HITACHI
Inspire the Next

Global-Active Device and VMware vSphere Metro Storage Cluster Configuration on Hitachi Storage

Implementation Guide

Poulomi Ganguly
Sayan Basu

August 2020

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

Revision	Changes	Date
v1.0	First Draft	08/18/2020
V1.1	Final PDF	8/27/20

Contents

Feedback 1

Revision History 1

Notices and Disclaimer 3

Executive Summary 4

1. Introduction: 5

1.1 Purpose: 5

1.2 vMSC block diagram: 5

1.3 Hardware requirement: 5

2. Global-Active Device (GAD) vMSC Configuration: 6

2.1 Host configuration: 6

2.2 Switch configuration: 7

2.3 Primary/ Secondary storage configuration: 7

2.4 Quorum storage configuration: 8

2.5 GAD configuration: 9

3. Failure scenarios: 19

Test simulation result of Storage TC ports failure (ALUA/NMP) for a particular Site (A or B): 19

Test simulation result of Storage TC ports failure (HDLM) for a particular Site (A or B): 19

Notices and Disclaimer

This document has been reviewed for accuracy as of the date of initial publication. Hitachi Vantara may make improvements and/or changes in product and/or programs at any time without notice.

The performance data contained herein was obtained in a controlled isolated environment. Results obtained in other operating environments may vary significantly. While Hitachi Vantara has reviewed each item for accuracy, there is no guarantee that similar results can be obtained elsewhere.

All designs, specifications, statements, information and recommendations (collectively, "designs") in this document are presented "AS IS", with all faults. Hitachi Vantara, its affiliates and their respective suppliers disclaim all warranties, including without limitation, the warranty of merchantability, fitness for a particular purpose and non-infringement or arising from a course of dealing, usage or trade practice. In no event shall Hitachi Vantara Corporation, its affiliates or their respective suppliers be liable for any indirect, special, consequential or incidental damages, including without limitation, lost profit or loss or damage to data arising out of the use or inability to use the designs, even if Hitachi Vantara, its affiliates or their respective suppliers have been advised of the possibility of such damages.

Virtual Storage Platform and Global-Active Device are trademarks of Hitachi Vantara.

Other company, product or service names may be trademarks or service marks of others.

WARNING: This document can only be used as Hitachi Vantara internal documentation for informational purposes only. This documentation is not meant to be disclosed to customers or discussed without a proper non-disclosure agreement (NDA).

Executive Summary

This document provides guidelines on how to configure VMware vSphere Metro Storage Cluster using Hitachi Storages.

1. Introduction:

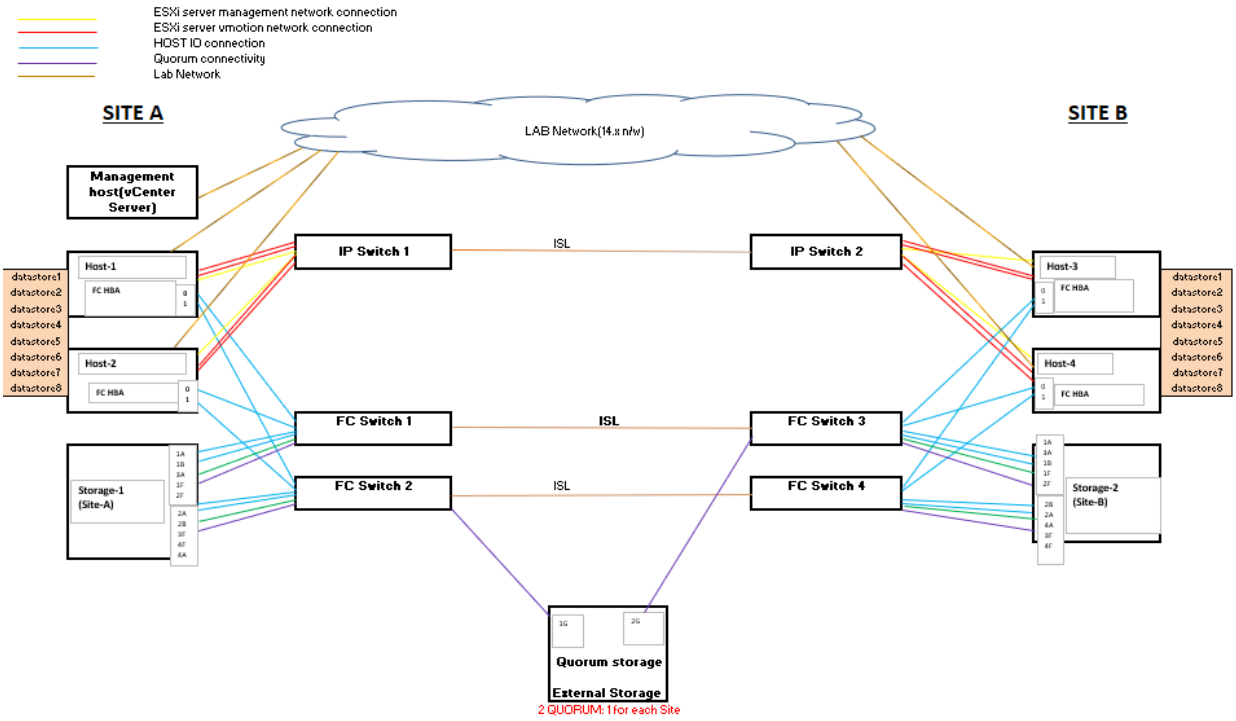
A VMware vSphere Metro Storage Cluster (vMSC) configuration is a specific storage configuration that combines replication with array-based clustering. These solutions are typically deployed in environments where the distance between data centers is limited, often metropolitan or campus environments.

1.1 Purpose:

This article provides step by step information for configuring/deploying a VMware metro storage cluster across 2 data centers using Hitachi Storage Array Platforms. It also includes various failure scenarios in used cases.

1.2 vMSC block diagram:

Figure 1: vMSC block diagram



1.3 Hardware requirement:

Hardware used for vMSC tests must be listed under VMware vCG. The following components create a VMware vSphere Metro Storage Cluster environment:

a. 4 ESXi hosts, 2 in each site, having 4 port NIC card and 2 port FC HBA or CNA as per requirement. All ESXi hosts will be running 2 VMs each (at minimum).

b. 2 *IP switches for management network and vMotion connections between hosts/VMs.

c. 4 *FC switch or iSCSI switch as per requirement for SAN connectivity to the datacenter storage network.

d. 2 *GAD Quorum Disks i.e. iSCSI disk from virtual machine or separate FC/iSCSI storage system such as Hitachi VSP F/G or other supported 3rd-party storage.

e. One management host i.e. vCenter server.

2. Global-Active Device (GAD) vMSC Configuration:

Configuration steps of vMSC with global-active device (GAD) can be divided in 5 areas. They are as follows:

2.1 Host configuration:

- vSphere 6.0U2, 6.5 or 6.7 ESXi host servers — The physical hosts on both data centers running the virtual machines and managed by vCenter Server.
- Host HBA/CNA connectivity to FC/iSCSI switches to be made as per block diagram for SAN connectivity to the datacenter storage network.
- 1 NIC port of 4 port NIC card of each host to be connected to free NIC port of another host on the same site for vMotion connectivity between hosts.
- 2 NIC ports of 4 port NIC card of each host to be connected to iSCSI switches of that site for management network connectivity.
- All the 4 hosts along with their VMs are part of a single vSphere Cluster under a VMware Datacenter.
- 2 VMs will reside in each host and RDM LUNs or vmfs volumes from datastore to be assigned to those VMs as per requirement.
- Host multipathing: All the four ESXi hosts must be configured with multipathing software, either with NMP (VMware Native Multipathing) or with HDLM (Hitachi Dynamic Link Manager) to load-balance I/O between all available preferred & non preferred paths.
- NMP/HDLM multipathing software integrates with GAD to provide load balancing, path optimization, path failover, and path failback capabilities for vSphere hosts.
- When HDLM is chosen as multipathing software, first HDLM for VMware zip file is copied to the required server. Then after unzipping the file, 5 no. of vib files will be there under 5 different folders. All the vib files needs to be installed first to get HDLM installed on the server. Commands are as below:
  - esxcli software vib install -v /<HDLM location>/vib20/hex-hdlm-dlnkmgr/*.vib
  - esxcli software vib install -v /<HDLM location>/vib20/psp-hdlm-exlbk/*.vib
  - esxcli software vib install -v /<HDLM location>/vib20/psp-hdlm-exlio/*.vib
  - esxcli software vib install -v /<HDLM location>/vib20/psp-hdlm-exrr/*.vib
  - esxcli software vib install -v /<HDLM location>/vib20/satp-hdlm/*.vib

Now, to check whether Hitachi LUNs are managed by HDLM, run the following commands:

- esxcli storage nmp device list

Output for a LUN should be like below:

```
Device Display Name: HITACHI Fibre Channel Disk (naa.60060e8008753e000050753e00000133)
   Storage Array Type: HTI_SATP_HDLM
   Storage Array Type Device Config: {device config options }
   Path Selection Policy: HTI_PSP_HDLM_EXLIO
   Path Selection Policy Device Config:
   Path Selection Policy Device Custom Config:
   Working Paths: vmhba65:C0:T0:L0, vmhba64:C0:T0:L0
   Is USB: false
```

- When NMP is chosen as multipathing software, ALUA rules must be set on ESXi hosts prior to version ESXi6.7 update 1 for Hitachi LUNs as below:
  - esxcli storage nmp satp rule add -V HITACHI -M "OPEN-V" -P VMW_PSP_RR -s VMW_SATP_ALUA -c tpgs_on
  - esxcli storage core claimrule load

For ESXi 6.7update1 and above hosts, ALUA rule is already enabled on OS and no additional command is required to enable/configure ALUA on hosts.

2.2 Switch configuration:

- 2 * network switches are configured for LAN connectivity to the datacenter network from each host and VMs (ref. vMSC block diagram on 1.2).
- 4 * FC switches are configured to connect ESXi hosts of both sites to the datacenter storage network(FC). For configuring metro cluster environment, FC switch 1 in site-1 needs to be in cascade with FC switch-3 in site-2, similarly FC switch 2 in site-1 with FC switch-4 in site-2.
- For iSCSI vMSC configuration, 4 iSCSI switches are needed, and they are configured similarly in a cascade switch connection.

2.3 Primary/ Secondary storage configuration:

- 2 * Hitachi storages are configured one in each site and connected to FC/iSCSI switches of respective site as shown in block diagram.
- Each site storage will have 2 sets of LUNs, one set for primary volumes and another set for secondary volumes. All these LUNs will be made available to all the 4 hosts on both sites.
- Each site storage will have two pair of storage ports for MCU(initiator)-RCU(target) pair. These connections will act as the storage replication link between the primary-secondary storages (ref. vMSC block diagram).
- Site-1 storage primary volumes will be in GAD pair with secondary volumes of Site-2 storage. Similarly, Site-2 storage primary volumes will be in GAD pair with secondary volumes of Site-1 storage.
- For NMP-ALUA configurations, ALUA setting must be enabled on PVOLs of both the sites. Following is the command for enabling ALUA on Hitachi LUNs:
  - raidcom modify ldev -ldev_id <ldev_id> -alua enable -fx -IH<horcm_instance>

For example,

- raidcom modify ldev -ldev_id 08:10 -alua enable -fx -IH4545

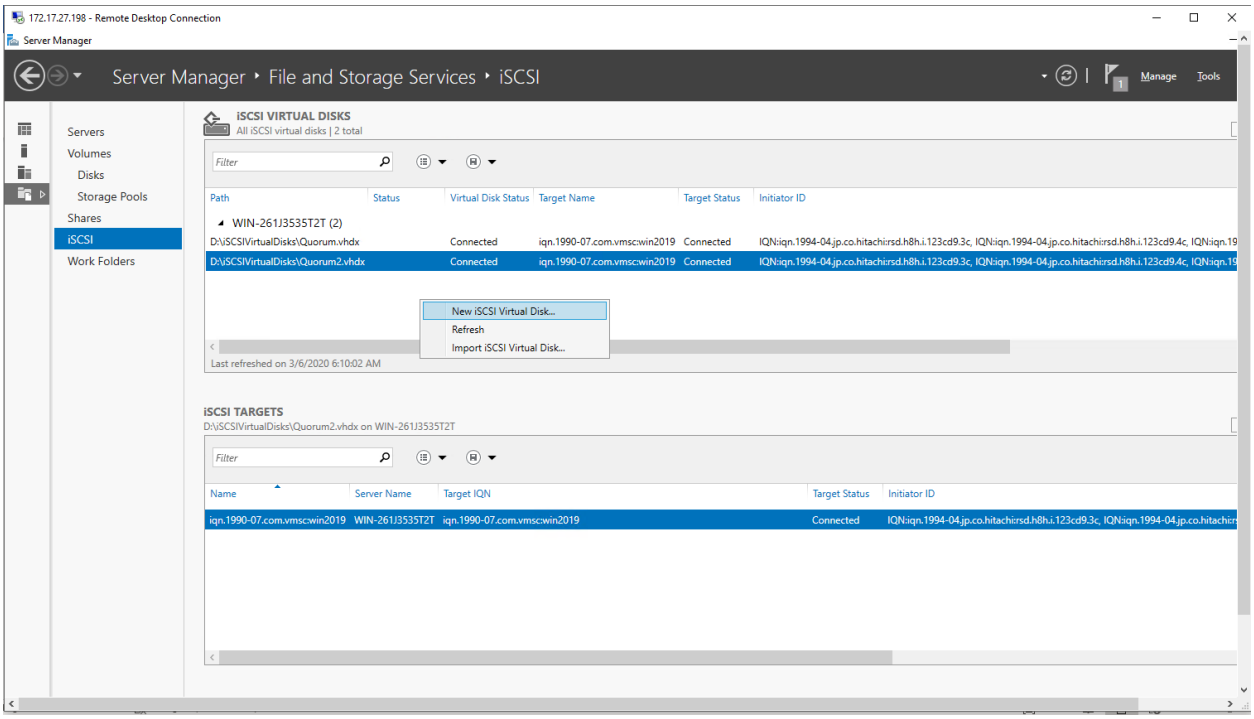
- Also, path optimization settings need to be done on primary and secondary storage hostgroups as mentioned below:
  - raidcom modify lun -port cl1-d HOSTGROUP -lun_id all -asymmetric_access_state optimized -I10 (On PVOL host group)
  - raidcom modify lun -port cl1-d HOSTGROUP -lun_id all -asymmetric_access_state non_optimized -I10 (On SVOL host group)
- HMO78 needs to be set on Host group having SVOLs for all HDLM configurations. This option is not required for NMP-ALUA configurations.

2.4 Quorum storage configuration:

- 2 *quorum disks are assigned for each site GAD pair set.
- Quorum disk can be configured by either assigning an iSCSI disk from the local disk of server hosting a Microsoft Windows Server or separate storage system such as Hitachi VSP F/G or other supported 3rd-party storage

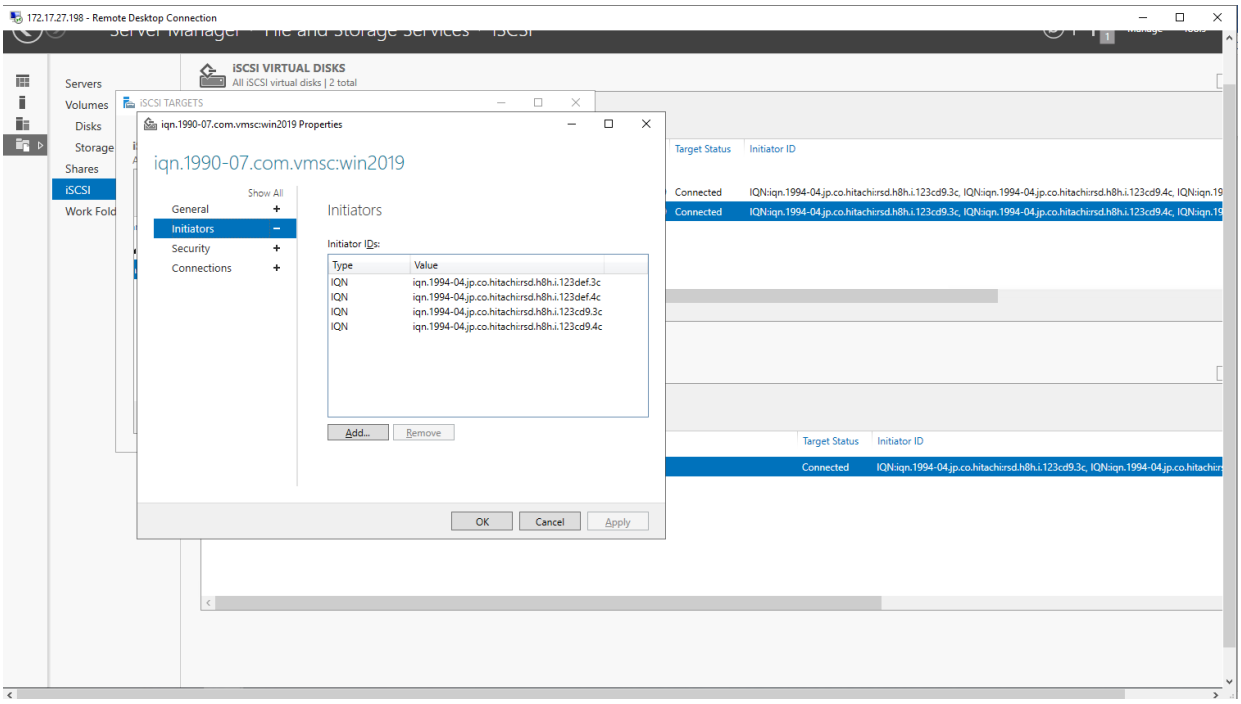
1. Steps to configure Windows Server local disk as iSCSI Quorum disk for VSP storage units Go to Server Manager -> File and Storage Services -> iSCSI -> New iSCSI Virtual Disk. Follow the steps to create iSCSI disks.

Figure 2: Create iSCSI disks



2. Under iSCSi Targets, Right click ‘View all Targets’ ->Properties -> Initiators. Add the IQN of the storage ports that will be used as External Ports for Quorum.

Figure 3: Add External Ports for Quorum



2.5 GAD configuration:

- 2 sets of GAD pair are configured in each site storage for vMSC environment.
- GAD pair can be created using storage GUI or from CCI.
- To create GAD pair from Raid Manager CCI server, steps are as follows:
  - On Primary storage:

1. First, create DP pool for Primary volumes on site-1 storage. Repeat the same steps to create DP pool for Primary volumes on site-2 storage.

   raidcom add dp_pool -pool_name <pool_name> -ldev_id <ldev_id> -I<Primary storage horcm instance>

   Create LDEVs from this DP pool for Primary volumes in site-1 storage. Repeat the same steps to create Primary volumes on site-2 storage. Command to create each LDEV is as follows:

   raidcom add ldev -pool <pool_id> -ldev_id <ldev_id> -capacity <pool_capacity> -I<primary_storage_horcm_instance>

   Format the newly created LDEV:

   raidcom initialize ldev -ldev_id <ldev_id> -operation fmt -I<primary_storage_horcm_instance >

2. Create host group, set host mode options for Primary volumes host group in site-1 storage. Repeat these steps to create Primary volumes host group in site-2 storage as well.

Command to create host group:

raidcom add host_grp -port <host_group_id> -host_grp_name <host_group_name> -I<primary_storage_horcm_instance>

Command to set host mode and host mode options:

raidcom modify host_grp -port <host_group_id> -host_mode 21 -host_mode_opt 54 63 114 78 -I<primary_storage_horcm_instance>

Set Port topology and add HBA port wwns(for all 4 hosts) at Host Group:

raidcom modify port -port <port_id> -port_speed 0 -topology f_port -security_switch y -I<primary_storage_horcm_instance>
raidcom add hba_wwn -port <host_group_id> -hba_wwn <HBA_WWN> -I<primary_storage_horcm_instance>

Assign LDEVs to host group:

raidcom add lun -port <host_group_id> -lun_id 0 -ldev_id <ldev_id> -I<primary_storage_horcm_instance>

- On Secondary storage:

1. First, create resource group of primary storage(site-1) type on secondary storage (site-2) and assign respective secondary resources to it. Then repeat these steps to create resource group of primary storage(site-2) type on secondary storage (site-1) and assign respective secondary resources to it.

raidcom add resource -resource_name <resource_group_name> -virtual_type <primary_storage_serial_number> <storage_model_type> -IH<secondary_storage_horcm_instance>

Reserve the host group ID in the resource group of the storage system at the secondary site.

raidcom add resource -resource_name <resource_group_name> -port <secondary_hostgroup_ID> -IH<secondary_storage_horcm_instance>

Delete the virtual LDEV ID of the volumes from secondary storage which will be used for creating GAD pairs.

raidcom unmap resource -ldev_id <LDEV_ID> -virtual_ldev_id <virtual_LDEV_ID> -IH<secondary_storage_horcm_instance>

Reserve the LDEV ID's in the resource group.

raidcom add resource -resource_name <resource_group_name> -ldev_id <LDEV_ID> -I<secondary_storage_horcm_instance>

Setting the reservation attribute to the volume for the secondary volume of GAD pair.
Set the reservation attribute for GAD to the LDEV ID's

raidcom map resource -ldev_id <LDEV_ID> -virtual_ldev_id reserve -IH<secondary_storage_horcm_instance>

For the LDEV ID to which the reservation attribute was set, ffff is displayed for VIR_LDEV (virtual LDEV ID)

2. Create host group of GAD Secondary site Storage and set Host Mode Options:

raidcom add host_grp -port <Host_group_ID> -host_grp_name <host_group_name> -IH<secondary_storage_horcm_instance>

raidcom modify host_grp -port <Host_group_ID> -host_mode 21 -host_mode_opt 54 63 78 114 -IH<secondary_storage_horcm_instance>

Set Port topology and add HBA port wwn of all the 4 hosts of both sites at Host Group

raidcom modify port -port <Port_ID> -port_speed 0 -topology f_port -security_switch y -IH<secondary_storage_horcm_instance>

raidcom add hba_wwn -port <Host_group_ID> -hba_wwn <HBA_WWN> -IH<secondary_storage_horcm_instance>

3. Create DP Pool and LDEVs for secondary volumes of Site-1 GAD pair. Repeat these steps for creating DP Pool and LDEVs for secondary volumes of Site-2 GAD pair.

raidcom add dp_pool -pool_name <Secondary_pool_name> -ldev_id <Pool_volume_LDEV_ID> -I<secondary_storage_horcm_instance>

Create secondary volumes with the same capacity as the primary volumes.

raidcom add ldev -pool <pool_id> -ldev_id <LDEV_ID> -capacity <LDEV_size> -IH<secondary_storage_horcm_instance>

Adding an LU path to the secondary volume.

raidcom add lun -port <Host_grp_ID> -lun_id 0 -ldev_id <LDEV_ID> -IH<secondary_storage_horcm_instance>

- GAD Pair Creation:

Use the below command in site-1 storage to create site-1 GAD Pair. Repeat this step to create site-2 GAD pair on Site-2 storage.

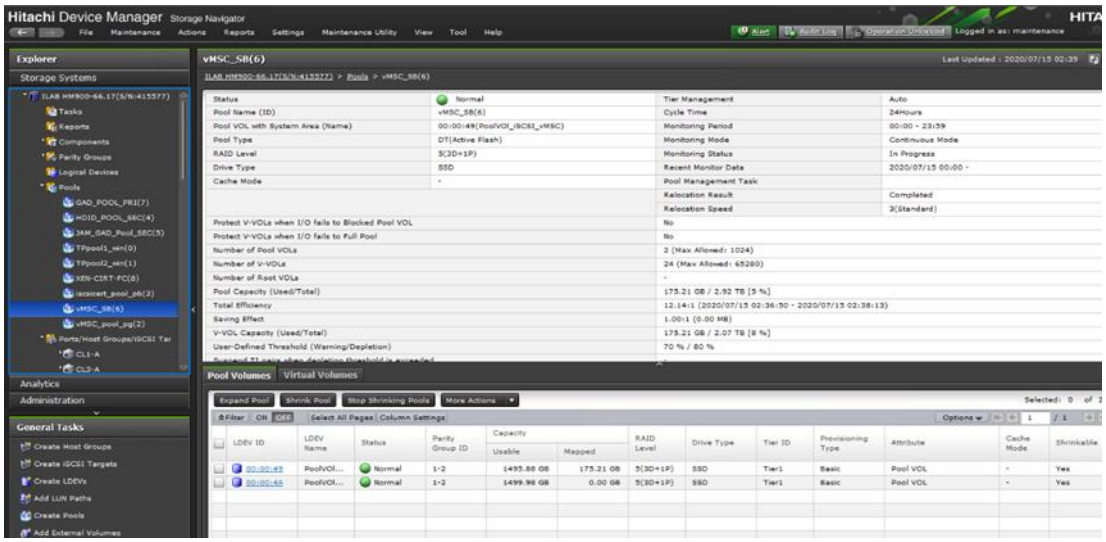
paircreate -g <GAD_PAIR_Name> -f never -vl -jq 7 -IH<primary_storage_horcm_instance>

- To create GAD pair from SVP, steps are as follows:

- On Primary storage:

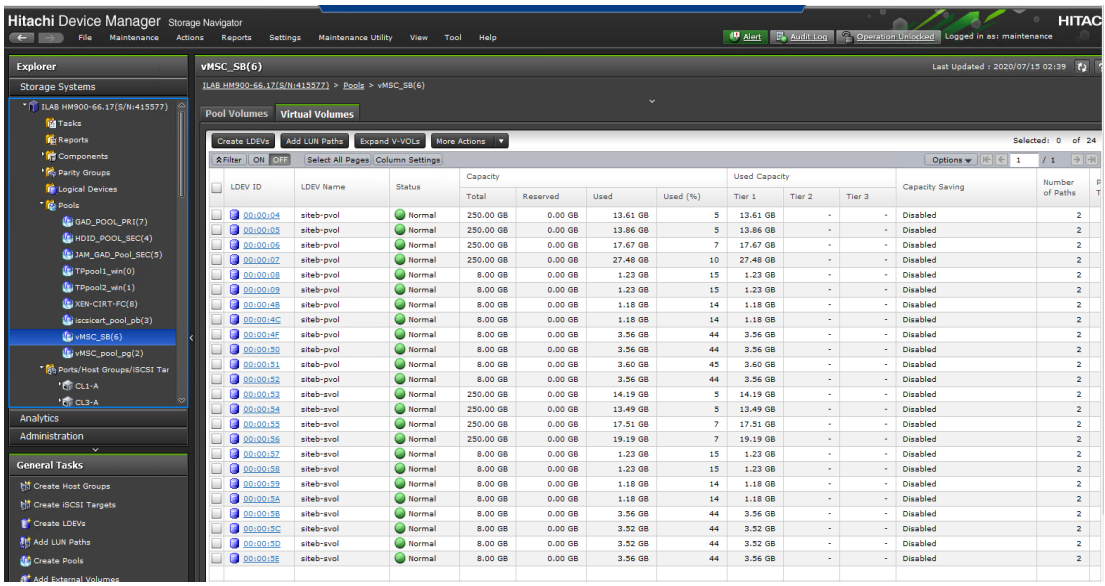
1. First, create DP pool for Primary volumes on site-1 storage. Repeat the same steps to create DP pool for Primary volumes on site-2 storage.

Figure 4: Create DP Pools



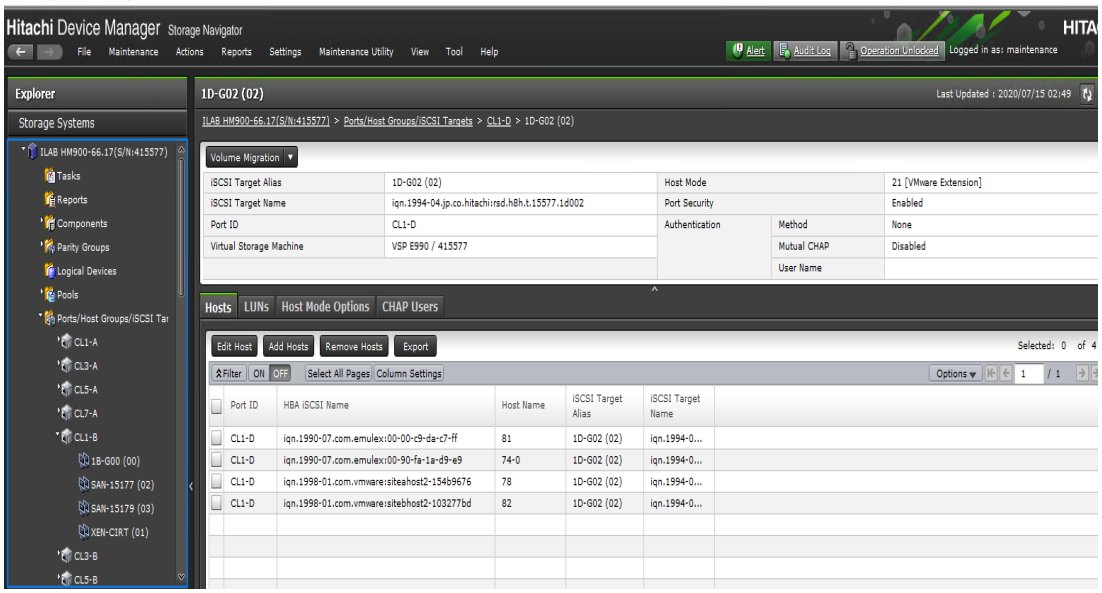
Create LDEVs from this DP pool for Primary volumes in site-1 storage. Repeat the same steps to create Primary volumes on site-2 storage. Command to create each LDEV is as follows:

Figure 5: Create LDEVs



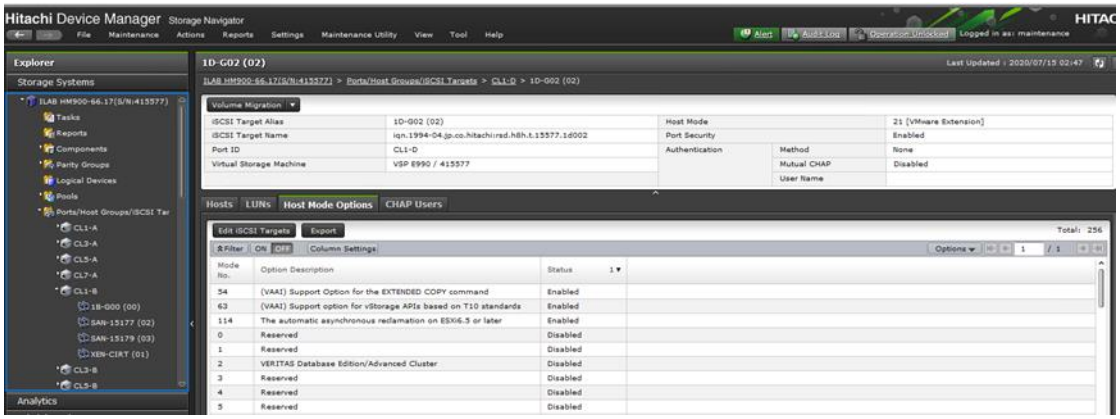
2. Create host group, set host mode options for Primary volumes host group in site-1 storage. Repeat these steps to create Primary volumes host group in site-2 storage as well.

Figure 6: Create Host Groups



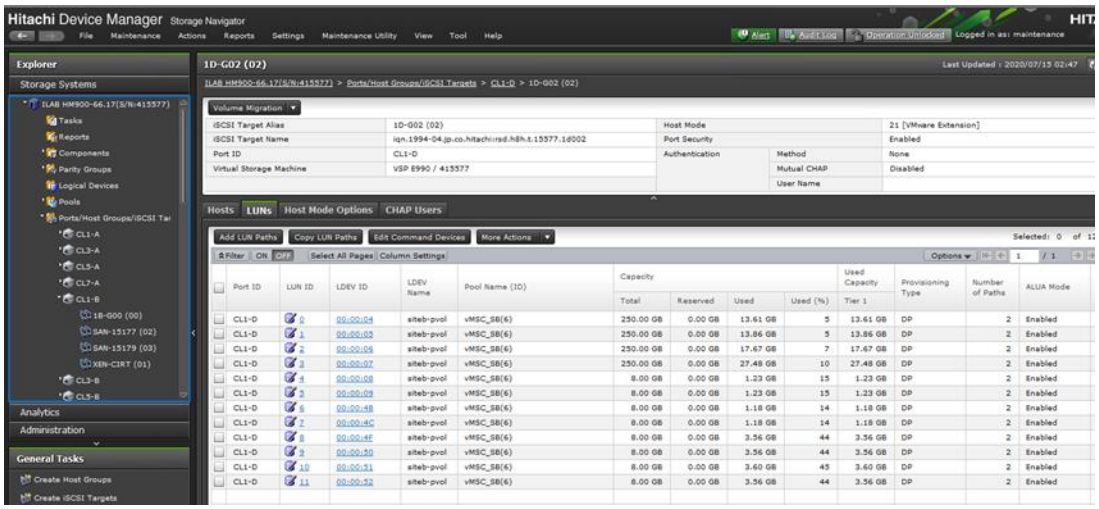
Set host mode and host mode options:

Figure 7: Set Host Mode Options



Assign LDEVs to host group:

Figure 8: Assign LDEVs

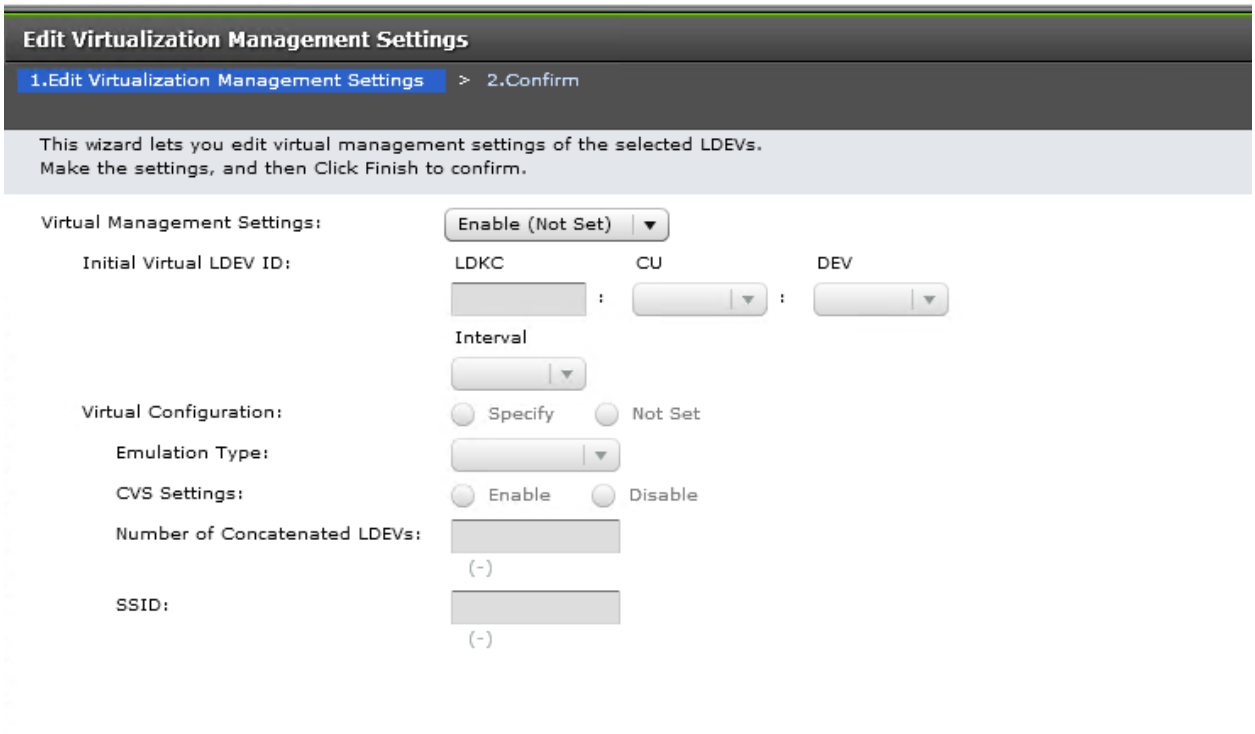


- On Secondary storage:

1. First, create resource group of primary storage(site-1) type on secondary storage (site-2) and assign respective secondary resources to it. Then repeat these steps to create resource group of primary storage(site-2) type on secondary storage (site-1) and assign respective secondary resources to it.

   Delete the virtual LDEV ID of the volumes from secondary storage which will be used for creating GAD pairs.

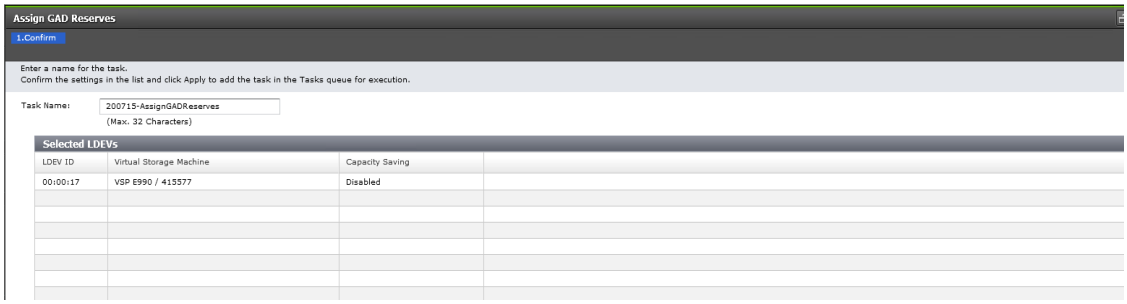
Figure 9: Delete Virtual LDEV ID



Setting the reservation attribute to the volume for the secondary volume of GAD pair.

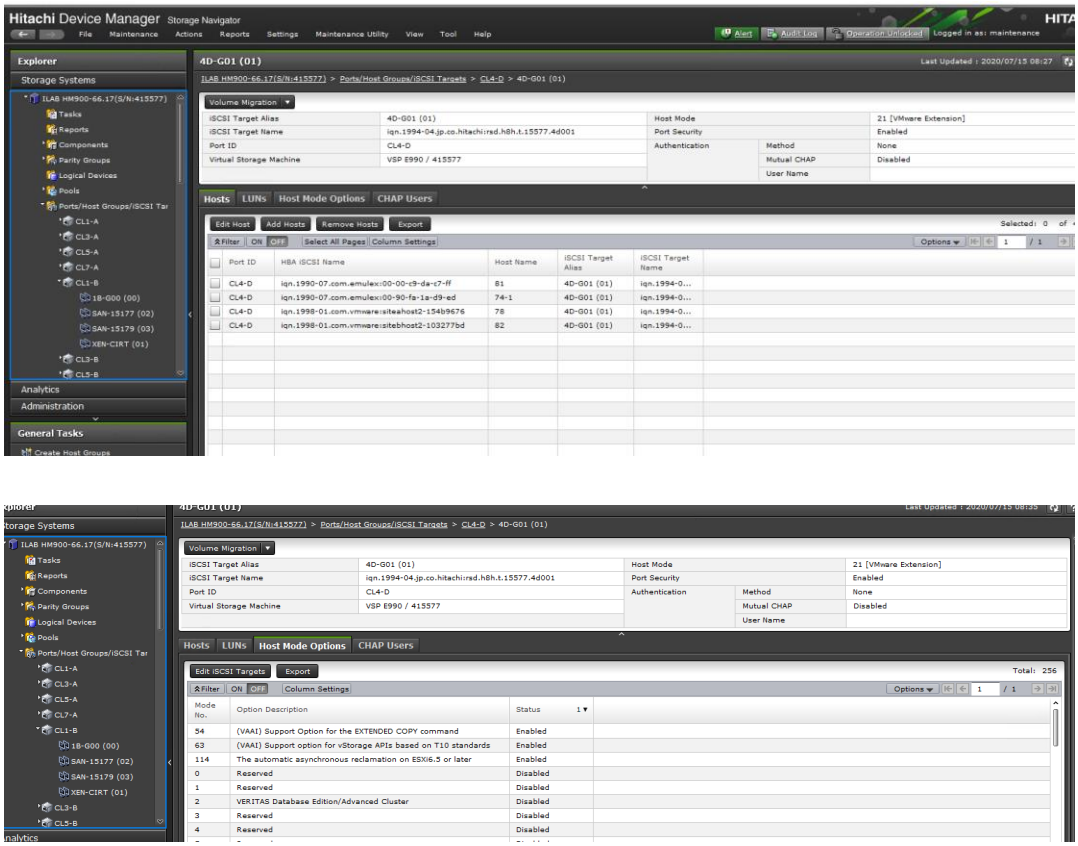
Set the reservation attribute for GAD to the LDEV ID's

Figure 10: Set Reservation Attribute



2. Create host group of GAD Secondary site Storage and set Host Mode Options

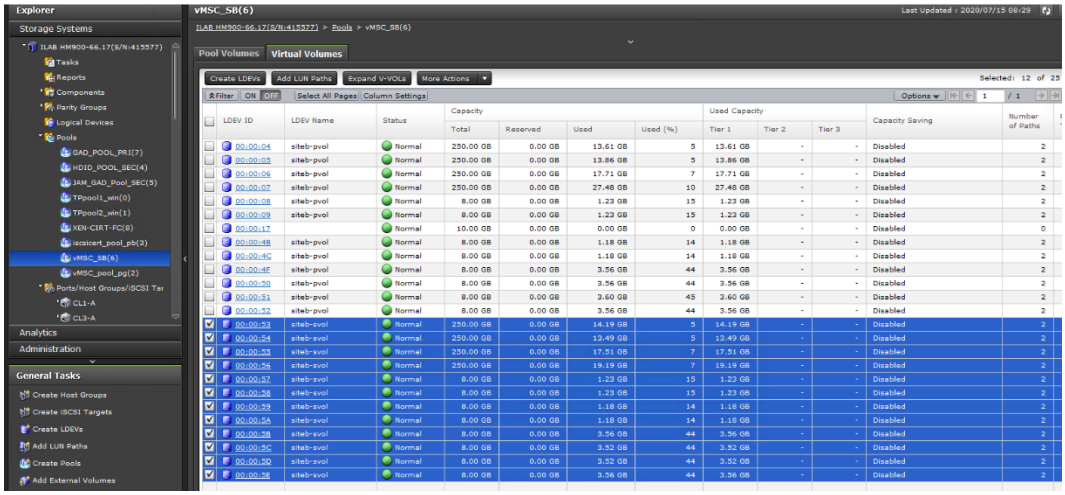
Figure 11: Secondary Site Host Group and Host Mode Options



:

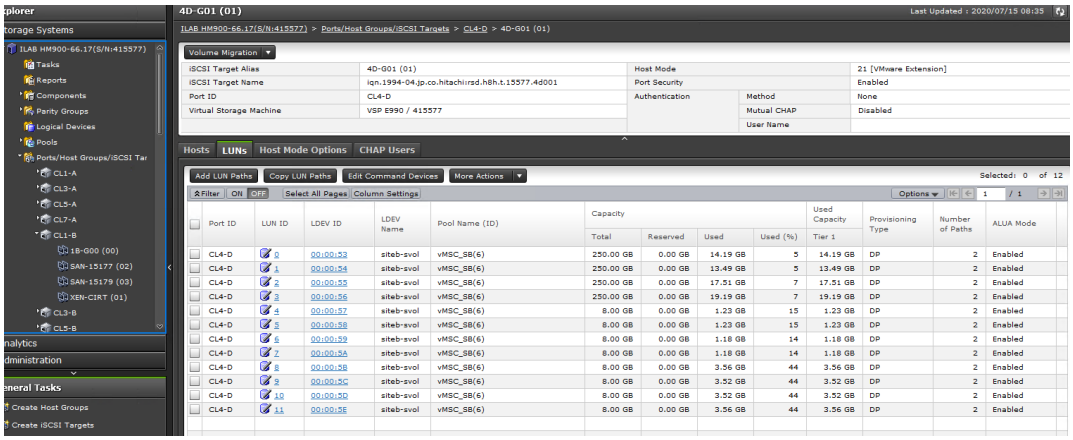
3. Create DP Pool and LDEVs for secondary volumes of Site-1 GAD pair. Repeat these Steps for creating DP Pool and LDEVs for secondary volumes of Site-2 GAD pair. Create secondary volumes with the same capacity as the primary volumes.

Figure 12: Create DP Pool and LDEVs



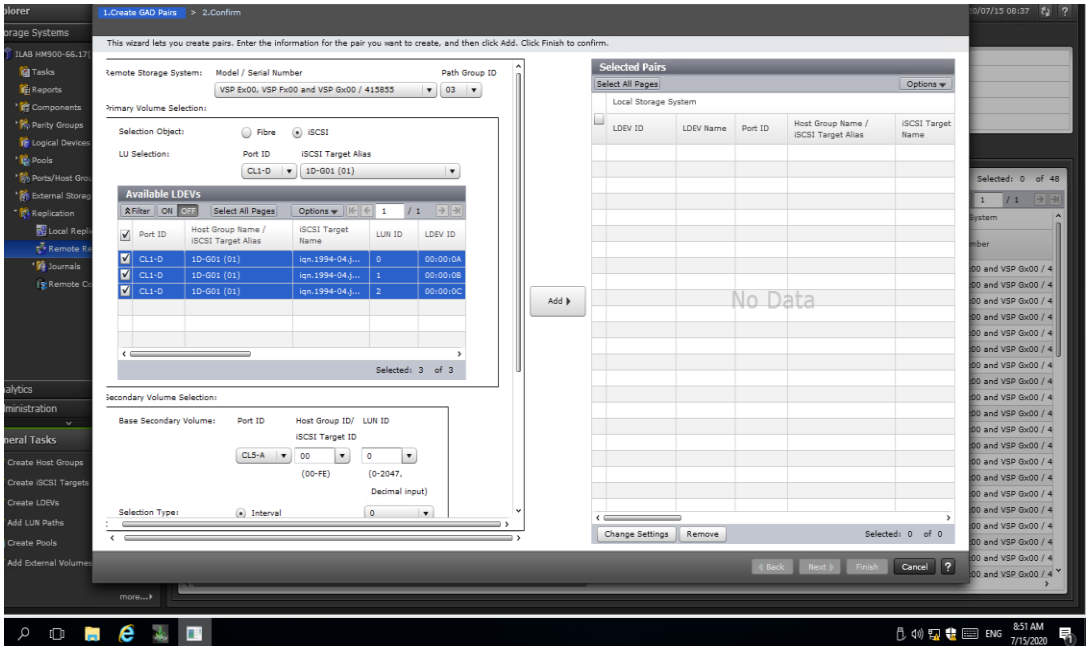
Adding an LU path to the secondary volume.

Figure 13: Add LU Path



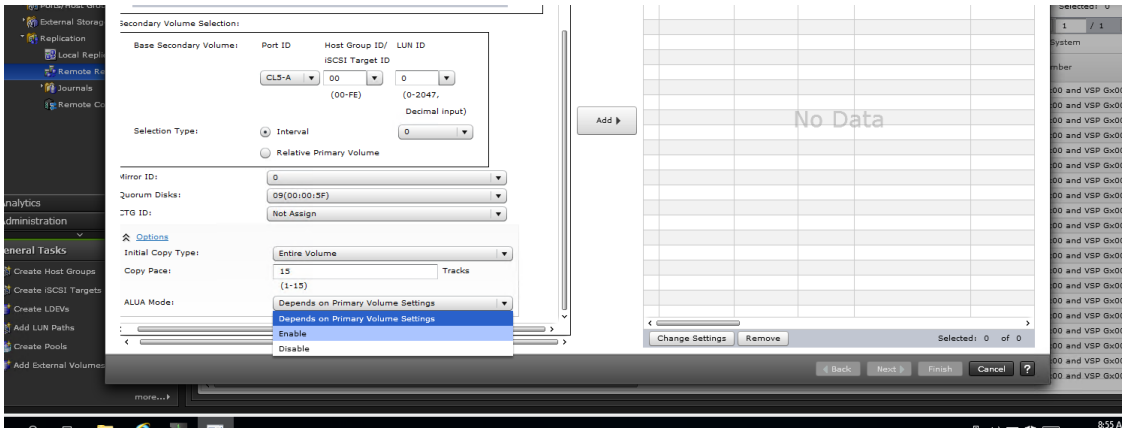
- GAD Pair Creation:

Figure 14: GAD Pair Creation



For NMP-ALUA configuration, while creating ALUA pair ALUA mode must be enabled.

Figure 15: Enable ALUA Mode



3. Failure scenarios:

This section deals with the typical failure scenarios in a GAD vMSC environment and the results for each scenario

Remote connection failure for a particular site

GAD pair behaves differently for failing remote connections for each site.

Test simulation result of Storage TC ports failure (ALUA/NMP) for a particular Site (A or B):

1. Disabled Site A Storage TC ports:
   → Site A PVOLs win and Site B corresponding SVOLs block.
   → Site B PVOLs win and Site A corresponding SVOLs block.

2. Disabled Site B Storage TC ports:
   → Site A PVOLs block and Site B corresponding SVOLs win.
   → Site B PVOLs win and Site A corresponding SVOLs block.

Test simulation result of Storage TC ports failure (HDLM) for a particular Site (A or B):

1. Disabled Site A Storage TC ports:
   → Site A PVOLs win and Site B corresponding SVOLs block.
   → Site B PVOLs win and Site A corresponding SVOLs block.

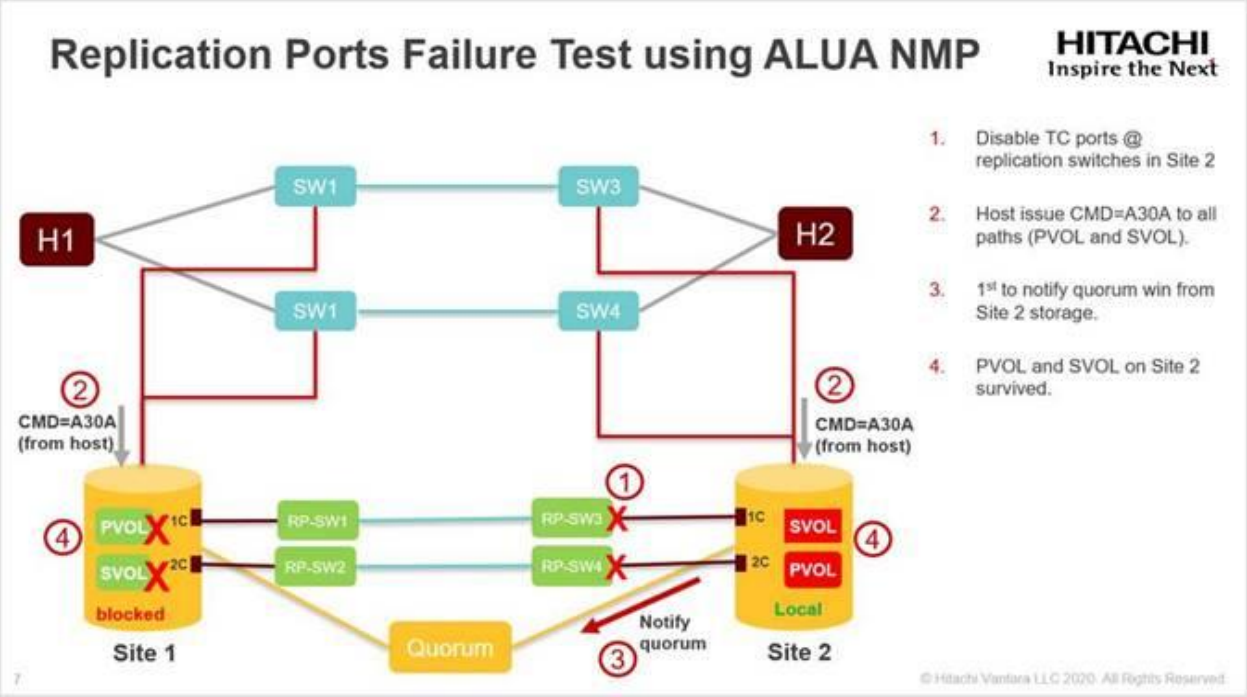
2. Disabled Site B Storage TC ports:
   → Site A PVOLs win and Site B corresponding SVOLs block.
   → Site B PVOLs win and Site A corresponding SVOLs block.

Note:

- For ESXi 6.7, the parameter for action_OnRetryErrors is ON by default.
- For ESXi 6.7U3B, the same parameter is OFF by default.
- When using NMP/ALUA as multipath, set HMO78=OFF
- Ensure for NMP/ALUA, ALUA enabled per LUN / Dedicated Ports for PVOLs and SVOLs enabled with HG Optimized and Non-optimized Paths.

- vSphere GUI on all ESXi hosts showing LUN status: “Active (IO)” à PVOLs and “Active” à SVOLs.
- Zero IOPS observed on SVOLs Storage Ports and generated IO workload observed on PVOLs Storage Ports.
- For NMP/ALUA the host sends CMD=A30A to all the paths, and the storage that notifies Quorum first, survives.
- With HDLM, it was confirmed no ALUA RTPG A3h command send and therefore both PVOLs survived on both storages.

Figure 16: Replication Ports Failure Test



Hitachi Vantara LLC

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

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

P Ganguly August 2020