4-port ▪ Mellanox MT27710 Connectx-4-LX NIC 4-port	All components updated to versions supplied with media kit 4.6	4
Brocade 6510	▪ SAN Switch with 48 × 16 Gbps Fibre Channel ports	V7.3.1d	2

Table 3 External storage for quorum

Hardware	Description	Version	Quantity
Hitachi Virtual Storage Platform G600	Single 50 GB LDEV mapped to primary and secondary storage systems	83-05-39-40/01	1

Key software components

This section describes the software components used in this solution.

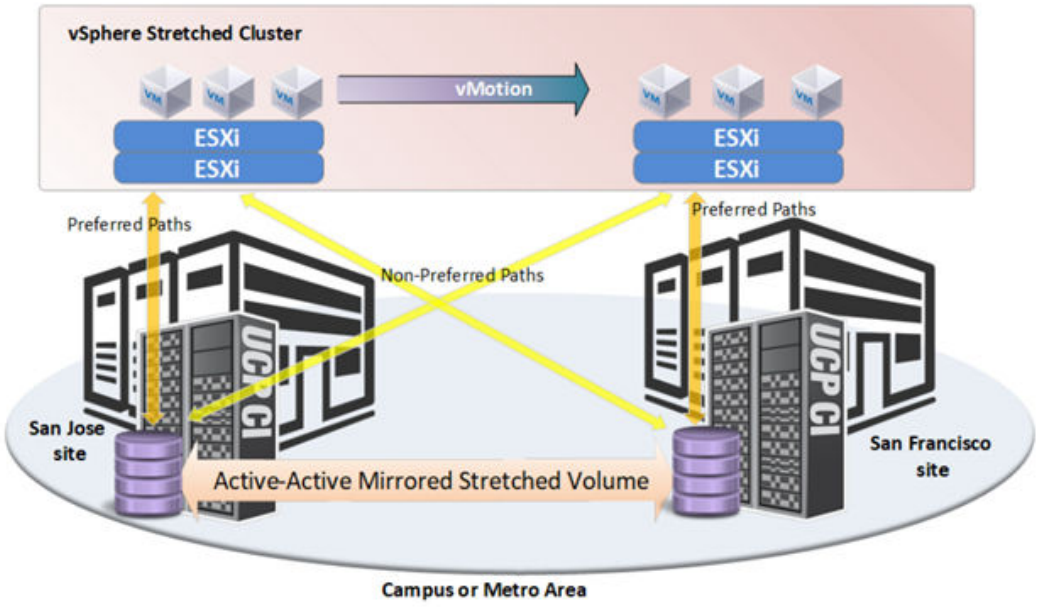
Software	Version
VMware Cloud Foundation (VCF)	4.3.1

The following is the breakdown of software component versions included in VCF 4.3.1.

Software	Version
SDDC Manager	4.3.1.0-18624509
VMware vCenter Server Appliance	7.0U2d
VMware ESXi	7.0U2c

Test environment configuration

The following is a logical diagram of the test environment leveraging UCP RS with VCF.



Hardware details can be found in Key hardware components (on page 4).

See the Unified Compute Platform CI with VMware vSphere Metro Storage Cluster (vMSC) Reference Architecture Guide for more information.

Create GAD replication pairs

Using global-active device, we created an active-active mirrored stretched volume between storage systems. See the Global-Active Device User Guide for more information.

The following methods are available to create and manage GAD replication pairs.

Method	Description	Notes
CCI	The Command Control Interface software enables you to perform storage system configuration and data management operations by issuing commands to the Hitachi RAID storage systems.	Command Control Interface User and Reference Guide
vRealize Orchestrator Storage Connector Plugin	Hitachi Storage Connector for VMware vRealize Orchestrator adds workflows to vRealize Orchestrator that automate configuration and provisioning operations on Hitachi storage systems.	Storage Connector for VMware vRealize Orchestrator User's Guide
Hitachi Ops Center Protector	Ops Center Protector provides a modern, holistic approach to data protection, recovery and retention. It has a unique workflow-based policy engine, presented in an easy-to-use whiteboard-style user interface that helps map copy-data management processes to business priorities. A wide range of fully integrated hardware storage-based and host-based incremental-forever data capture capabilities are included that can be combined into complex workflows to automate and simplify copy-data management.	Hitachi Ops Center Protector User Guide
REST API	Using the REST API, you can get information from the storage system or change the configuration of the storage system. The REST API can perform operations on the following storage systems: ▪ VSP 5000 series ▪ VSP E series ▪ VSP F350, VSP F370, VSP F700, VSP F900 ▪ VSP G350, VSP G370, VSP G700, VSP G900	REST API Reference Guide

Note: The CCI method was used for the test environment.

CCI server

A CCI server is needed to communicate to the replication storage systems and manage replication pairs. There are two methods for the CCI server to establish connectivity to the replication storage systems.

Connectivity Method	Description	Notes
In-band	▪ The command device must be mapped to the SCSI/fibre using LUN Manager. ▪ Can be presented to a VM (must be an RDM device type). ▪ Can be outside of the VCF environment.	▪ This option poses a problem in a VCF environment because Fibre Channel HBAs are atypical on the hosts. ▪ The CCI server should be located in the management domain cluster in a VCF environment because deploying on a VI workload domain could be decommissioned and cause inadvertent teardown of the CCI server.
Out-of-band	▪ This method leverages the IP address of the service processor (SVP) and does not require a Fibre Channel connection. ▪ Requires any host that has an IP route to both SVPs.	▪ For a VCF environment, we leveraged this method of pair management.

Create a remote connection

To connect the primary and secondary storage systems, you must first set the port attributes on both storage systems, physically connect the storage systems, and then add the remote paths between the storage systems.

Depending on the storage subsystem model, there are different requirements for setting the attributes on both the primary and secondary storage systems. See chapter 4 of the Global-Active Device User Guide.

Planning the quorum disk

The quorum disk is used to determine the storage system on which server I/O should continue when a path or storage system failure occurs. There are several methods to implement the quorum disk.

Deployment Method	Description	Notes
On-Prem Externalized Storage	▪ Requires a third storage subsystem. ▪ iSCSI or Fibre Channel implementation.	Global-Active Device User Guide
Amazon Linux VM on AWS Cloud	▪ Removes the need for a third datacenter/storage. ▪ iSCSI ▪ No OS Licensing.	Self-service build guide (Amazon Linux)
SUSE Virtual Machine on Azure Cloud	▪ Removes the need for a third datacenter. ▪ iSCSI ▪ No OS Licensing.	Self-service build guide (Azure SUSE)
AWS/Azure VM Image	▪ Removes need for self-service or manual deployment of VMs. ▪ Ensures correct configuration with fewer manual steps. ▪ Free on AWS/Azure.	Image is available on respective marketplace

Setting up the secondary storage system

For the secondary storage system to correctly mirror the characteristics of the primary storage system, a Virtual Storage Machine (VSM) needs to be created.

- Create a resource group.
- Reserve host group ID.
- Delete virtual LDEV ID of the S-VOL.
- Reserve an LDEV ID for the S-VOL.
- Set the GAD reserve attribute on the S-VOL.
- Create host group.
- Create a pool.
- Create the S-VOL.
- Add an LU path to the S-VOL.

This procedure is documented in Chapter 4 "Setting up the secondary storage system" of the Global-Active Device User Guide.

CCI configuration definition files

In a Global-Active Device environment, there are 4 HORCM configuration files that need to be defined:

- Primary Storage System and P-VOLs
- Secondary Storage System and S-VOLs
- Virtual Storage Machine information for Primary Storage
- Virtual Storage Machine information for Secondary Storage

See "Creating the configuration definition files" in Chapter 4 of the Global-Active Device User Guide.

See the Command Control Interface User's Guide for details.

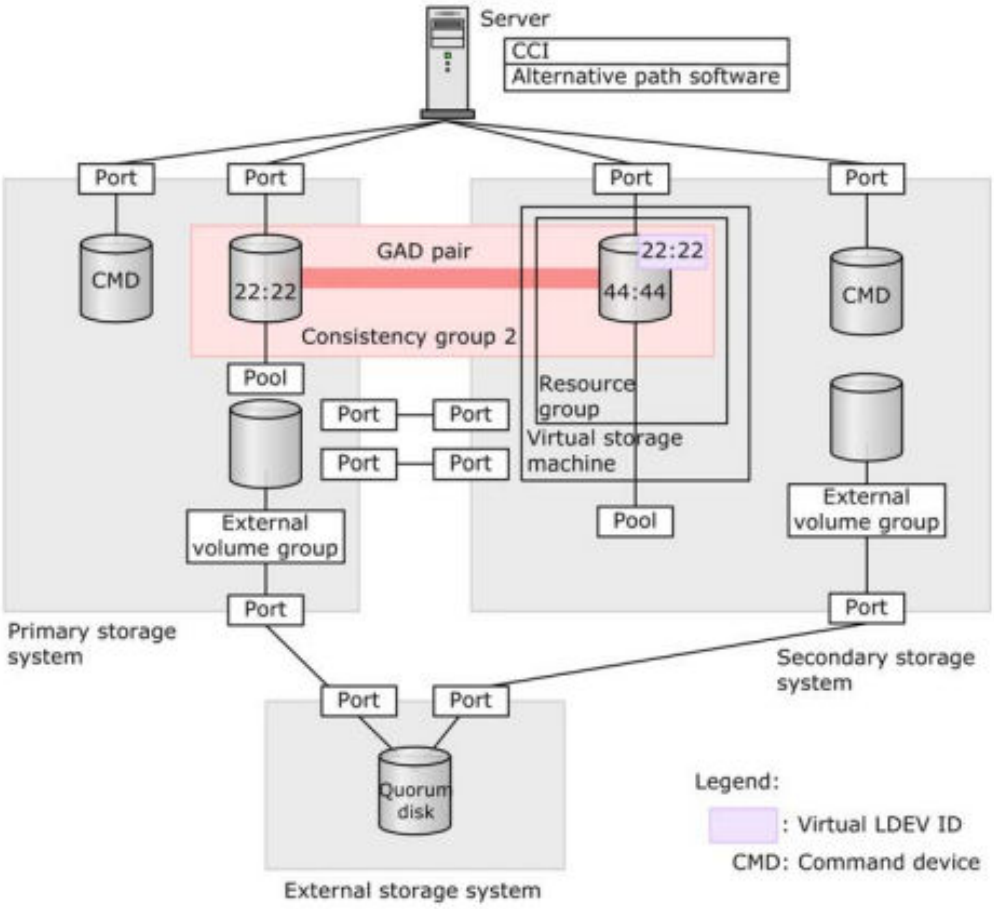
Set the ALUA mode

To specify the preferred path in a cross-path configuration, you must enable the ALUA mode. Before creating a GAD pair, enable the ALUA mode of the P-VOL. If the ALUA mode of the P-VOL is enabled, the ALUA mode of the S-VOL is also enabled when a GAD pair is created.

See "Setting the ALUA mode" section in Chapter 4 of the Global-Active Device User Guide.

Create GAD pairs

When GAD configuration is complete, you can start creating GAD pairs. When a pair is created, the P-VOL LDEV ID is set as the S-VOL's virtual LDEV ID. When the paircreate operation completes, the pair status becomes PAIR, and the P-VOL and S-VOL can accept I/O from the host.



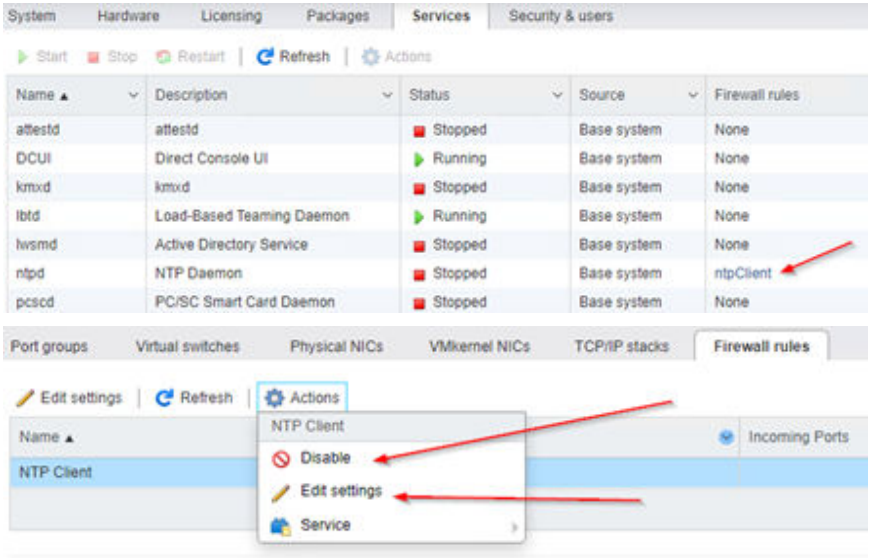
See the Global-Active Device User Guide.

Deploy workload domains

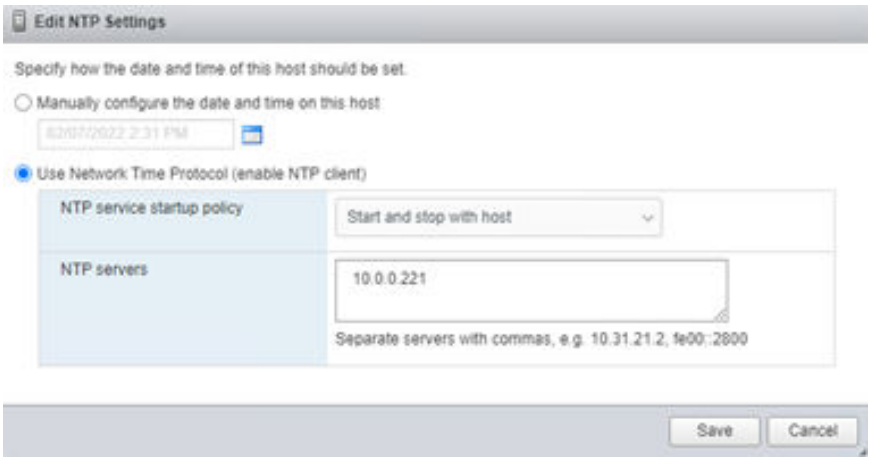
Leveraging the UCP RS with VMware Cloud Foundation SDDC Manager application, VI Workload Domains were deployed to validate the functionality of VCF and GAD. Before leveraging the hosts for deployment there are pre-requisites that need to be met in order to deploy the hosts.

Prepare ESXi Hosts for VMware Cloud Foundation

- DNS — All records must pre-exist before deployment.
  - ESXi hosts
  - vCenter
  - NSX Manager
  - NSX Edge
- Host Preparations
  1. Verify that the ESXi version is supported by SDDC Manager.
  2. Set NTP. There is a caveat to enabling NTP. By default a firewall rule prevents NTP traffic on the host. Either disable the rule or edit the rule to allow the traffic to pass to the NTP source.



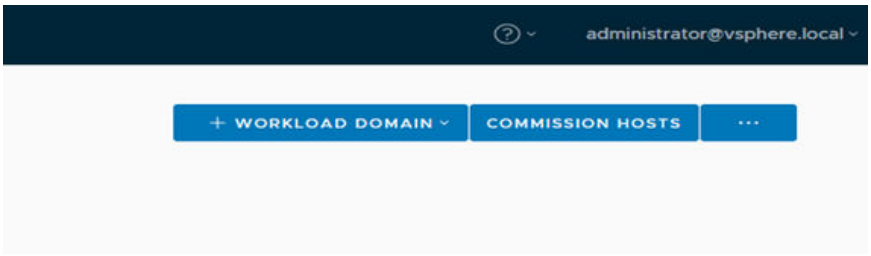
  3. After the firewall rule change is made, set the NTP.



  4. Confirm that NTP is started.

5. Enable SSH.
6. Clear all partitions on storage devices to be used by the vSAN.
7. Assign certificates.
    - Generate a self-signed certificate: Run the following command in the ESXi host / sbin/generate-certificate and then restart the hostd and vpxa services by issuing `/etc/init.d/hostd restart && /etc/init.d/vpxa restart`.
    - External CA-signed certificates: Copy certificates to */etc/vmware/ssl* and rename them to *rui.crt* and *rui.key* respectively and then restart the hostd and vpxa services by issuing `/etc/init.d/hostd restart && /etc/init.d/vpxa restart`.

- Commission Hosts by using the SDDC Manager UI to onboard the ESXi hosts in the upper right hand corner of the Dashboard or the Hosts inventory list.



- After the host commissioning task is complete, deploy a workload domain by clicking the +WORKLOAD DOMAIN button.

Clean up decommissioned hosts

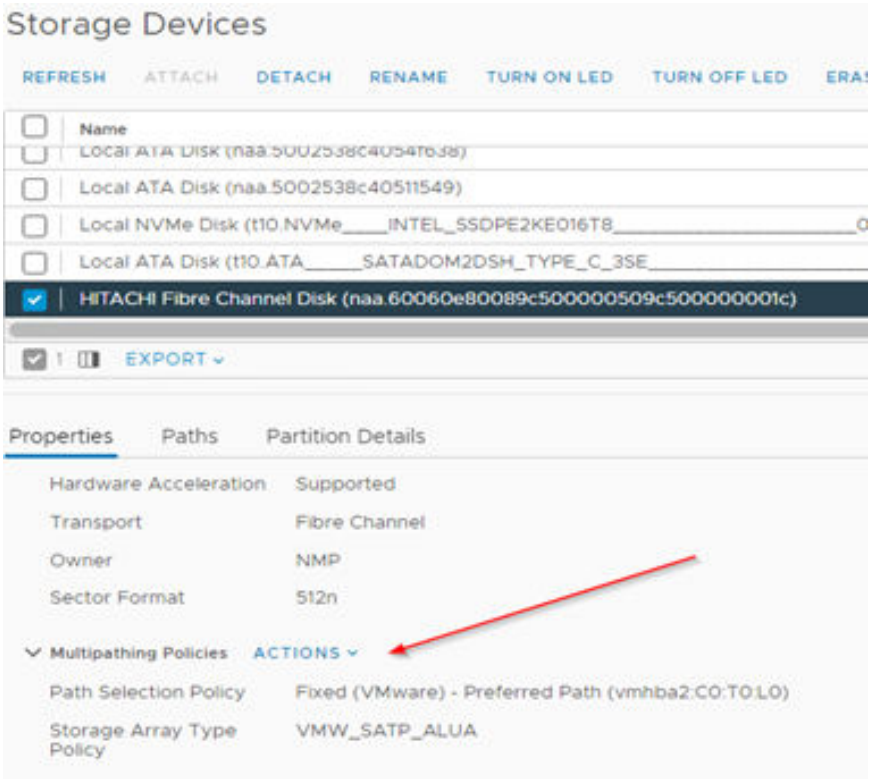
Decommissioning the hosts does not put the host in a state where it can be recommissioned. To clean up decommissioned hosts, follow one of the two methods listed in the VMware documentation at Cleaning up Decommissioned Hosts.

Set the multipathing policy

Use the following procedure to set multipathing.

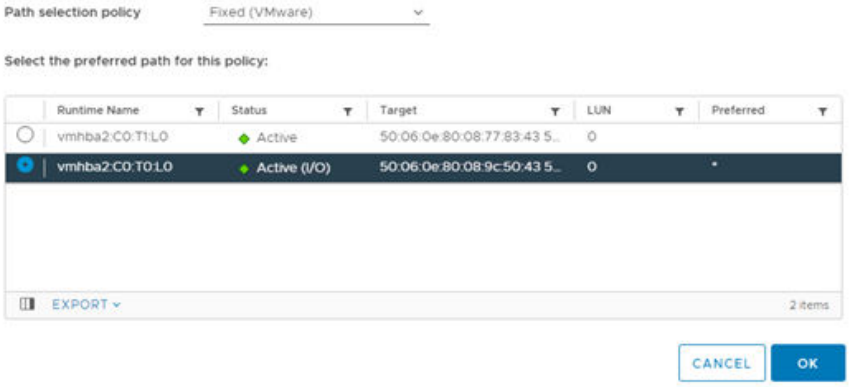
Procedure

1. Navigate to the storage device setting on the ESXi host. Select the GAD LUN and edit the multipathing policy.



2. Click **Actions** and select **Edit Multipathing**.

   You will see the following screenshot. In our example, we used Fixed I/O as we have a single path, but Round Robin and Most Recently Used are supported as well. Hitachi Dynamic Link Manager can also be leveraged to handle the I/O path. In our case, we use VMware's Native Multipathing and we observe that ALUA is functional.



Deploy a test VM

We deployed a VM running Ubuntu 20.04 on the primary cluster to validate the operating state of the VM during our test cases.

Test methodology

The following test cases were used for validation.

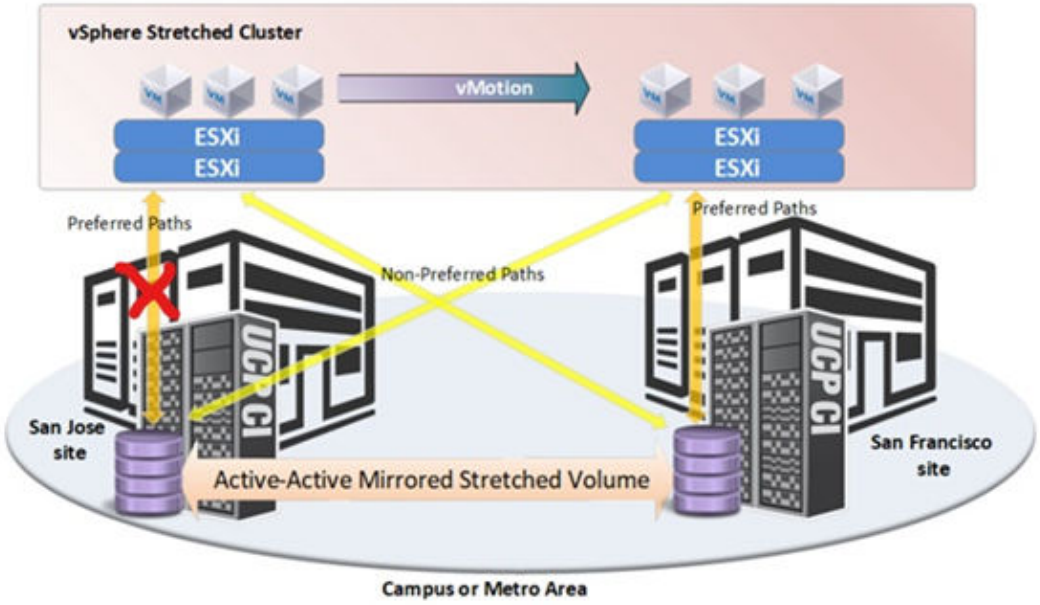
Test Case 1 - vSAN as primary storage

Leveraging an all-flash configuration on the hosts, we deployed a workload with vSAN as the primary storage. We mapped the shared volume to all hosts. ALUA was enabled on the LDEV and native multipathing was used.

In the following figure, we demonstrate that there are two targets identified by the host, yet they are the same volume. This is expected behavior as the Active (I/O) path is the PVOL (Primary) and the Active path represents the SVOL (Secondary).

Runtime Name	Target	LUN	Status
vmhba2:C0:T2:L0	50:06:0e:80:08:77:83:43 50:06:0e:80:08:77:83:43	0	Active
vmhba2:C0:T1:L0	50:06:0e:80:08:9c:50:43 50:06:0e:80:08:9c:50:43	0	Active (I/O)

- We performed a storage vMotion from the VMFS datastore to the vSAN datastore and back to the VMFS datastore without issues.
- We performed a compute only vMotion from the primary site to the secondary site without issues.
- We induce a storage failure by disabling a port on the Fibre Channel SAN switch that handles the traffic to the host on the primary site as shown in the following figure.



- The host detects the path failure and switches the flow of I/O to the non-optimal/preferred path. The host detects the path failure and switches the flow of I/O to the non-optimal/ preferred path.

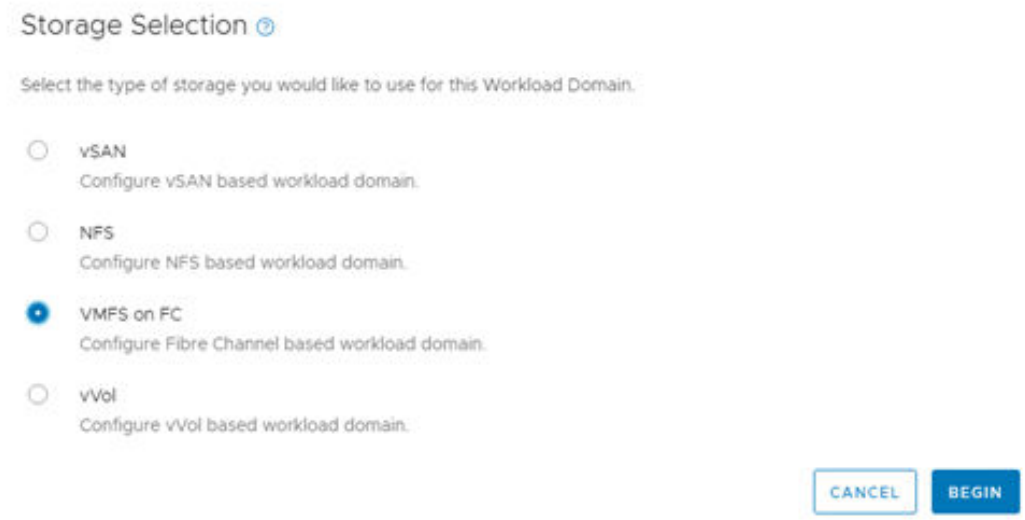
Runtime Name	Target	LUN	Status
vmhba2:C0:T2:L0	50:06:0e:80:08:77:83:43 50:06:0e:80:08:77:83:43	0	Active (I/O)

- After the path is restored, the preferred path is restored as it was before the path failure.

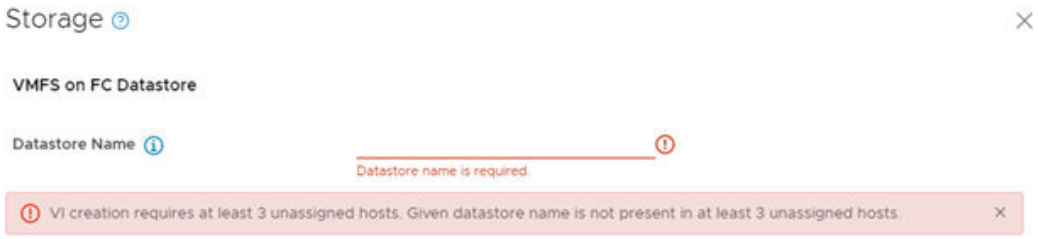
Runtime Name	Target	LUN	Status
vmhba2:C0:T2:L0	50:06:0e:80:08:77:83:43 50:06:0e:80:08:77:83:43	0	Active
vmhba2:C0:T1:L0	50:06:0e:80:08:9c:50:43 50:06:0e:80:08:9c:50:43	0	Active (I/O)

Test Case 2 – VMFS on FC as Primary Storage

We cleaned up the hosts in SDDC Manager to redeploy a new Workload Domain on the primary site. During the deployment process, we select VMFS on FC as primary storage as shown in the following figure.



Note that the GAD-enabled Datastore must exist prior to deployment.



When the deployment was complete, we confirmed that there were two targets identified by the host on the same hosts as seen in Test Case 1.

We performed the same set of tests as Test Case 1 with the same results.

Analysis

Our testing validated the functionality of Global Active Device in a UCP RS environment with VMware Metro Stretch Cluster leveraging VMware Cloud Foundation.

Test Case 1 – Workload domain with vSAN as primary storage using stretched VMFS datastore as secondary storage

In a UCP RS environment with vSAN as primary storage, we validated that Active/Active I/O from two Metro Stretched Clusters were able to complete I/O on both sites concurrently. We demonstrated that a vMotion from one cluster to another while the VM resides on the VMFS datastore does not require storage vMotion because all the I/O is synchronous to the replicated volume on the respective site. We confirmed that a storage or path failure to one site would allow the cluster to resume I/O on the secondary site. We also confirmed that there is no need for the VMFS volume/datastore to be presented to the clusters prior to the deployment of VCF; this means that a GAD volume can be presented to the clusters after the VI workload domain has been already deployed.

Test Case 2 – Workload domain with stretched VMFS datastore as primary storage

As with Test Case 1, we validated that Active/Active I/O from two Metro Stretched Clusters were able to complete I/O on both sites concurrently. We demonstrated that a vMotion from one cluster to another while the VM resides on the VMFS datastore does not require storage vMotion as all the I/O is synchronous to the replicated volume on the respective site. We confirmed that a storage or path failure to one site would allow the cluster to resume I/O on the secondary site. Unlike the vSAN as primary storage use case, in order for a VI workload domain to deploy using VMFS on Fibre Channel as primary storage a datastore created from the VMFS volume must exist and specified during the deployment process.

In both test cases, we validated that non-GAD volumes can co-exist in the environment.

Hitachi Vantara

Corporate Headquarters
2535 Augustine Drive
Santa Clara, CA 95054 USA
HitachiVantara.com | community.HitachiVantara.com

Contact Information
USA: 1-800-446-0744
Global: 1-858-547-4526
HitachiVantara.com/contact