Hitachi Vantara

Configuring Global-Active Device and VMware vSphere Metro Storage Cluster on a Hitachi Storage System

v3.0

Implementation Guide

This document provides guidelines for configuring GAD and VMware vSphere Metro Storage Cluster (vMSC) using a Hitachi storage system.

Hitachi Vantara

October 2023

© 2022, 2023 Hitachi Vantara LLC. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or stored in a database or retrieval system for commercial purposes without the express written permission of Hitachi, Ltd., or Hitachi Vantara Corporation (collectively, "Hitachi"). Licensee may make copies of the Materials provided that any such copy is: (i) created as an essential step in utilization of the Software as licensed and is used in no other manner; or (ii) used for archival purposes. Licensee may not make any other copies of the Materials. "Materials" mean text, data, photographs, graphics, audio, video and documents.

Hitachi reserves the right to make changes to this Material at any time without notice and assumes no responsibility for its use. The Materials contain the most current information available at the time of publication.

Some of the features described in the Materials might not be currently available. Refer to the most recent product announcement for information about feature and product availability, or contact Hitachi Vantara Corporation at https://support.HitachiVantara.com/ en_us/contact-us.html.

Notice: Hitachi products and services can be ordered only under the terms and conditions of the applicable Hitachi agreements. The use of Hitachi products is governed by the terms of your agreements with Hitachi Vantara Corporation.

By using this software, you agree that you are responsible for:

1) Acquiring the relevant consents as may be required under local privacy laws or otherwise from authorized employees and other individuals to access relevant data; and
2) Verifying that data continues to be held, retrieved, deleted, or otherwise processed in accordance with relevant laws.

Notice on Export Controls. The technical data and technology inherent in this Document may be subject to U.S. export control laws, including the U.S. Export Administration Act and its associated regulations, and may be subject to export or import regulations in other countries. Reader agrees to comply strictly with all such regulations and acknowledges that Reader has the responsibility to obtain licenses to export, re-export, or import the Document and any Compliant Products.

EXPORT CONTROLS - Licensee will comply fully with all applicable export laws and regulations of the United States and other countries, and Licensee shall not export, or allow the export or re-export of, the Software, API, or Materials in violation of any such laws or regulations. By downloading or using the Software, API, or Materials, Licensee agrees to the foregoing and represents and warrants that Licensee is not located in, under the control of, or a national or resident of any embargoed or restricted country.

Hitachi is a registered trademark of Hitachi, Ltd., In the United States and other countries.

AIX, AS/400e, DB2, Domino, DS6000, DS8000, Enterprise Storage Server, eServer, FICON, Flash Copy, IBM, Lotus, MVS, OS/390, PowerPC, RS6000, S/390, System z9, System z10, Tivoli, z/OS, z9, z10, z13, z/VM, BCPii™ and z/VSE are registered trademarks or trademarks of International Business Machines Corporation.

Active Directory, ActiveX, Bing, Excel, Hyper-V, Internet Explorer, the Internet Explorer logo, Microsoft, the Microsoft Corporate Logo, MS-DOS, Outlook, PowerPoint, SharePoint, Silverlight, SmartScreen, SQL Server, Visual Basic, Visual C++, Visual Studio, Windows, the Windows logo, Windows Azure, Windows PowerShell, Windows Server, the Windows start button, and Windows Vista are registered trademarks or trademarks of Microsoft Corporation. Microsoft product screen shots are reprinted with permission from Microsoft Corporation.

All other trademarks, service marks, and company names in this document or web site are properties of their respective owners.

Table of Contents

Table of Contents **3**
Preface **4**
About this document 4
Document conventions 4
Intended audience 4
Revision History 4
Accessing product downloads 4
Getting Help 5
Comments 5
Introduction **6**
Purpose 6
vMSC Block Diagram 6
Hardware Requirements 7
Configuring GAD vMSC **8**
Configuring the Host 8
Configuring the Switch 10
Configuring Primary and Secondary Storage Systems 10
Configuring the Quorum Storage 11
Configuring GAD 12
Creating a GAD Pair 15
Remote Connection Failure Scenarios **22**

Preface

About this document

This document provides guidelines for configuring GAD and VMware vSphere Metro Storage Cluster (vMSC) using a Hitachi storage system.

Document conventions

This document uses the following typographic convention:

Convention	Description
Bold	• Indicates text in a window, including window titles, menus, menu options, buttons, fields, and labels. Example: **Click OK**. • Indicates emphasized words in list items.
Italic	Indicates a document title or emphasized words in text.
`Monospace`	Indicates text that is displayed on screen or entered by the user. Example: `pairdisplay -g oradb`

Intended audience

This document is intended for public.

Revision History

Revision	Changes	Date
v1.0	Initial release	August 2020
v2.0	iSCSI host connection support removal for GAD	October 2022
v3.0	ESXi8.0 support addition	October 2023

Accessing product downloads

Product software, drivers, and firmware downloads are available on Hitachi Vantara Support Connect: https://support.hitachivantara.com/.

Log in and select Product Downloads to access the most current downloads, including updates that may have been made after the release of the product.

Getting Help

Hitachi Vantara Support Connect is the destination for technical support of products and solutions sold by Hitachi Vantara. To contact technical support, log on to Hitachi Vantara Support Connect for contact information: https://support.hitachivantara.com/en_us/contact-us.html.

Hitachi Vantara Community is a global online community for customers, partners, independent software vendors, employees, and prospects. It is the destination to get answers, discover insights, and make connections. **Join the conversation today!** Go to community.hitachivantara.com, register, and complete your profile.

Comments

Please send us your comments on this document to GPSE-Docs-Feedback@hitachivantara.com. Include the document title and number, including the revision level (for example, -07), and refer to specific sections and paragraphs whenever possible. All comments become the property of Hitachi Vantara LLC.

Introduction

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

Purpose

This document provides instructions for configuring a VMware vMSC across two data centers using Hitachi storage systems. Additionally, it includes various failure scenarios based on the use case.

vMSC Block Diagram

The following diagram shows the metro cluster between two datacenters and the quorum storage placed at a third site:

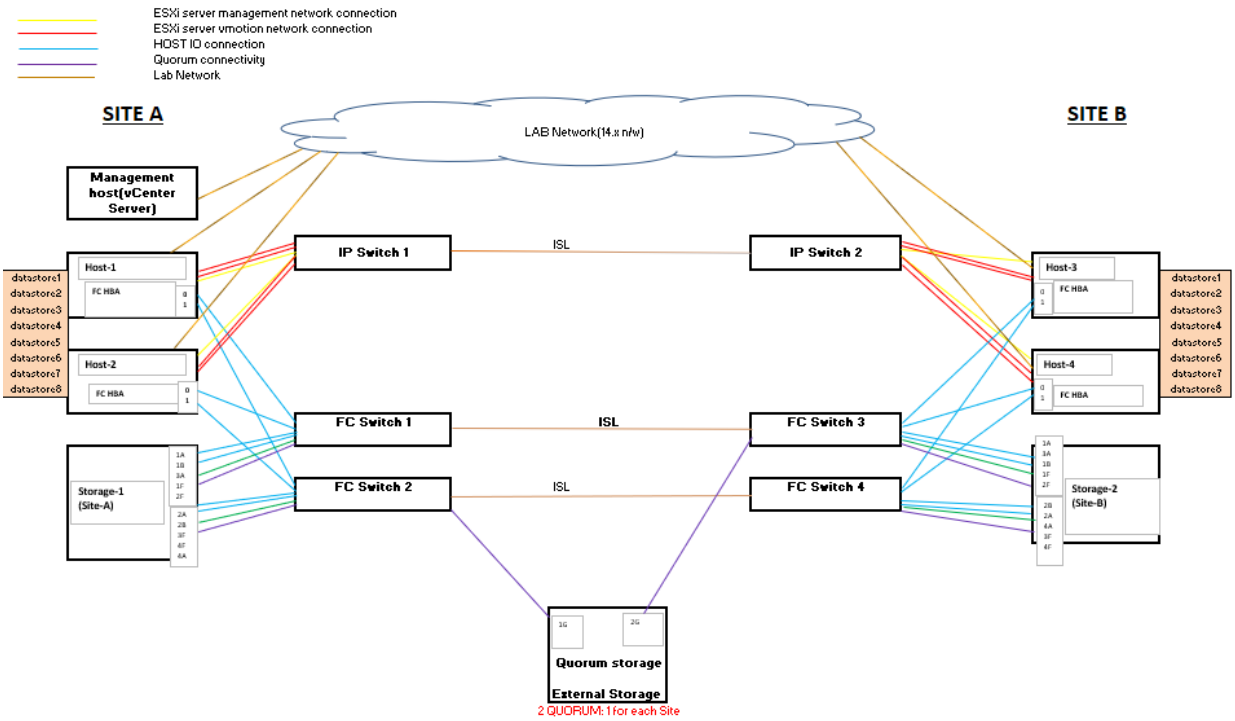

Figure 1: vMSC block diagram

Hardware Requirements

The hardware used for vMSC tests must be listed as supported in the VMware Compatibility Guide (VCG). The following components are required to create a VMware vMSC environment:

- Four ESXi hosts with two in each site, having four port NIC card and two port FC HBA as per requirement. All ESXi hosts will be running two VMs each (minimum).
- Two IP switches for management network and vMotion connections between hosts/VMs.
- Four FC switches for SAN connectivity to the datacenter storage network.
- Two GAD Quorum Disks (iSCSI disk from virtual machine or separate FC/iSCSI storage system such as the Hitachi VSP F/G or other supported third-party storage systems).
- One management host (vCenter server).

Configuring GAD vMSC

Configuring vMSC with GAD can be divided into the following high-level steps:

1. Configuring the Host
2. Configuring the Switch
3. Configuring the primary and secondary storage systems
4. Configuring the Quorum storage
5. Configuring GAD

Configuring the Host

The following components are required for host configuration:

- For vSphere 6.5, 6.7, 7.0 or 8.0 ESXi host servers, the physical hosts running the virtual machines in both data centers must be managed by the vCenter Server.
- For connecting SAN to the datacenter storage network, you must connect the Host HBAs to FC switches.
- For connecting vMotion between hosts, you must connect one NIC port of a 4 port NIC card on each host to the free NIC port of another host at the same site.
- For connecting the management network, you must connect two NIC ports of a 4 port NIC card on each host to iSCSI switches of at the same site.
- ESXi hosts using the iSCSI protocol with GAD is not supported by any Hitachi Virtual Storage Platform (VSP) storage system.
- All four hosts along with the VMs must be a part of a single vSphere Cluster under a VMware Datacenter.
- Two VMs are located in each host. RDM LUNs or Virtual Machine File System (VMFS) volumes from the datastore must be assigned to those VMs.
- Host multipathing: You must configure all four ESXi hosts with multipathing software, either with VMware Native Multipathing (NMP) or with Hitachi Dynamic Link Manager (HDLM) to load-balance I/O between all available preferred and non-preferred paths.
    - NMP or HDLM multipathing software integrates with GAD to provide load balancing, path optimization, path failover, and path failback capabilities for vSphere hosts.
- When HDLM is selected as the multipathing software, the HDLM for VMware zip file is copied to the required server. Then, after unzipping the file, five vib files will be present under five different folders. You must install all vib files first to install HDLM on the server as follows:
    - `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`

- To verify whether Hitachi LUNs are managed by HDLM, run the following command:

  ```
  esxcli storage nmp device list
  ```

  The following shows the output for a LUN:

  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 selected as the multipathing software, you must set ALUA rules on ESXi hosts prior to version ESXi 6.7 update 1 for Hitachi LUNs as follows:

  ```
  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 6.7 update 1 and later hosts, the ALUA rule is already enabled on the OS and no additional command is required to enable/configure ALUA on hosts.

Configuring the Switch

The following components are required for switch configuration:

- For connecting LAN to the datacenter network from each host and VM, two network switches must be configured.
- For connecting ESXi hosts of both sites to the datacenter storage network (FC), four FC switches must be configured.
- For configuring the metro cluster environment, FC switch 1 in site-1 must be in cascade with FC switch-3 in site-2, and similarly, FC switch 2 in site-1 must be in cascade with FC switch-4 in site-2.

Configuring Primary and Secondary Storage Systems

The following components are required for primary and secondary storage system configuration:

- Two Hitachi storage systems configured (one in each site) and connected to FC/iSCSI switches of the respective site.
- Remote connectivity for GAD is supported for both the FC and iSCSI protocol.
- Each site storage must have two pairs of LUNs, one set for primary volumes and another set for secondary volumes. These LUNs will be made available to all the four hosts on both sites.
- Each site storage must have two pairs of storage ports for MCU (initiator) and RCU (target) pair. These connections act as the storage replication link between the primary and secondary storage systems.
- Site-1 storage primary volumes are in GAD pair with secondary volumes of the site-2 storage system. Similarly, site-2 storage primary volumes are in GAD pair with secondary volumes of the site-1 storage system.
- For NMP-ALUA configurations, the ALUA setting must be enabled on the P-VOLs for both sites. For enabling ALUA on Hitachi LUNs, run the following command:

  ```
  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
  ```

- Path optimization settings must be configured on the primary and secondary storage host groups as follows:

  ```
  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 must be set on the Host group having SVOLs for all HDLM configurations. This is not required for NMP-ALUA configurations.

Configuring the Quorum Storage

The following components are required for Quorum storage configuration:

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

To configure a Windows Server local disk as an iSCSI Quorum disk for VSP storage systems, complete the following steps:

1. Navigate to **Server Manager > File and Storage Services > iSCSI > New iSCSI Virtual Disk**.

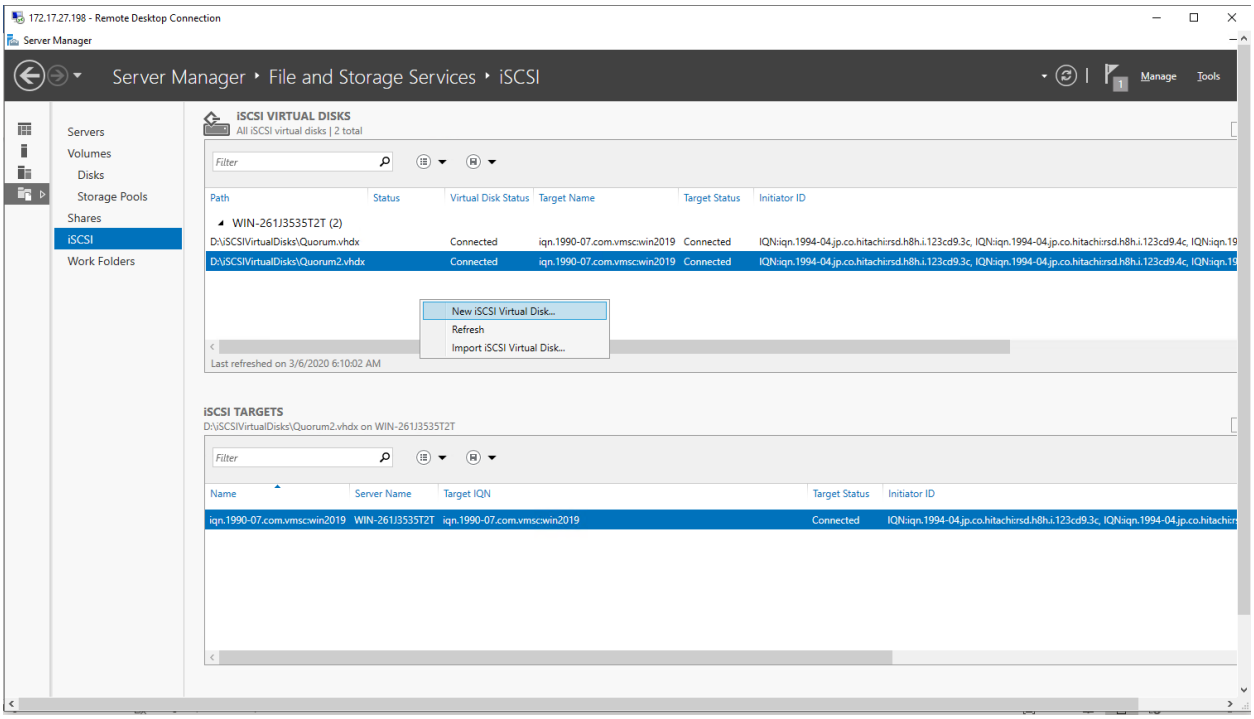

Figure 2: Creating iSCSI disks

2. Under iSCSi Targets, right-click **View all Targets** > **Properties** > **Initiators** and add the IQN of the storage ports that will be used as External Ports for the Quorum.

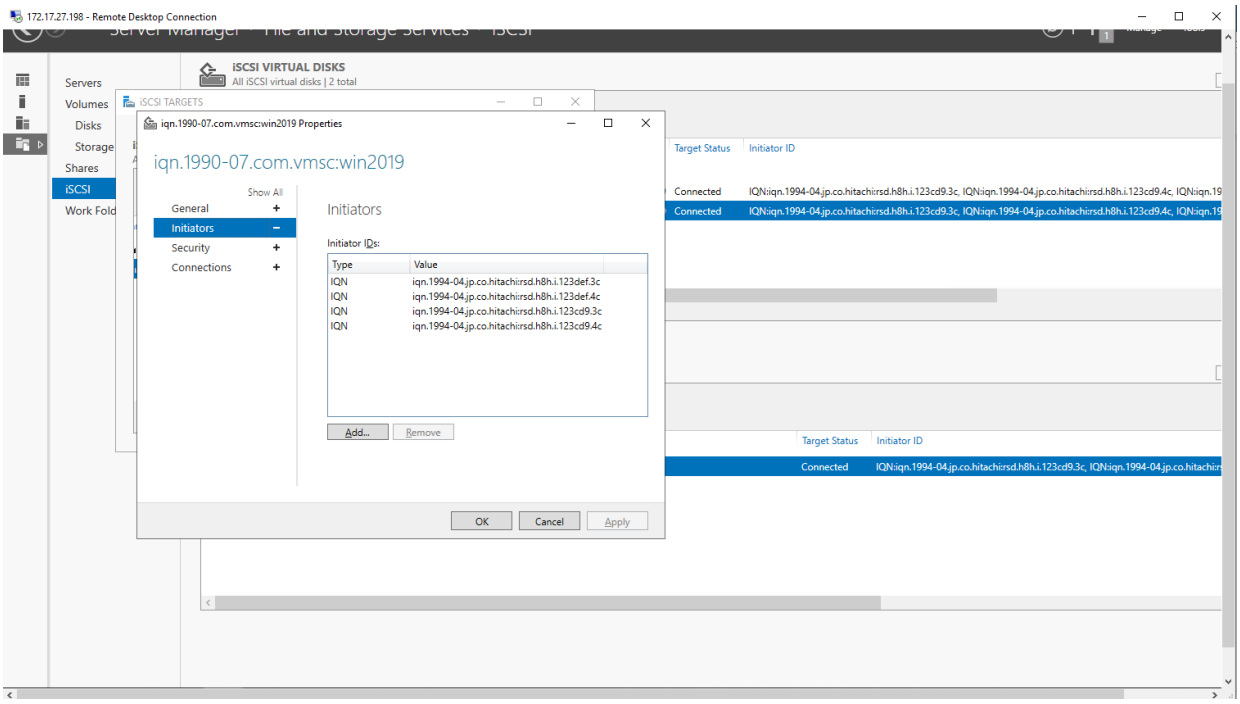

Figure 3: Adding External Ports for Quorum

Configuring GAD

The following components are required for GAD configuration:

Two sets of GAD pairs configured using the storage UI or CCI for each site storage for the vMSC environment.

To create a GAD pair from Raid Manager CCI server on the primary storage system, complete the following steps:

1. Create a DP pool by running the following command for primary volumes in the site-1 storage system. Repeat the procedure to create a DP pool for primary volumes in the site-2 storage system.

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

2. Create LDEVs from this DP pool for primary volumes in the site-1 storage system by running the following command. Repeat the procedure to create primary volumes in the site-2 storage system.

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

3. Format the newly created LDEV by running the following command:

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

4. Create a host group and set host mode options for primary volumes host group in the site-1 storage system. Repeat the procedure to create primary volumes host group in the site-2 storage system.

5. To create a host group, run the following command:

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

6. To set host mode and host mode options, run the following command:

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

7. To set Port topology and add HBA port wwns (for all four hosts) at Host Group, run the following commands:

```
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>
```

8. To assign LDEVs to the host group, run the following command:

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

To create a GAD pair from Raid Manager CCI server on the secondary storage system, complete the following steps:

1. Create a resource group of the primary storage system (site-1) on the secondary storage system (site-2) and assign the respective secondary resources. Repeat the procedure to create a resource group of the primary storage system (site-2) on the secondary storage system (site-1) and assign the respective secondary resources. Run the following command:

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

2. Reserve the host group ID in the resource group of the storage system at the secondary site by running the following command:

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

3. Delete the virtual LDEV ID of the volumes from the secondary storage system that will be used for creating GAD pairs by running the following command:

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

4. Reserve the LDEV IDs in the resource group by running the following command:

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

5. Set the reservation attribute for GAD to the LDEV IDs by running the following command:

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

6. For the LDEV ID where the reservation attribute was set, ffff is displayed for VIR_LDEV (virtual LDEV ID)

   You can verify this by running the following command:

```
raidcom get ldev -ldev_id <LDEV_ID> -fx -
IH<secondary_storage_horcm_instance>
```

7. Create a host group of the GAD secondary site storage system and set Host Mode Options by running the following commands:

```
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>
```

8. Set Port topology and add HBA port wwn of all the four hosts of both sites at the Host Group by running the following command:

```
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>
```

9. Create DP pool and LDEVs for secondary volumes of Site-1 GAD pair by running the following command. Repeat the procedure to create 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>
```

10. Create secondary volumes with the same capacity as the primary volumes by running the following command:

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

11. Add an LU path to the secondary volume by running the following command:

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

Creating a GAD Pair

To create a GAD Pair in site-1, run the following command in the site-1 storage system. Repeat the procedure to create site-2 GAD pair on the site-2 storage system.

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

To create GAD pair from SVP on the primary storage system, complete the following steps:

1. Create a DP pool for primary volumes on the site-1 storage system. Repeat the procedure to create a DP pool for primary volumes on the site-2 storage system.

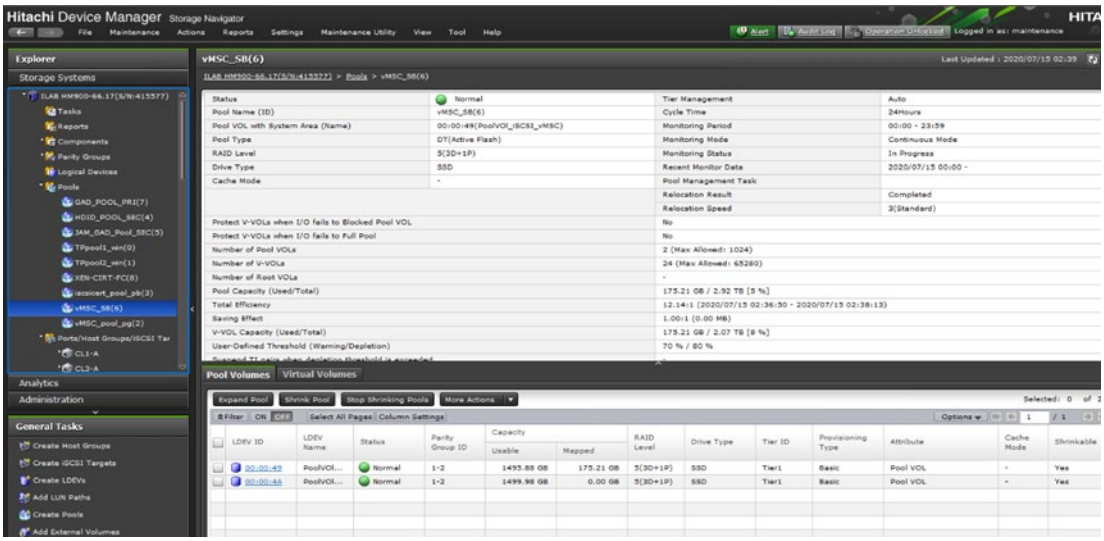

Figure 4: Creating DP Pools

2. Create LDEVs from this DP pool for primary volumes in the site-1 storage system. Repeat the procedure to create primary volumes in the site-2 storage system.

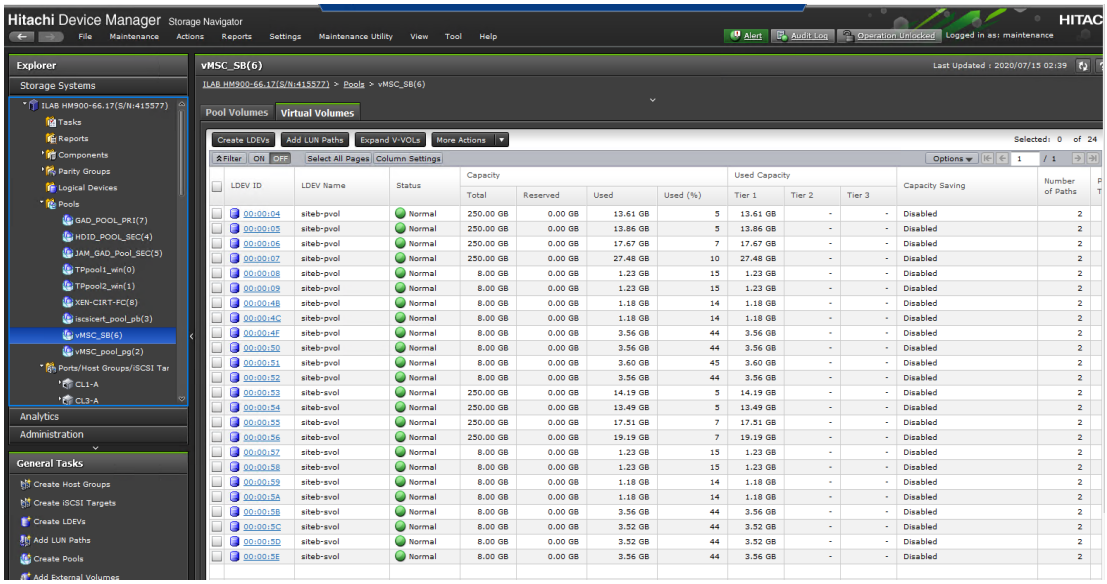

Figure 5: Creating LDEVs

3. Create a host group and set the host mode options for primary volumes host group in the site-1 storage system. Repeat the procedure to create primary volumes host group in the site-2 storage system.

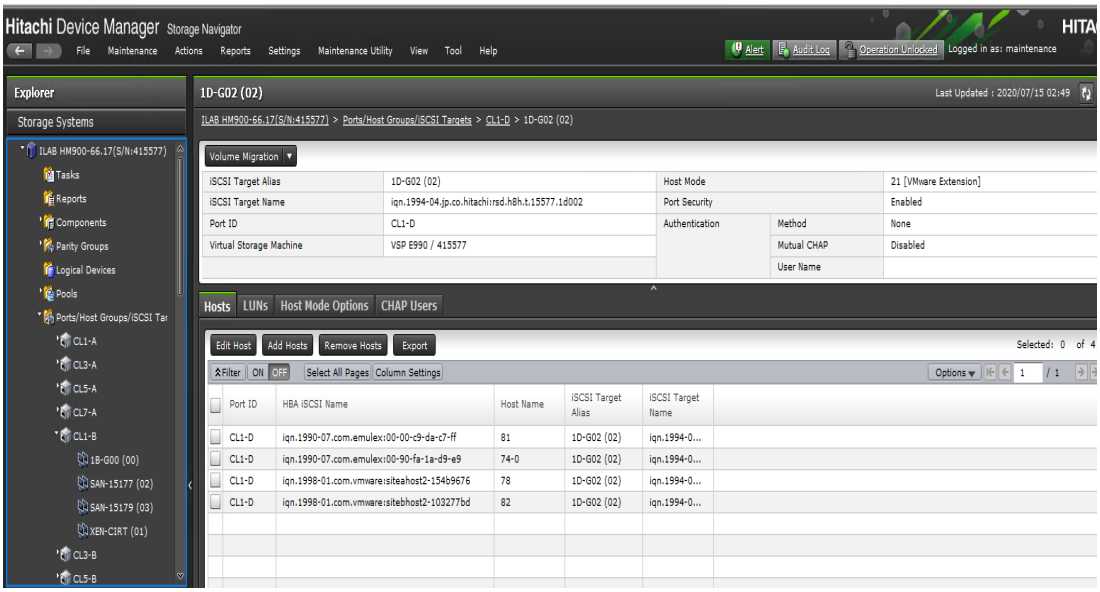

Figure 6: Creating Host Groups

4. Set host mode and host mode options:

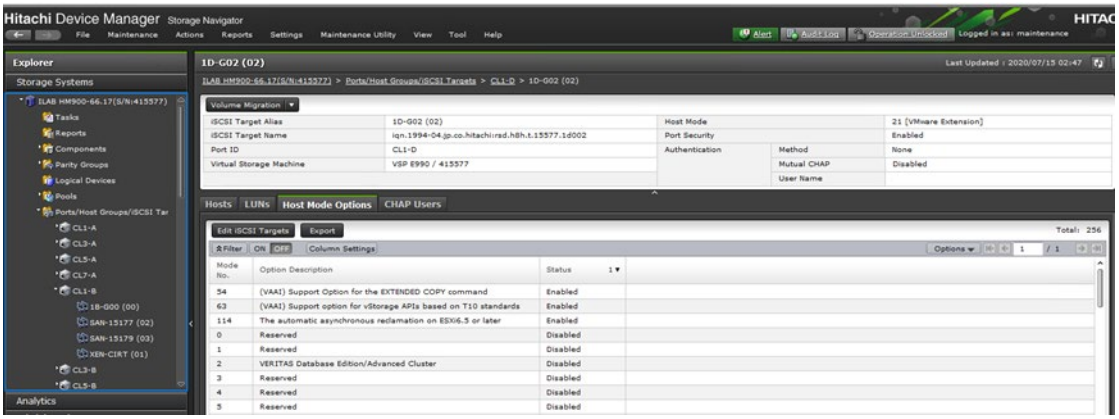

Figure 7: Setting Host Mode Options

5. Assign LDEVs to the host group:

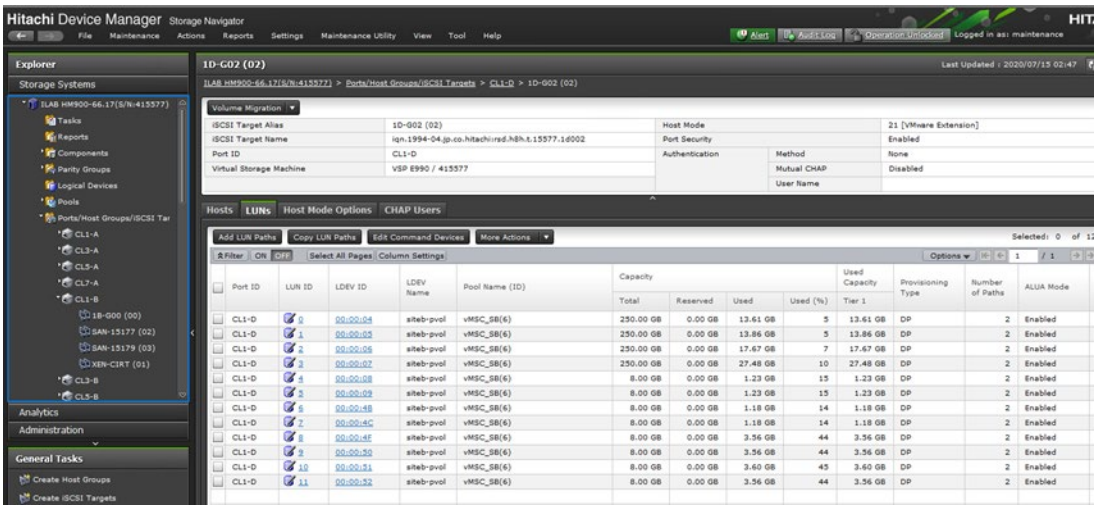

Figure 8: Assigning LDEVs

To create GAD pair from SVP on the secondary storage system, complete the following steps:

1. Create a resource group of the primary storage (site-1) type on the secondary storage system (site-2) and assign the respective secondary resources. Repeat the procedure to create a resource group of the primary storage (site-2) type on the secondary storage system (site-1) and assign the respective secondary resources.

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

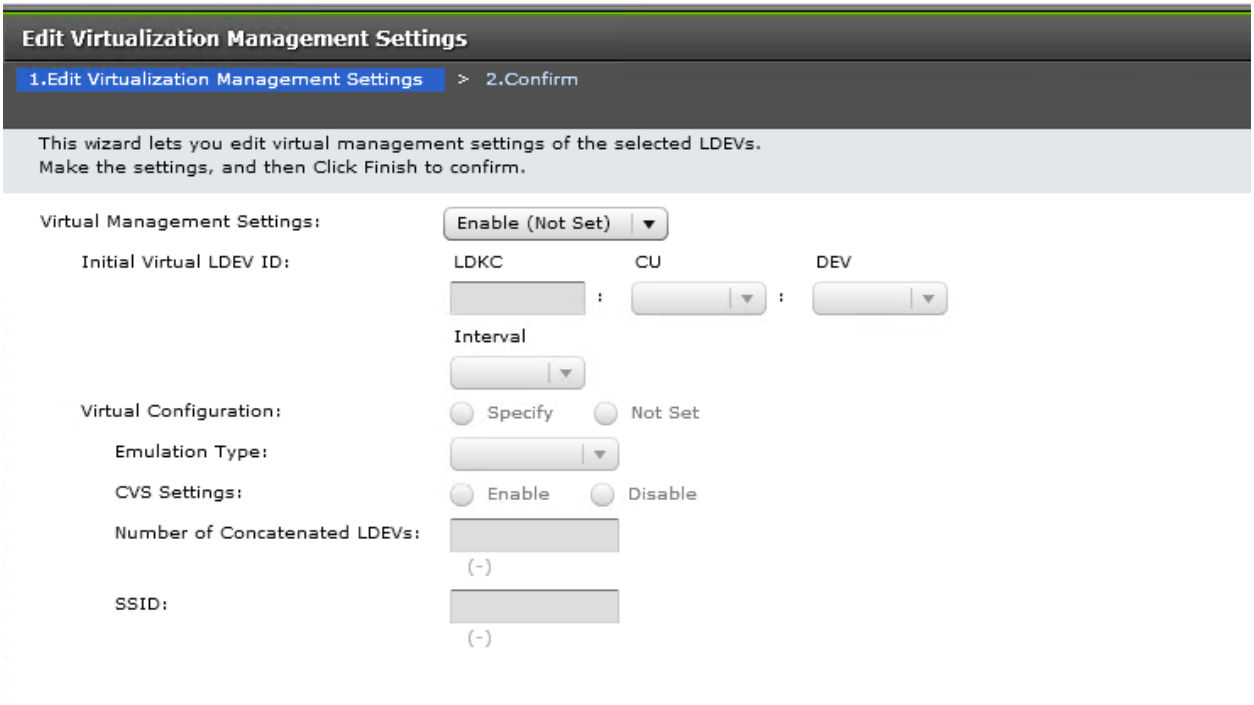

Figure 9: Deleting Virtual LDEV IDs

3. Set the reservation attribute to the volume for the secondary volume of the GAD pair.
4. Set the reservation attribute for GAD to the LDEV IDs.

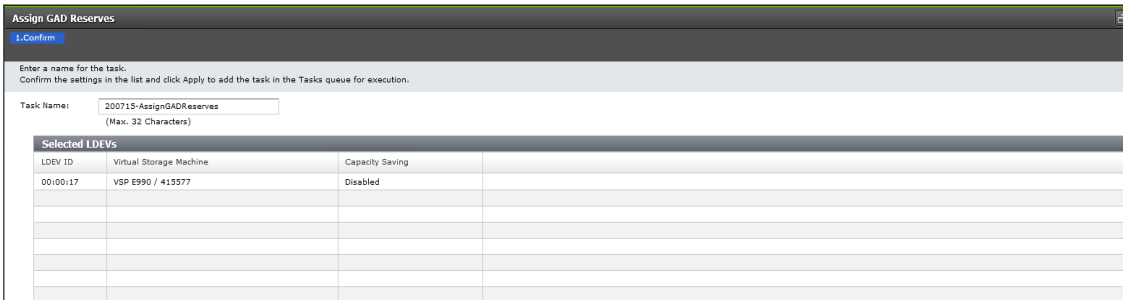

Figure 10: Setting the Reservation Attribute

5. Create a host group of the GAD secondary site storage system and set Host Mode options.

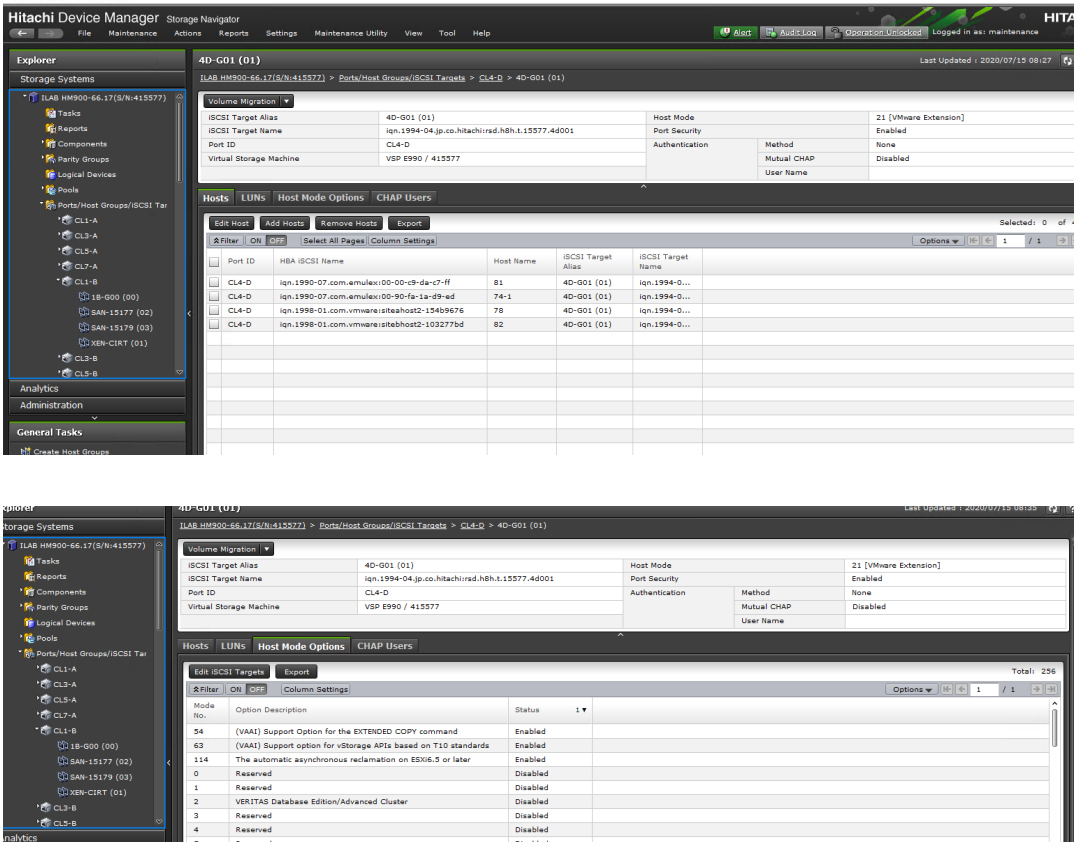

Figure 11: Secondary Site Host Group and Host Mode Options

6. Create DP Pool and LDEVs for secondary volumes of site-1 GAD pair. Repeat the procedure for creating DP Pool and LDEVs for secondary volumes of site-2 GAD pair.

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

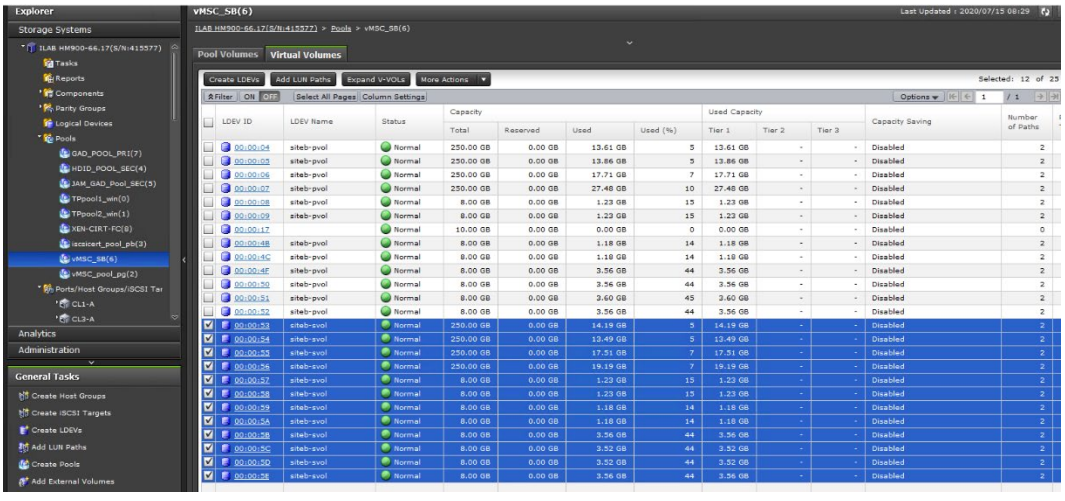

Figure 12: Create DP Pool and LDEVs.

8. Add an LU path to the secondary volume.

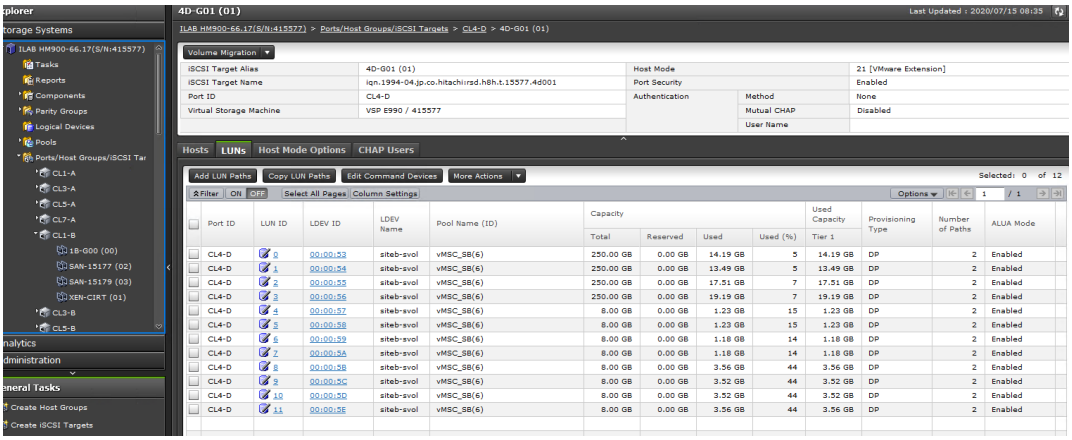

Figure 13: Adding LU Path

9. Create a GAD Pair.

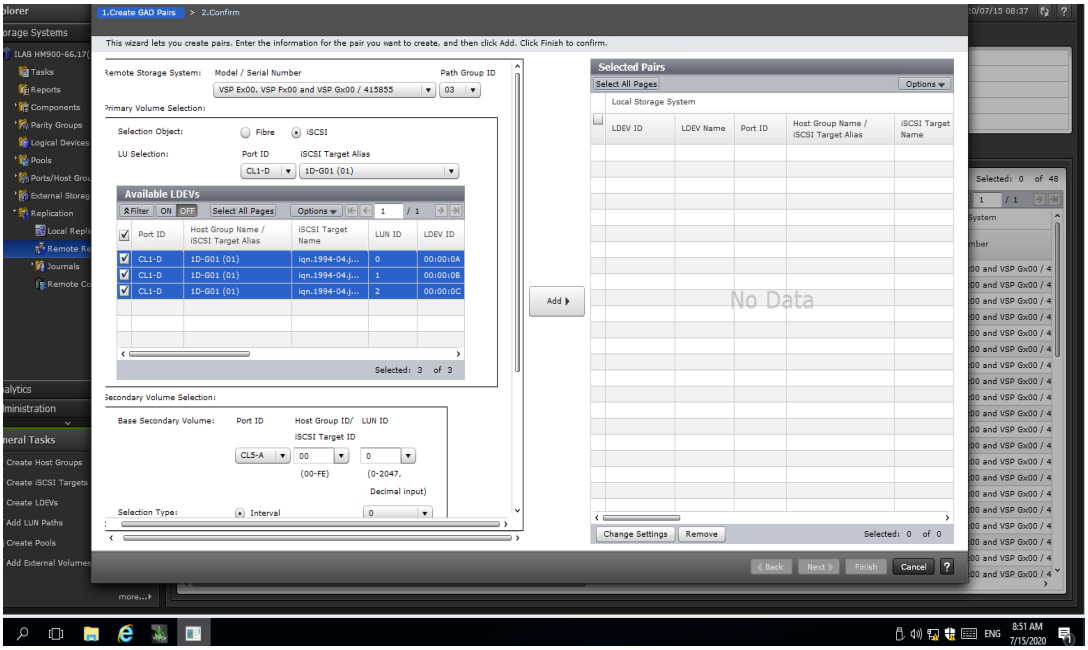

Figure 14: GAD Pair Creation

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

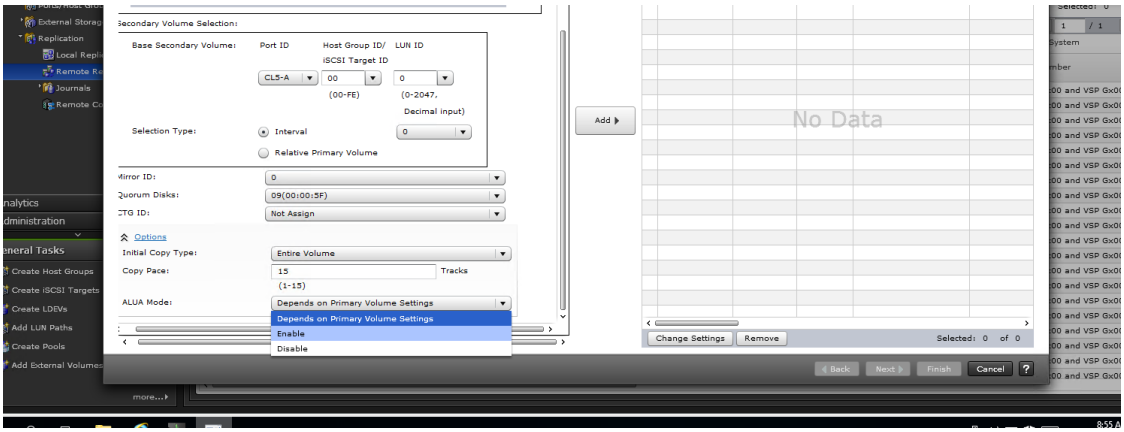

Figure 15: Enabling ALUA Mode

Remote Connection Failure Scenarios

This section covers the typical failure scenarios in a GAD vMSC environment, along with the results for each scenario.

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.
- For NMP/ALUA multipath, set HMO78=OFF.
- For NMP/ALUA, ensure that ALUA is enabled per LUN / Dedicated Ports for PVOLs, and SVOLs enabled with HG Optimized and Non-optimized Paths.
- For vSphere, the UI on all ESXi hosts shows the LUN status: “Active (IO)” à PVOLs and “Active” à SVOLs.
- For SVOLs Storage Ports, zero IOPS was observed, and for PVOLs Storage Ports, generated IO workload was observed.
- For NMP/ALUA, the host sends CMD=A30A to all the paths, and the storage that notifies the Quorum first survives.

- HDLM confirmed that no ALUA RTPG A3h command send occurred; therefore, both PVOLs survived on both storage systems.

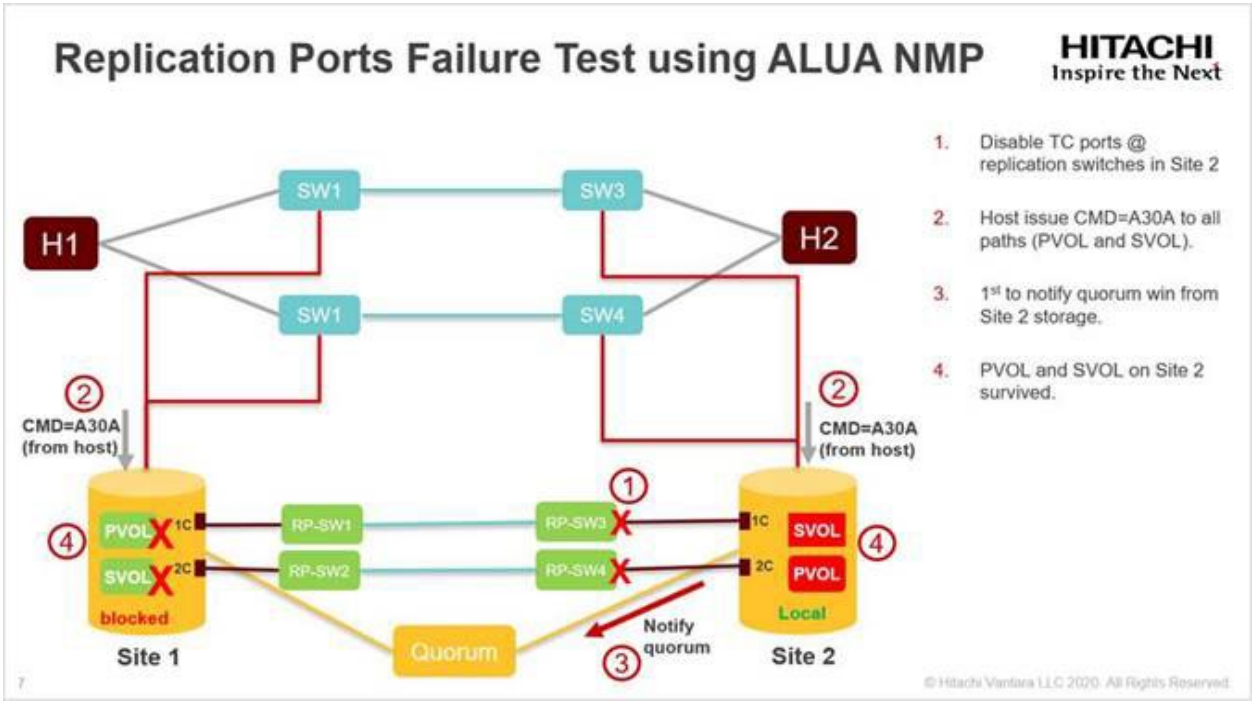

Figure 16: Replication Ports Failure Test

Hitachi Vantara

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

Regional Contact Information
Americas: +1 866 374 5822 or info@hitachivantara.com
Europe, Middle East and Africa: +44 (0) 1753 618000 or info.emea@hitachivantara.com
Asia Pacific: +852 3189 7900 or info.marketing.apac@hitachivantara.com