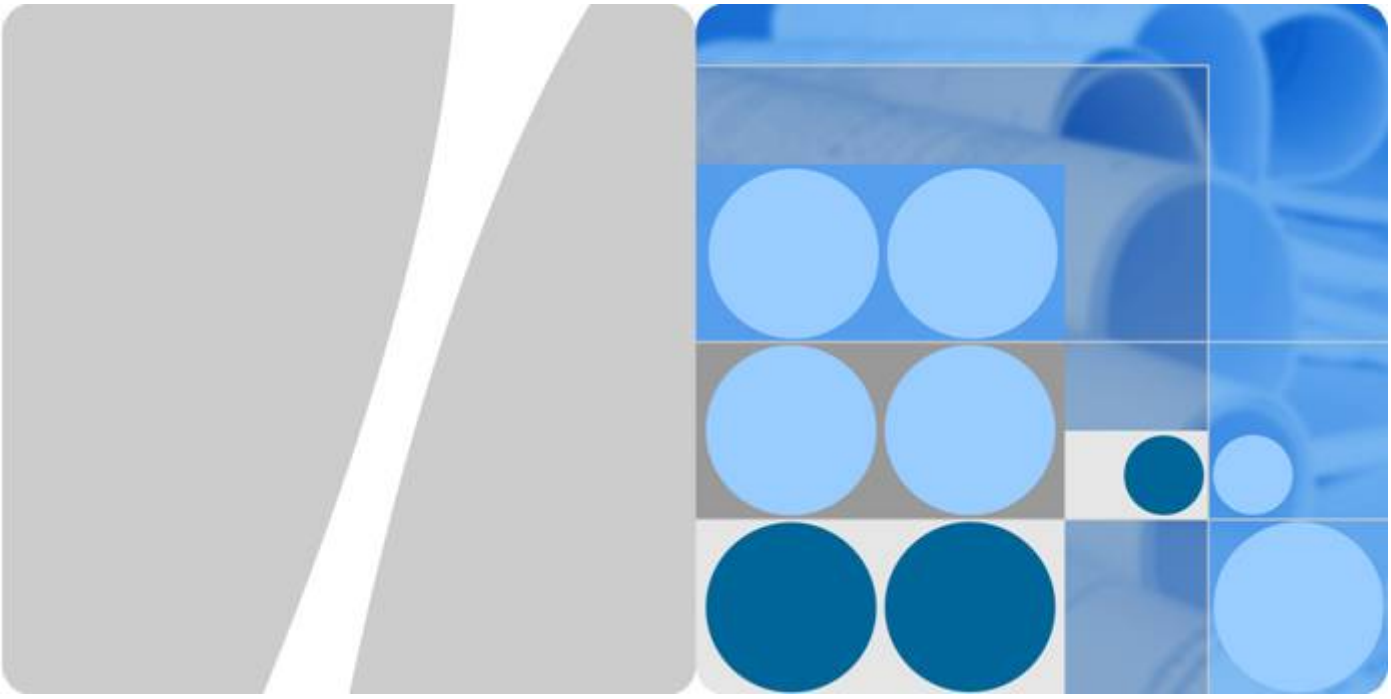


Huawei SAN Storage Host Connectivity Guide for Red Hat

Issue 04

Date 2018-04-10

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://e.huawei.com

Contents

1 About This Document

Overview

This document details the configuration methods and precautions for connecting Huawei SAN storage devices to Red Hat Enterprise Linux (Red Hat for short) hosts.

Intended Audience

This document is intended for:

- Huawei technical support engineers
- Technical engineers of Huawei's partners
- Other personnel who are involved in interconnecting Huawei SAN and Red Hat servers or who are interested in the interconnection.

Readers of this guide are expected to be familiar with the following topics:

- Huawei OceanStor V3, OceanStor V5, and Dorado V3
- Red Hat

Related Documents

For the hosts, host bus adapters (HBAs), and operating systems that are compatible with Huawei storage devices, go to **support-open.huawei.com**.

For the latest Huawei storage product documentation, go to **support.huawei.com**.

For Red Hat documents or support, go to **www.redhat.com/en/services/support**.

Conventions

Symbol Conventions

Symbol	Description
 DANGER	Indicates an imminently hazardous situation which, if not avoided, will result in death or serious injury.

Symbol	Description
	Indicates a potentially hazardous situation which, if not avoided, could result in death or serious injury.
	Indicates a potentially hazardous situation which, if not avoided, may result in minor or moderate injury.
	Indicates a potentially hazardous situation which, if not avoided, could result in equipment damage, data loss, performance deterioration, or unanticipated results. NOTICE is used to address practices not related to personal injury.
	Calls attention to important information, best practices and tips. NOTE is used to address information not related to personal injury, equipment damage, and environment deterioration.

General Conventions

Convention	Description
Times New Roman	Normal paragraphs are in Times New Roman.
Boldface	Names of files, directories, folders, and users are in **boldface**. For example, log in as user **root**.
Italic	Book titles are in *italics*.
Courier New	Examples of information displayed on the screen are in Courier New.

Command Conventions

Format	Description
Boldface	The keywords of a command line are in **boldface**.
Italic	Command arguments are in *italics*.

Where To Get Help

Huawei support and product information can be obtained on the Huawei Online Support site.

Product Information

For documentation, release notes, software updates, and other information about Huawei products and support, go to the Huawei Online Support site (registration required) at **http://support.huawei.com/enterprise/**.

Technical Support

Huawei has a global technical support system, able to offer timely onsite and remote technical support service.

For any assistance, contact:

- Your local technical support

  http://e.huawei.com/en/branch-office-query
- Huawei company headquarters.

  Huawei Technologies Co., Ltd.

  Address: Huawei Industrial Base Bantian, Longgang Shenzhen 518129 People's Republic of China

  Website: **http://enterprise.huawei.com/**

Document Feedback

Huawei welcomes your suggestions for improving our documentation. If you have comments, send your feedback to infoit@huawei.com.

2 Introduction

2.1 Basic Concepts

2.1.1 Introduction to Red Hat

Red Hat Enterprise Linux is an open-source Linux operating system developed by Red Hat. It supports all major hardware platforms and thousands of commercial and custom applications. By combining technology innovation brought by open source code with the stability of enterprise-level platforms, it delivers the highest reliability and best cost-effectiveness.

2.1.2 File Systems in Red Hat

Red Hat offers a variety of file systems from which to choose:

- ext4

  The fourth extended file system (ext4) is a journaling file system for Linux, developed as the successor to ext3. This file system is the fourth edition of ext or extfs for Linux and is supported since Linux kernel 2.6.28. As an optimized version of ext3, ext4 modifies some major data structures in ext3. The maximum size of the ext4 file system is 1 EB and the maximum file size is 16 TB.

- ext3

  The third extended file system (ext3) is a journaling file system developed by the open-source community. This file system supports multiple log types and is highly available. As an extension of ext2, ext3 is compatible with ext2. The maximum size of the ext3 file system is 16 TB and the maximum file size is 2 TB.

- ext2

  The second extended file system (ext2) is a standard file system for Linux. ext2 is an extension of the Minix file system. ext2 has outstanding file access capability,

particularly in processing small and medium-sized files. This file system is gradually replaced by ext3.

- tmpfs

  tmpfs is a memory-based file system similar to a virtual disk. It can use RAM and swap space for storage. Different from virtual disks, tmpfs is available immediately after being installed. tmpfs is the best RAM-based file system.

NOTE

Virtual disks are block devices available only after being formatted by mkfs.

- cramfs

  The compressed ROM file system (cramfs) does not compress all contents in it to the memory at a time. During data access, this file system first locates the requested data and then decompresses the data to the memory in real time. The data is accessed in the memory instead of in the file system.

You can run the following command to view the types of the mounted file systems:

```
[root@localhost ~]# df -Th
Filesystem    Type    Size  Used Avail Use% Mounted on
/dev/sda1     ext4    272G 34G  224G  13% /
tmpfs        tmpfs    7.8G 96K  7.8G 1% /dev/shm
/dev/sdd      ext3     20G  173M 19G 1% /mnt/file_sdd
/dev/sdf      ext3     20G  173M 19G 1% /mnt/file_sdf
[root@localhost ~]#
```

The preceding output shows that the mounted file systems are ext4, ext3, and tmpfs.

2.2 Host-SAN Connectivity

2.2.1 FC Connectivity

A Fibre Channel (FC) SAN is a specialized high-speed network that connects host servers to storage systems. The FC SAN components include HBAs in the host servers, switches that help route storage traffic, cables, storage processors (SPs), and storage disk arrays.

To transfer traffic from host servers to shared storage, the FC SAN uses the Fibre Channel protocol that packages SCSI commands into Fibre Channel frames.

- Ports in FC SAN

  Each node in the SAN, such as a host, a storage device, or a fabric component has one or more ports that connect it to the SAN. Ports are identified in a number of ways, such as by:

  - World Wide Port Name (WWPN)

    A globally unique identifier for a port that allows certain applications to access the port. The FC switches discover the WWPN of a device or host and assign a port address to the device.

  - Port_ID (or port address)

    Within a SAN, each port has a unique port ID that serves as the FC address for the port. This unique ID enables routing of data through the SAN to that port. The FC switches assign the port ID when the device logs in to the fabric. The port ID is valid only when the device is logged on.

- Zoning

  Zoning provides access control in the SAN topology. Zoning defines which HBAs can connect to which targets. When you configure a SAN by using zoning, the devices outside a zone are not visible to the devices inside the zone.

  Zoning has the following effects:

  - Reduces the number of targets and LUNs presented to a host.
  - Controls and isolates paths in a fabric.
  - Separates different environments, for example, a test from a production environment.

2.2.2 iSCSI Connectivity

In computing, Internet Small Computer Systems Interface (iSCSI) is an IP-based storage networking standard for linking data storage systems.

By carrying SCSI commands over IP networks, iSCSI is used to access remote block devices in the SAN, providing hosts with the illusion of locally attached devices.

A single discoverable entity on the iSCSI SAN, such as an initiator or a target, represents an iSCSI node.

Each iSCSI node can be identified in a number of ways, such as by.

- IP address

  Each iSCSI node can have an IP address associated with it so that routing and switching equipment on your network can establish the connection between the server and storage. This address is just like the IP address that you assign to your computer to get access to your company's network or the Internet.

- iSCSI name

  A worldwide unique name for identifying the node. iSCSI uses the iSCSI Qualified Name (IQN) and Extended Unique Identifier (EUI).

  By default, Red Hat generates unique iSCSI names for your iSCSI initiators, for example, iqn.1994-05.com.redhat:876ee1a1014. Usually, you do not have to change the default value, but if you do, make sure that the new iSCSI name you enter is worldwide unique.

2.2.3 Multipath Connectivity

UltraPath

UltraPath is a Huawei-developed multipathing software. It can manage and process disk creation/deletion and I/O delivery of operating systems.

UltraPath provides the following functions:

- Masking of redundant LUNs

  In a redundant storage network, an application server with no multipathing software detects a LUN on each path. Therefore, a LUN mapped through multiple paths is mistaken for two or more different LUNs. UltraPath installed on the application server masks redundant LUNs on the operating system driver layer to provide the application server with only one available LUN, the virtual LUN. In this case, the application server

only needs to deliver data read and write operations to UltraPath that masks the redundant LUNs, and properly writes data into LUNs without damaging other data.

- Optimum path selection

  In a multipath environment, the owning controller of the LUN on the storage system mapped to an application server is the prior controller. With UltraPath, an application server accesses the LUN on the storage system through the prior controller, thereby obtaining the highest I/O speed. The path to the prior controller is the optimum path.

- Failover and failback
  - Failover

    When a path fails, UltraPath fails over its services to another functional path.
  - Failback

    UltraPath automatically delivers I/Os to the first path again after the path recovers from the fault. There are two methods to recover a path:

- I/O Load balancing

  UltraPath provides load balancing within a controller and across controllers.
  - For load balancing within a controller, I/Os poll among all the paths of the controller.
  - For load balancing across controllers, I/Os poll among the paths of all these controllers.

- Path test

  UltraPath tests the following paths:
  - Faulty paths

    UltraPath tests faulty paths with a high frequency to detect the path recover as soon as possible.
  - Idle paths

    UltraPath tests idle paths to identify faulty paths in advance, preventing unnecessary I/O retires. The test frequency is kept low to minimize impact on service I/Os.

DM-Multipath

DM-Multipath is built-in multipathing software in Red Hat.

DM-Multipath allows you to configure multiple I/O paths between a host and a storage system as one device. These I/O paths may contain independent physical devices such as cables, switches, and controllers.

DM-Multipath supports redundant paths and improves system performance.

- Redundancy

  DM-Multipath supports active/standby path configuration. This configuration creates a redundant path for each active path. The redundant paths are not used when the active paths work properly. Once an element (such as a cable, switch, or controller) on an active I/O path becomes faulty, DM-Multipath switches I/Os to a standby path.

- Performance enhancement

  DM-Multipath supports active-active paths, that is, I/Os are distributed to all paths based on the I/O scheduling algorithm. DM-Multipath can check I/O loads on paths and dynamically balance I/Os among the paths using the round-robin algorithm.

Table 2-1 describes DM-Multipath components.

Table 2-1 DM-Multipath components

Component	Description
Kernel module	Redirects I/Os on paths and path groups and provides redundant paths.
mpathconf	A command used to configure and manage DM-Multipath (applicable in some operating systems)
multipath	A management command use to list and configure multipathing devices
multipathd	A daemon process that monitors paths. It initiates path switchover upon a path fault. This process also interactively modifies multipathing devices. This process is started before the **/etc/multipath.conf** file is modified.

ALUA

- ALUA definition:

  Asymmetric Logical Unit Access (ALUA) is a multi-target port access model. In a multipathing state, the ALUA model provides a way of presenting active/passive LUNs to a host and offers a port status switching interface to switch over the working controller. For example, when a host multipathing program that supports ALUA detects a port status change (the port becomes unavailable) on a faulty controller, the program will automatically switch subsequent I/Os to the other controller.

- Support by Huawei storage

  Old-version Huawei storage supports ALUA only in dual-controller configuration, but not in multi-controller or HyperMetro configuration.

  New-version Huawei storage supports ALUA in dual-controller, multi-controller, and HyperMetro configurations.

  Table 2-2 defines old- and new-version Huawei storage.

Table 2-2 Old- and new-version Huawei storage

Storage Type	Version	Remarks
Old-version Huawei storage (namely, storage that does not support multi-controller ALUA or ALUA HyperMetro)	T V1, T V2, 18000 V1, V300R001, V300R002, V300R003C00, V300R003C10, V300R005, and Dorado V300R001C00	-

Storage Type	Version	Remarks
New-version Huawei storage (namely, storage that supports multi-controller ALUA and ALUA HyperMetro)	V300R003C20, V300R006C00, V500R007C00, Dorado V300R001C01, and later versions	V300R003C20: refers to only V300R003C20SPC200 and later versions. V300R006C00: refers to only V300R006C00SPC100 and later versions. Dorado V300R001C01: refers to only V300R001C01SPC100 and later versions.

- ALUA impacts

  ALUA is mainly applicable to a storage system that has only one prior LUN controller. All host I/Os can be routed through different controllers to the working controller for execution. ALUA will instruct the hosts to deliver I/Os preferentially from the LUN working controller, thereby reducing the I/O routing-consumed resources on the non-working controllers.

  If all I/O paths of the LUN working controller are disconnected, the host I/Os will be delivered only from a non-working controller and then routed to the working controller for execution.

- Suggestions for using ALUA on Huawei storage

  To prevent I/Os from being delivered to a non-working controller, you are advised to ensure that:

  - LUN home/working controllers are evenly distributed on storage systems so that host service I/Os are delivered to multiple controllers for load unbalancing.
  - Hosts always try the best to select the optimal path to deliver I/Os even after an I/O path switchover.

2.2.4 SAN Boot

SAN Boot is a network storage management system that stores data (including servers' operating systems) totally on storage systems. Specifically, operating systems are installed on and booted from SAN storage devices. SAN Boot is also called Remote Boot or boot from SAN.

SAN Boot can help to improve system integration, enable centralized management, and facilitate recovery.

- Server integration: Blade servers are used to integrate a large number of servers within a small space. There is no need to configure local disks.
- Centralized management: Boot disks of servers are centrally managed on a storage device. All advanced management functions of the storage device can be fully utilized. For example, the snapshot function can be used for backup. Devices of the same model can be quickly deployed using the snapshot function. In addition, the remote replication function can be used for disaster recovery.

- Quick recovery: Once a server that is booted from SAN fails, its boot volume can be quickly mapped to another server, achieving quick recovery.

2.3 Interoperability Query

When connecting a storage system to a Red Hat host, consider the interoperability of upper-layer applications and components (such as storage systems, Red Hat systems, HBAs, and switches) in the environment.

You can query the latest compatibility information by performing the following steps:

Step 1 Log in to the website **support-open.huawei.com**.

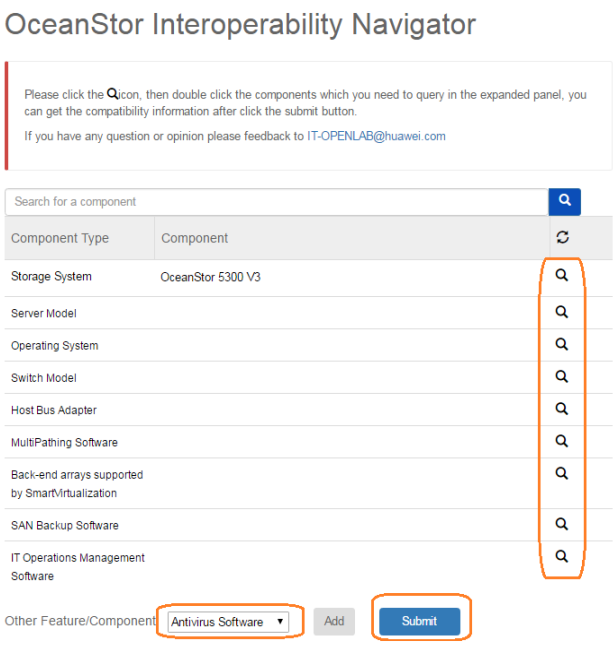
Step 2 On the home page, choose **Interoperability Center > Storage Interoperability**.

Figure 2-1 Interoperability query page

Then, the **OceanStor Interoperability Navigator** is displayed.

Step 3 Select the components to query and click **Submit**.

Figure 2-2 Query on OceanStor Interoperability Navigator



----End

2.4 Specifications

Red Hat has different specifications for LUNs and file systems in various versions. **Table 2-3** lists major Red Hat Linux specifications.

Table 2-3 Major Red Hat specifications

Category	Operating System	Layer	
		HBA (per Initiator)	Kernel
Number of LUNs	RHEL4	256 – Default, 32768 - Max (Emulex) 256 (QLogic) 256 (Brocade)	32768
	RHEL5	256 – Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	65536

	RHEL6	256 – Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	65536
	RHEL7	256 – Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	65536
File system	Type	Max. File System Size	Max. File Size
	ext3	8 TB (RHEL4) 16 TB (RHEL5) 16 TB (RHEL6) 16 TB (RHEL7)	2 TB (RHEL4) 2 TB (RHEL5) 2 TB (RHEL6) 2 TB (RHEL7)
	ext4	16 TB (RHEL5) 16 TB (RHEL6) 50 TB (RHEL7)	16 TB (RHEL5) 16 TB (RHEL6) 16 TB (RHEL7)
	XFS	100 TB (RHEL5) 300 TB (RHEL6) 500 TB (RHEL7)	100 TB (RHEL5) 100 TB (RHEL6) 500 TB (RHEL7)

For more information, go to:

http://cn.redhat.com/resourcelibrary/articles/articles-red-hat-enterprise-linux-6-technology-capabilities-and-limits

Alternatively, go to **https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/** and search Storage Administration Guide of the corresponding version.

2.5 Common Red Hat Commands

Table 2-4 lists the common management commands used in Red Hat hosts.

Table 2-4 Red Hat commands

Command	Function
df	Views the file system size and usage.
fdisk /dev/sd#	Partitions **sd#** disks.

Command	Function
cat /sys/class/scsi_host/host*/modelname	Views the Fibre Channel HBA model.
cat /sys/class/scsi_host/host*/fwrev	Views the Fibre Channel HBA firmware.
ifconfig	Configures network port parameters.
lsscsi	Displays the hardware address, type, and manufacturer of each disk.
lvdisplay -v /dev/vgname/lvname	Views details about **lvname**.
mount	Mounts a logical volume.
shutdown -h now	Shuts down the host.
shutdown –ry 0	Restarts the host.
vgdisplay -v vgname	Views details about **vgname**.
vgscan	Scans for volume groups in the system.

NOTE

The pound (**#**) in the table indicates a number that can be specified based on actual conditions.

3 Planning Connectivity

Red Hat hosts and storage systems can be connected based on different criteria. **Table 3-1** describes the typical connection modes.

Table 3-1 Connection modes

Criteria	Connection Mode
Interface module type	Fibre Channel connection/iSCSI connection
Whether switches are used	Direct connection (no switches are used)/Switch-based connection (switches are used)
Whether multiple paths exist	Single-path connection/Multi-path connection
Whether HyperMetro is configured	HyperMetro/Non-HyperMetro

Fibre Channel connections are the most widely used. To ensure service data security, both direct connections and switch-based connections require multiple paths.

The following details Fibre Channel and iSCSI connections in HyperMetro and non-HyperMetro scenarios.

3.1 HyperMetro Scenarios

For details about how to plan connectivity in HyperMetro scenarios, see the *BC&DR Solution Product Documentation (Active-Active Data Center)*.

3.2 Non-HyperMetro Scenarios

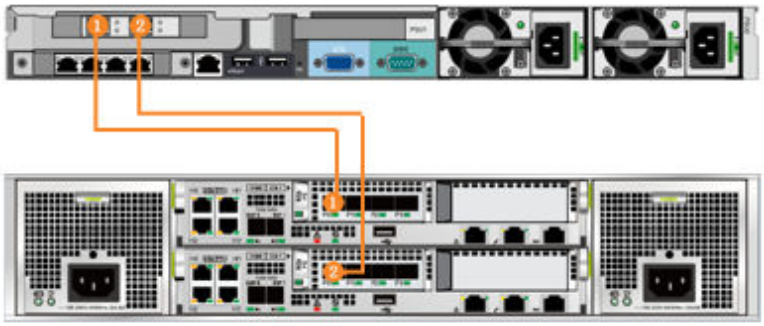
3.2.1 Direct FC Connections

Huawei provides dual-controller and multi-controller storage systems, which directly connect to Red Hat hosts through FC multi-path connections in different ways.

Two-Controller Storage

The following uses Huawei OceanStor 5500 V3 as an example to explain how to directly connect a Red Hat host to a two-controller storage system through FC multi-path connections, as shown in **Figure 3-1**.

Figure 3-1 Direct FC multi-path connections (two-controller storage)



NOTE

In this connection diagram, each of the two controllers is connected to a host HBA port with an optical fiber. The cable connections are detailed in **Table 3-2**.

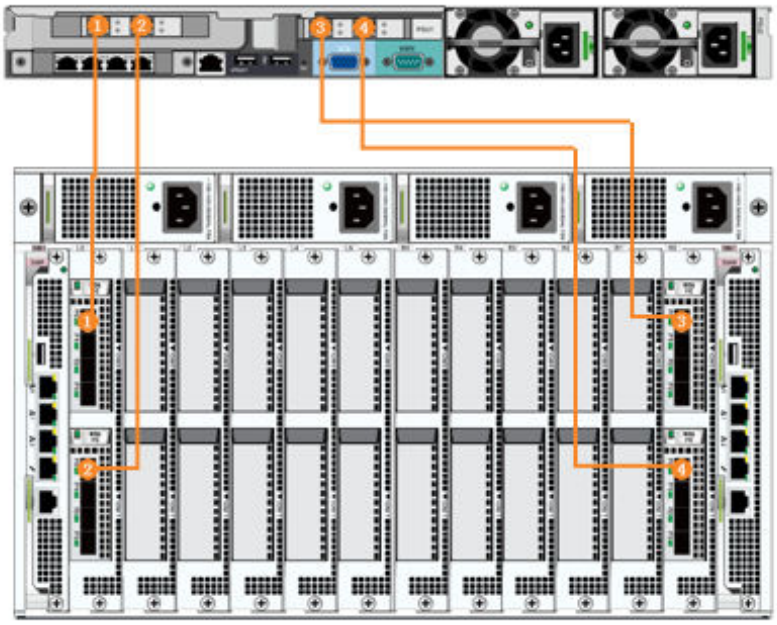
Table 3-2 Cable connection description (two-controller storage)

Cable No.	Description
1	Connects Port P0 on the Red Hat host to Controller A on the storage system.
2	Connects Port P1 on the Red Hat host to Controller B on the storage system.

Multi-Controller Storage

The following uses Huawei OceanStor 18800 V3 (four-controller) as an example to explain how to directly connect a Red Hat host to a multi-controller storage system through FC multi-path connections, as shown in **Figure 3-2**.

Figure 3-2 Direct FC multi-path connections (four-controller storage)



NOTE

In this connection diagram, each of the four controllers is connected to a host HBA port with an optical fiber. The cable connections are detailed in **Table 3-3**.

Table 3-3 Cable connection description (four-controller storage)

Cable No.	Description
1	Connects Port P0 on the Red Hat host to Controller A on the storage system.
2	Connects Port P1 on the Red Hat host to Controller B on the storage system.
3	Connects Port P2 on the Red Hat host to Controller C on the storage system.
4	Connects Port P3 on the Red Hat host to Controller D on the storage system.

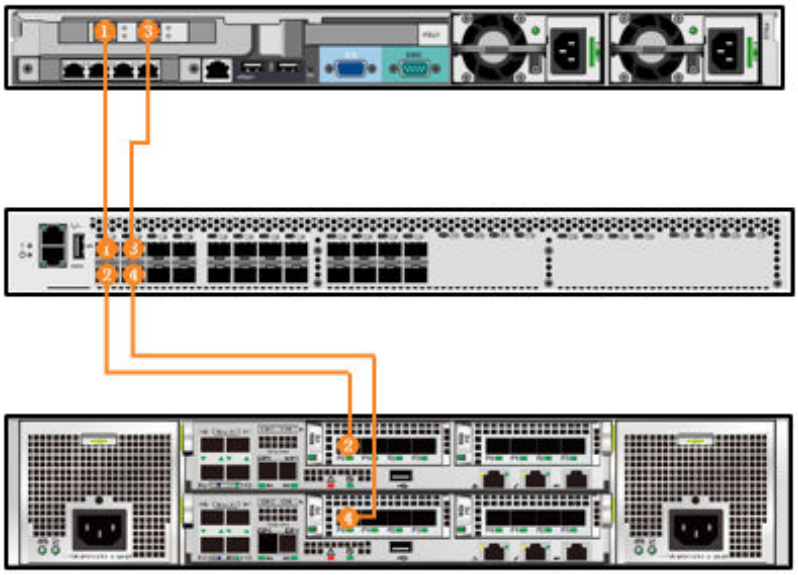
3.2.2 Switch-Based FC Connections

Huawei provides dual-controller and multi-controller storage systems, which connect to Red Hat hosts through FC multi-path connections using a switch in different ways.

Two-Controller Storage

The following uses Huawei OceanStor 5500 V3 as an example to explain how to connect a Red Hat host to a two-controller storage system through FC multi-path connections using a switch, as shown in **Figure 3-3**.

Figure 3-3 Switch-based FC multi-path connections (dual-controller storage)



NOTE

In this connection diagram, two controllers of the storage system and two ports of the Red Hat host are connected to the FC switch through optical fibers. On the FC switch, the ports connecting to the storage controllers and to the Red Hat host are grouped in a zone, ensuring connectivity between the host ports and the storage.

Table 3-4 Zone division on the FC switch (two-controller storage)

Zone Name	Zone Members	Zone Description
Zone001	Ports 1 and 2	Connects Port P0 on the Red Hat host to Controller A on the storage system.
Zone002	Ports 3 and 4	Connects Port P1 on the Red Hat host to Controller B on the storage system.
Zone003	Ports 1 and 4	Connects Port P0 on the Red Hat host to Controller B on the storage system.
Zone004	Ports 3 and 2	Connects Port P1 on the Red Hat host to Controller A on the storage system.

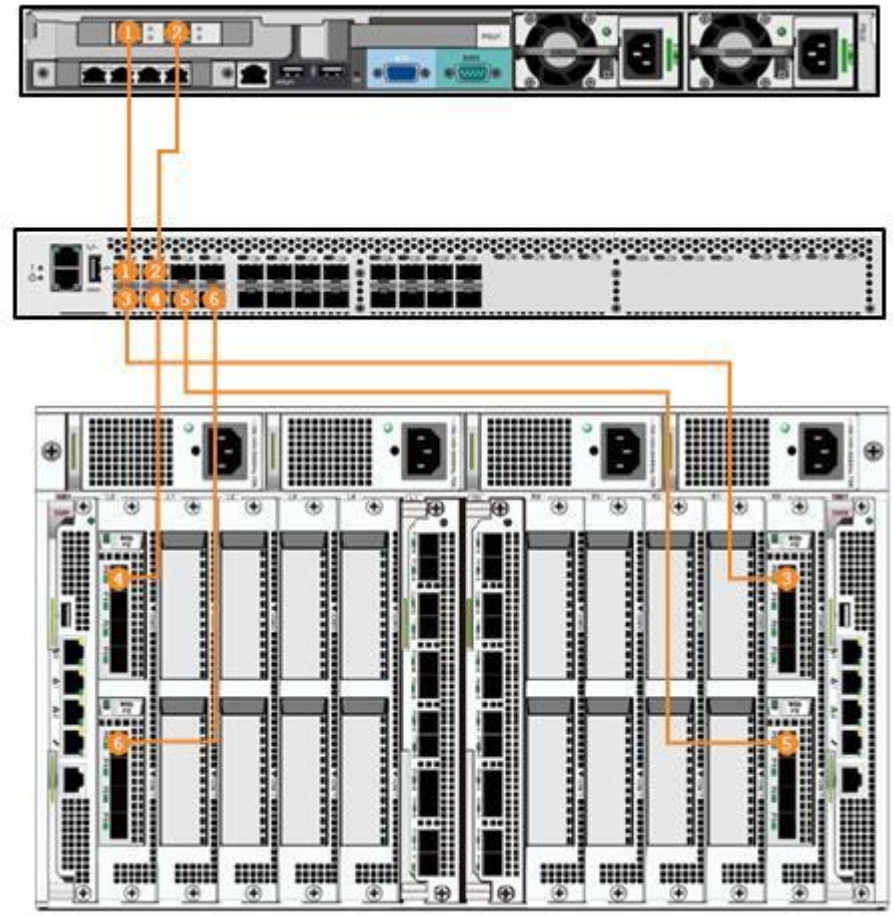
NOTE

Zone division in this table is for reference only. Plan zones based on site requirements.

Multi-Controller Storage

The following uses Huawei OceanStor 18800 V3 (four-controller) as an example to explain how to connect a Red Hat host to a multi-controller storage system through FC multi-path connections using a switch, as shown in **Figure 3-4**.

Figure 3-4 Switch-based FC multi-path connections (four-controller storage)



NOTE

In this connection diagram, four controllers of the storage system and two ports of the Red Hat host are connected to the FC switch through optical fibers. On the FC switch, the ports connecting to the storage controllers and to the Red Hat host are grouped in a zone, ensuring connectivity between the host ports and the storage.

Table 3-5 Zone division on the FC switch (four-controller storage)

Zone Name	Zone Members	Zone Description
Zone001	Ports 1 and 3	Connects Port P0 on the Red Hat host to Controller A on the storage system.
Zone002	Ports 1 and 4	Connects Port P0 on the Red Hat host to Controller B on the storage system.
Zone003	Ports 1 and 5	Connects Port P0 on the Red Hat host to Controller C on the storage system.
Zone004	Ports 1 and 6	Connects Port P0 on the Red Hat host to Controller D on the storage system.

Zone Name	Zone Members	Zone Description
Zone005	Ports 2 and 3	Connects Port P1 on the Red Hat host to Controller A on the storage system.
Zone006	Ports 2 and 4	Connects Port P1 on the Red Hat host to Controller B on the storage system.
Zone007	Ports 2 and 5	Connects Port P1 on the Red Hat host to Controller C on the storage system.
Zone008	Ports 2 and 6	Connects Port P1 on the Red Hat host to Controller D on the storage system.

NOTE

Zone division in this table is for reference only. Plan zones based on site requirements.

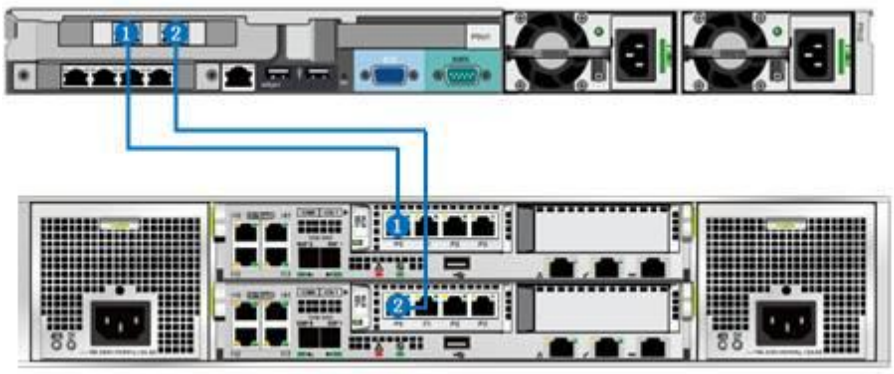
3.2.3 Direct iSCSI Connections

Huawei provides dual-controller and multi-controller storage systems, which directly connect to Red Hat hosts through iSCSI multi-path connections in different ways.

Two-Controller Storage

The following uses Huawei OceanStor 5500 V3 as an example to explain how to directly connect a Red Hat host to a two-controller storage system through iSCSI multi-path connections, as shown in **Figure 3-5**.

Figure 3-5 Direct iSCSI multi-path connections (two-controller storage)



NOTE

In this connection diagram, each of the two controllers is connected to a port on the host network adapter with a network cable. The IP address plan is detailed in **Table 3-6**.

Table 3-6 IP address plan for direct iSCSI multi-path connections (two-controller storage)

Port Name	Port Description	IP Address	Subnet Mask
Host.P0	Connects the Red Hat host to Controller A on the storage system.	192.168.5.5	255.255.255.0
Host.P1	Connects the Red Hat host to Controller B on the storage system.	192.168.6.5	255.255.255.0
Storage.A.P0	Connects Controller A on the storage system to the Red Hat host.	192.168.5.6	255.255.255.0
Storage.B.P0	Connects Controller B on the storage system to the Red Hat host.	192.168.6.6	255.255.255.0

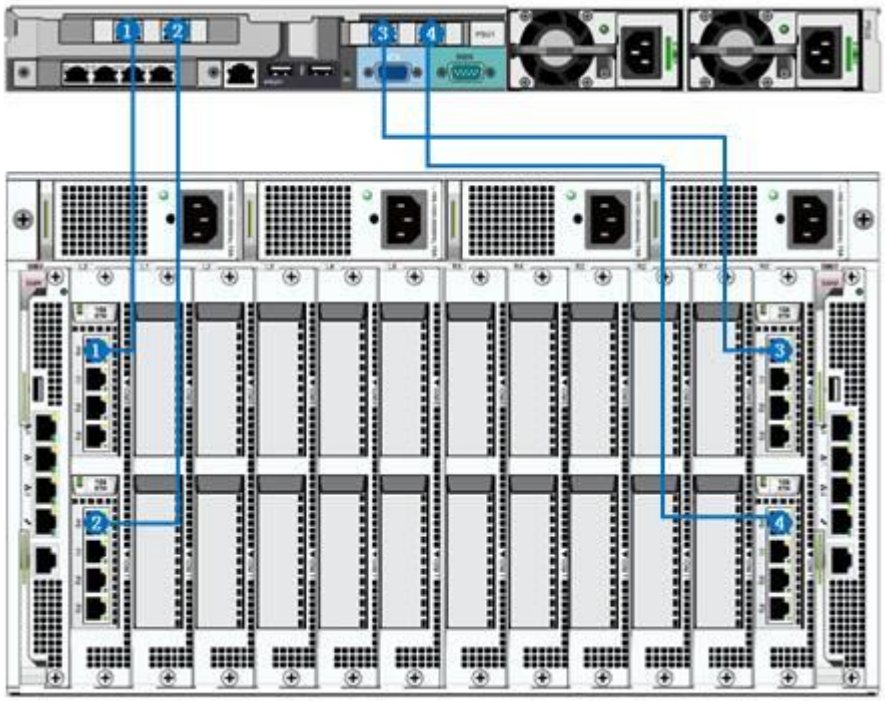
NOTE

IP addresses in this table are for reference only. Plan IP addresses based on site requirements.

Multi-Controller Storage

The following uses Huawei OceanStor 18800 V3 (four-controller) as an example to explain how to directly connect a Red Hat host to a multi-controller storage system through iSCSI multi-path connections, as shown in **Figure 3-6**.

Figure 3-6 Direct iSCSI multi-path connections (four-controller storage)



NOTE

In this connection diagram, each of the four controllers is connected to a port on host network adapters with a network cable. The IP address plan is detailed in **Table 3-7**.

Table 3-7 IP address plan for direct iSCSI multi-path connections (four-controller storage)

Port Name	Port Description	IP Address	Subnet Mask
Host.P0	Connects the Red Hat host to Controller A on the storage system.	192.168.5.5	255.255.255.0
Host.P1	Connects the Red Hat host to Controller B on the storage system.	192.168.6.5	255.255.255.0
Host.P2	Connects the Red Hat host to Controller C on the storage system.	192.168.7.5	255.255.255.0
Host.P3	Connects the Red Hat host to Controller D on the storage system.	192.168.8.5	255.255.255.0
Storage.A.P0	Connects Controller A on the storage system to the Red Hat host.	192.168.5.6	255.255.255.0
Storage.B.P0	Connects Controller B on the storage system to the Red Hat host.	192.168.6.6	255.255.255.0
Storage.C.P0	Connects Controller C on the storage system to the Red Hat host.	192.168.7.6	255.255.255.0
Storage.D.P0	Connects Controller D on the storage system to the Red Hat host.	192.168.8.6	255.255.255.0

NOTE

IP addresses in this table are for reference only. Plan IP addresses based on site requirements.

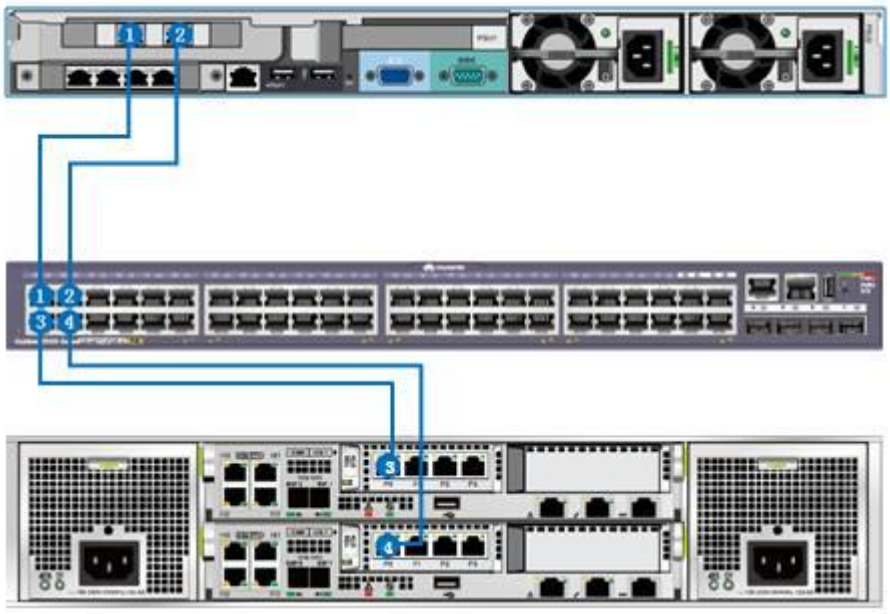
3.2.4 Switch-Based iSCSI Connections

Huawei provides dual-controller and multi-controller storage systems, which connect to Red Hat hosts through Ethernet switches in different ways.

Two-Controller Storage

The following uses Huawei OceanStor 5500 V3 as an example to explain how to connect a Red Hat host to a two-controller storage system through iSCSI multi-path connections using an Ethernet switch, as shown in **Figure 3-7**.

Figure 3-7 Switch-based iSCSI multi-path connections (two-controller storage)



NOTE

In this connection diagram, two controllers of the storage system and two ports of the Red Hat host network adapter are connected to the Ethernet switch through network cables. IP addresses of the ports on the storage and host are in the same subnet, ensuring connectivity between the host ports and the storage.

Table 3-8 IP address plan for switch-based iSCSI multi-path connections (two-controller storage)

Port Name	Port Description	IP Address	Subnet Mask
Host.P0	Connects the Red Hat host to Controller A on the storage system.	192.168.5.5	255.255.255.0
Host.P1	Connects the Red Hat host to Controller B on the storage system.	192.168.6.5	255.255.255.0
Storage.A.P0	Connects Controller A on the storage system to the Red Hat host.	192.168.5.6	255.255.255.0
Storage.B.P0	Connects Controller B on the storage system to the Red Hat host.	192.168.6.6	255.255.255.0

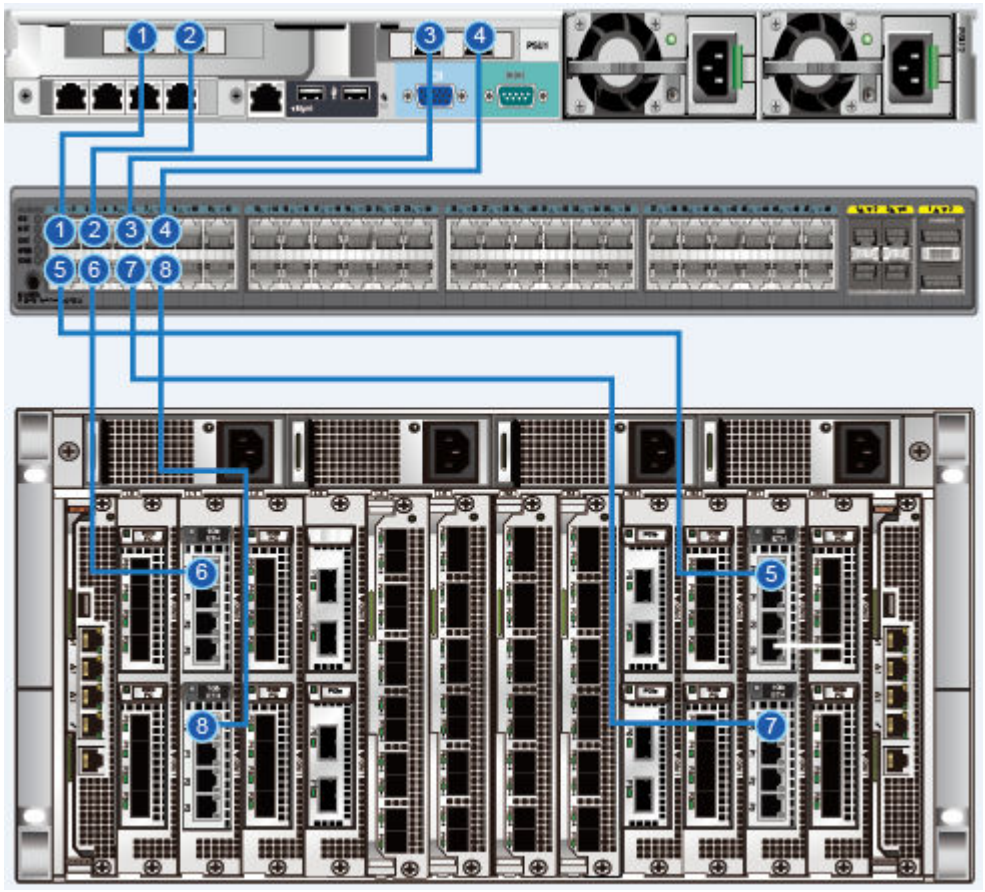
NOTE

IP addresses in this table are for reference only. Plan IP addresses based on site requirements.

Multi-Controller Storage

The following uses Huawei OceanStor 18800 V3 (four-controller) as an example to explain how to connect a Red Hat host to a multi-controller storage system through iSCSI multi-path connections using an Ethernet switch, as shown in **Figure 3-8**.

Figure 3-8 Switch-based iSCSI multi-path connections (four-controller storage)



NOTE

In this connection diagram, four controllers of the storage system and four ports of the Red Hat host network adapters are connected to the Ethernet switch through network cables. IP addresses of the ports on the storage and host are in the same subnet, ensuring connectivity between the host ports and the storage.

Table 3-9 IP address plan for switch-based iSCSI multi-path connections (four-controller storage)

Port Name	Port Description	IP Address	Subnet Mask
Host.P0	Connects the Red Hat host to Controller A on the storage system.	192.168.5.5	255.255.255.0
Host.P1	Connects the Red Hat host to Controller B on the storage system.	192.168.6.5	255.255.255.0
Host.P2	Connects the Red Hat host to Controller C on the storage system.	192.168.7.5	255.255.255.0
Host.P3	Connects the Red Hat host to Controller D on the storage system.	192.168.8.5	255.255.255.0
Storage.A.P0	Connects Controller A on the storage system to the Red Hat host.	192.168.5.6	255.255.255.0
Storage.B.P0	Connects Controller B on the storage system to the Red Hat host.	192.168.6.6	255.255.255.0
Storage.C.P0	Connects Controller C on the storage system to the Red Hat host.	192.168.7.6	255.255.255.0
Storage.D.P0	Connects Controller D on the storage system to the Red Hat host.	192.168.8.6	255.255.255.0

NOTE

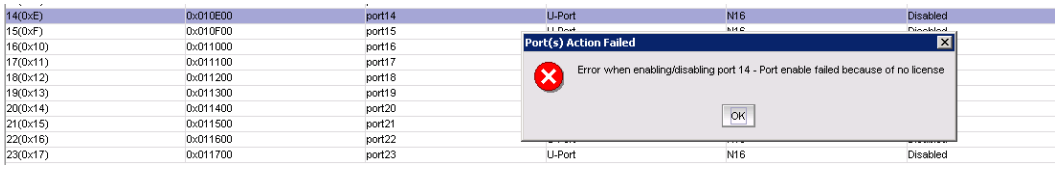
IP addresses in this table are for reference only. Plan IP addresses based on site requirements.

4 Preparations Before Configuration

4.1 Switch

Ensure that the switches are running properly and their ports have the necessary licenses and transmit data normally. **Figure 4-1** shows an example of a port failure due to lack of a license.

Figure 4-1 Switch port status



It is recommended that you obtain the product documentation of the switches for reference.

4.2 Storage System

Create disk domains, storage pools, LUNs, hosts, and mapping views on the storage system according to your service requirements. For details about these operations, see the *Basic Storage Service Configuration Guide* corresponding to your storage system.

4.3 Host

Before connecting a host to a storage system, make sure that the host HBAs are identified and working properly. You also need to obtain the WWNs of HBA ports. The WWNs will be used in subsequent configuration on the storage system.

4.3.1 Identifying HBAs

After an HBA is installed on a host, run the following command on the host to check whether the HBA is identified by the host.

```
[root@localhost ~]# lspci|grep Fibre
03:00.0 Fibre Channel: Emulex Corporation Saturn-X: LightPulse Fibre Channel Host
Adapter (rev 03)
03:00.1 Fibre Channel: Emulex Corporation Saturn-X: LightPulse Fibre Channel Host
Adapter (rev 03)
[root@localhost ~]#
[root@localhost ~]# cat /sys/class/scsi_host/host*/model*name
LPe12002
LPe12002
```

The output indicates that the host has identified two Fibre Channel host ports and the HBA model is Emulex LPe12002.

The method for viewing the HBA WWN varies according to the operating system. The following describes the methods on different operating systems.

Red Hat 4

Run the following command on the host:

```
# grep scsi /proc/scsi/qla2xxx/3
Number of reqs in pending_q= 0, retry_q= 0, done_q= 0, scsi_retry_q= 0
scsi-qla0-adapter-node=20000018822d7834;
scsi-qla0-adapter-port=21000018822d7834;
scsi-qla0-target-0=202900a0b8423858;
scsi-qla0-port-0=200800a0b8423858:202900a0b8423858:0000e8:1;
```

Red Hat 5 and Later Versions

Run the following command on the host:

```
# cat /sys/class/fc_host/hostx/port_name
```

4.3.2 Querying HBA Properties

This section describes how to query the attributes of QLogic and Emulex HBAs, including their models, driver versions, firmware versions, WWNs, port topologies, and port rates. To query other attributes or other vendors' HBAs, it is recommended that you use the management software provided by the respective HBA vendor. See the operation guide of the corresponding HBA management software for detailed operations.

QLogic

Figure 4-2 shows the command for querying the QLogic HBA model.

Figure 4-2 HBA model

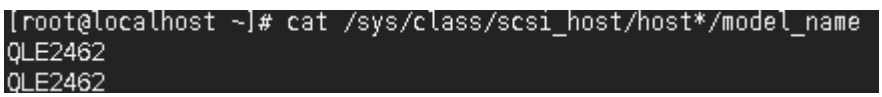


Figure 4-3 shows the command for querying the QLogic HBA driver version.

Figure 4-3 HBA driver version

```
[root@localhost ~]# cat /sys/class/scsi_host/host*/driver_version
8.07.00.33.07.3-k1
8.07.00.33.07.3-k1
```

Figure 4-4 shows the command for querying the QLogic HBA firmware version.

Figure 4-4 HBA firmware version

```
[root@localhost ~]# cat /sys/class/scsi_host/host*/optrom_fw_version
4.06.02 1154
4.06.02 1154
```

Figure 4-5 shows the command for querying the QLogic HBA WWN.

Figure 4-5 HBA WWN

```
[root@localhost ~]# cat /sys/class/fc_host/host*/port_name
0x2100001b328af20e
0x2101001b32aaf20e
```

Figure 4-6 shows the command for querying the QLogic HBA topology.

Figure 4-6 HBA topology

```
[root@localhost ~]# cat /sys/class/fc_host/host*/port_type
LPort (private loop)
LPort (private loop)
```

Figure 4-7 shows the command for querying the QLogic HBA port rate.

Figure 4-7 HBA port rate

```
[root@localhost ~]# cat /sys/class/fc_host/host*/speed
4 Gbit
4 Gbit
```

Emulex

Figure 4-8 shows the command for querying the Emulex HBA model.

Figure 4-8 HBA model

```
[root@localhost ~]# cat /sys/class/scsi_host/host*/modelname
LPe16002B-M6
LPe16002B-M6
```

Figure 4-9 shows the command for querying the Emulex HBA driver version.

Figure 4-9 HBA driver version

```
[root@localhost ~]# cat /sys/class/scsi_host/host*/lpfc_drvr_version
Emulex LightPulse Fibre Channel SCSI driver 8.3.7.21.4p
Emulex LightPulse Fibre Channel SCSI driver 8.3.7.21.4p
```

Figure 4-10 shows the command for querying the Emulex HBA firmware version.

Figure 4-10 HBA firmware version

```
[root@localhost ~]# cat /sys/class/scsi_host/host*/fwrev
11.2.156.27, sli-4:2:b
11.2.156.27, sli-4:2:b
```

Figure 4-11 shows the command for querying the Emulex HBA WWN.

Figure 4-11 HBA WWN

```
[root@localhost ~]# cat /sys/class/fc_host/host*/port_name
0x100000109b1c82f8
0x100000109b1c82f9
```

Figure 4-12 shows the command for querying the Emulex HBA topology.

Figure 4-12 HBA topology

```
[root@localhost ~]# cat /sys/class/fc_host/host*/port_type
NPort (fabric via point-to-point)
NPort (fabric via point-to-point)
```

Figure 4-13 shows the command for querying the Emulex HBA port rate.

Figure 4-13 HBA port rate

```
[root@localhost ~]# cat /sys/class/fc_host/host*/speed
16 Gbit
16 Gbit
```

5 Configuring Connectivity

5.1 Establishing Fibre Channel Connections

5.1.1 Host Configuration

Query the HBA WWN. **Figure 5-1** provides an example.

Figure 5-1 Querying the WWN of the host HBA

```
[root@localhost ~]# cat /sys/class/fc_host/host*/port_name
0x100000109b1c82f8
0x100000109b1c82f9
```

5.1.2 (Optional) Switch Configuration

The commonly used Fibre Channel switches are mainly from Brocade, Cisco, and QLogic. This section uses Brocade switches as an example to explain how to configure them.

5.1.2.1 Querying the Switch Model and Version

Perform the following steps to query the switch model and version:

Step 1 Log in to the Brocade switch on a web browser.

1. On the web browser, enter the IP address of the Brocade switch and press **Enter**.
2. In the **Web Tools** switch login dialog box is displayed, enter the account and password.

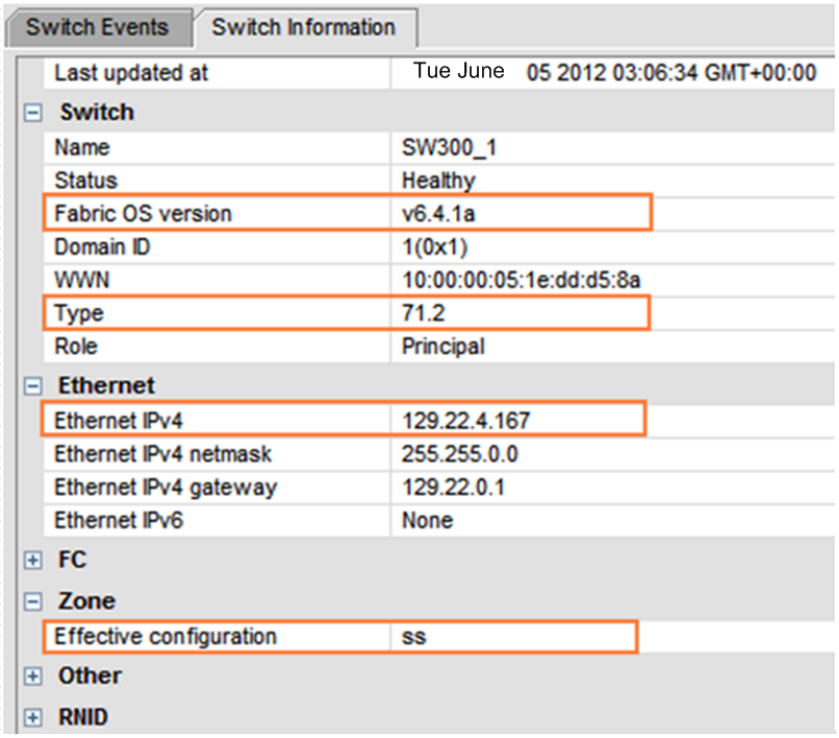
   The default account and password are **admin** and **password**.

CAUTION

Web Tools works properly only when Java is installed on the host. Java 1.6 or later is recommended.

Step 2 On the switch management page that is displayed, click **Switch Information**.

Figure 5-2 Switch information



Major parameters are describes as follows:

- **Fabric OS version**: indicates the switch version. The interoperability between switches and storage systems varies with the switch version. Only switches of authenticated versions can interconnect correctly with storage systems.
- **Type**: This parameter is a decimal consisting of an integer and a decimal fraction. The integer indicates the switch model and the decimal fraction indicates the switch template version. You only need to pay attention to the switch model. **Table 5-1** describes switch model mapping.
- **Ethernet IPv4**: indicates the switch IP address.
- **Effective Configuration**: indicates the currently effective configurations. This parameter is critical to subsequent zone configurations. In this example, the currently effective configuration is **ss**.

Table 5-1 Mapping between switch types and names

Switch Type	Switch Name	Switch Type	Switch Name
1	Brocade 1000 Switch	64	Brocade 5300 Switch
2,6	Brocade 2800 Switch	66	Brocade 5100 Switch
3	Brocade 2100, 2400 Switches	67	Brocade Encryption Switch
4	Brocade 20x0, 2010, 2040, 2050 Switches	69	Brocade 5410 Blade
5	Brocade 22x0, 2210, 2240, 2250 Switches	70	Brocade 5410 Embedded Switch
7	Brocade 2000 Switch	71	Brocade 300 Switch
9	Brocade 3800 Switch	72	Brocade 5480 Embedded Switch
10	Brocade 12000 Director	73	Brocade 5470 Embedded Switch
12	Brocade 3900 Switch	75	Brocade M5424 Embedded Switch
16	Brocade 3200 Switch	76	Brocade 8000 Switch
17	Brocade 3800VL	77	Brocade DCX-4S Backbone
18	Brocade 3000 Switch	83	Brocade 7800 Extension Switch
21	Brocade 24000 Director	86	Brocade 5450 Embedded Switch
22	Brocade 3016 Switch	87	Brocade 5460 Embedded Switch
26	Brocade 3850 Switch	90	Brocade 8470 Embedded Switch
27	Brocade 3250 Switch	92	Brocade VA-40FC Switch
29	Brocade 4012 Embedded Switch	95	Brocade VDX 6720-24 Data Center Switch
32	Brocade 4100 Switch	96	Brocade VDX 6730-32 Data Center Switch
33	Brocade 3014 Switch	97	Brocade VDX 6720-60 Data Center Switch
34	Brocade 200E Switch	98	Brocade VDX 6730-76 Data Center Switch

Switch Type	Switch Name	Switch Type	Switch Name
37	Brocade 4020 Embedded Switch	108	Dell M8428-k FCoE Embedded Switch
38	Brocade 7420 SAN Router	109	Brocade 6510 Switch
40	Fibre Channel Routing (FCR) Front Domain	116	Brocade VDX 6710 Data Center Switch
41	Fibre Channel Routing, (FCR) Xlate Domain	117	Brocade 6547 Embedded Switch
42	Brocade 48000 Director	118	Brocade 6505 Switch
43	Brocade 4024 Embedded Switch	120	Brocade DCX 8510-8 Backbone
44	Brocade 4900 Switch	121	Brocade DCX 8510-4 Backbone
45	Brocade 4016 Embedded Switch	124	Brocade 5430 Switch
46	Brocade 7500 Switch	125	Brocade 5431 Switch
51	Brocade 4018 Embedded Switch	129	Brocade 6548 Switch
55.2	Brocade 7600 Switch	130	Brocade M6505 Switch
58	Brocade 5000 Switch	133	Brocade 6520 Switch
61	Brocade 4424 Embedded Switch	134	Brocade 5432 Switch
62	Brocade DCX Backbone	148	Brocade 7840 Switch

----End

5.1.2.2 Configuring Zones

Skip this section if you use direct connections.

Zone configuration is important for Fibre Channel switches. The configurations differ with the switch vendor, model, and version. For details, refer to the specific switch's *Configuration Guide*. The following explains the zone configuration procedure by using the Brocade 6510 switch as an example.

Step 1 Log in to the Brocade switch on a web browser.

On the web browser, enter the IP address of the Brocade switch and press **Enter**. The **Web Tools** switch login dialog box is displayed. Enter the account and password (**admin** and **password** by default) to log in.

Step 2 Check the port status on the switch.

In normal conditions, port indicators on the switch are steady green after the corresponding ports have been connected to hosts and storage arrays using optical fibers. This example uses ports 0, 1, 4, and 5, as shown in **Figure 5-3**.

Figure 5-3 Port status

Step 3 Go to the **Zone Admin** page.

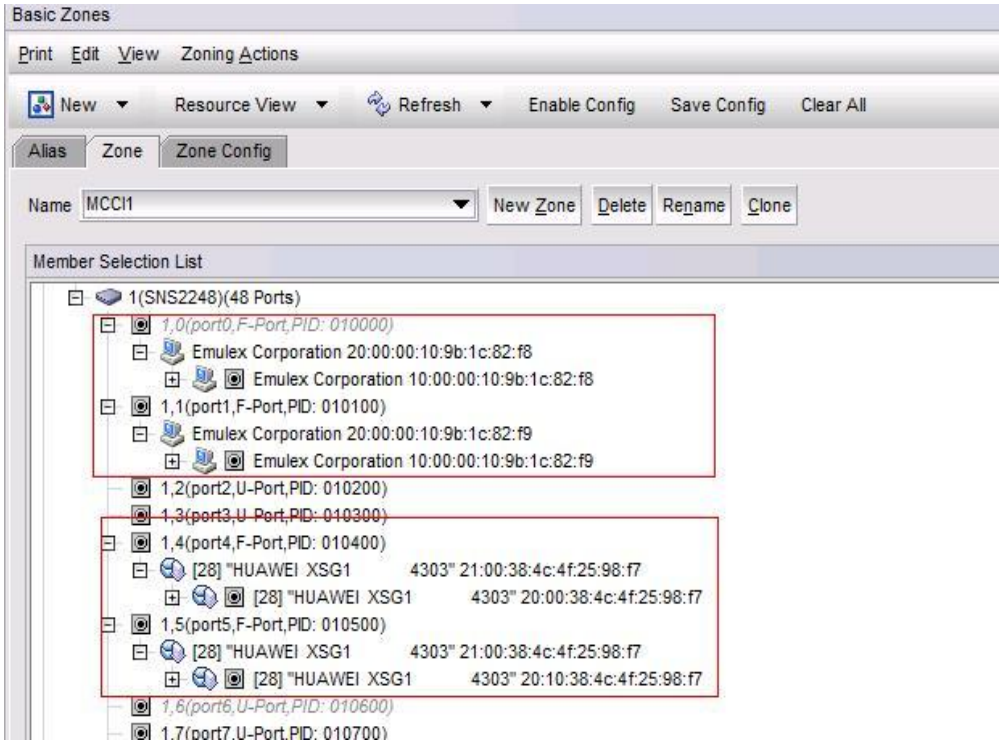
Choose **Configure** > **Zone Admin** from the main menu of **Web Tools**.

Figure 5-4 Zone Admin page

Step 4 Check whether the switch has identified hosts and storage systems.

On the **Zone Admin** page, click the **Zone** tab. In **Member Selection List**, check whether all related ports have been identified, as shown in **Figure 5-5**.

Figure 5-5 Identified ports

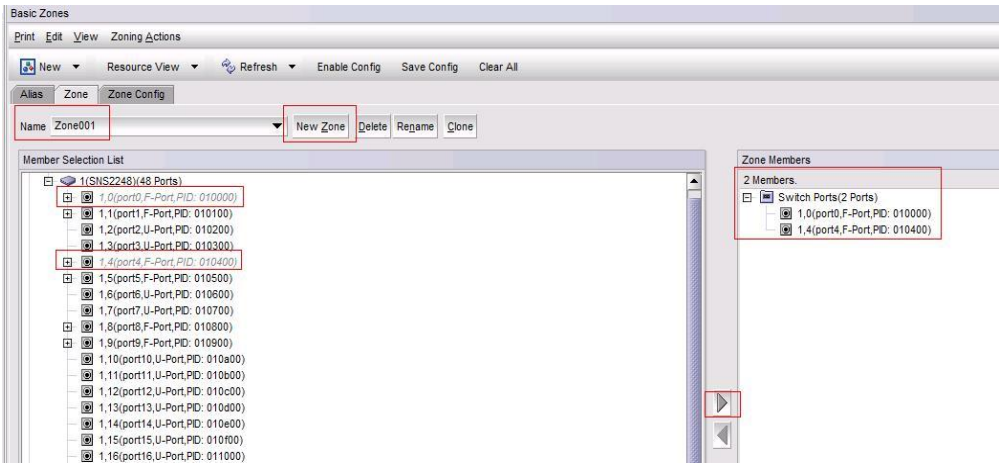


In this example, the hosts use ports 0 and 1, while the storage systems use ports 4 and 5. The display indicates that the switch has correctly identified the devices connected by the four ports.

Step 5 Create a zone.

On the **Zone** tab page, click **New Zone** and enter a name (**Zone001** in this example). Add port 0 (connecting to port P0 of a host) and port 4 (connecting to controller A of a storage system) to this zone, as shown in **Figure 5-6**.

Figure 5-6 Creating a zone

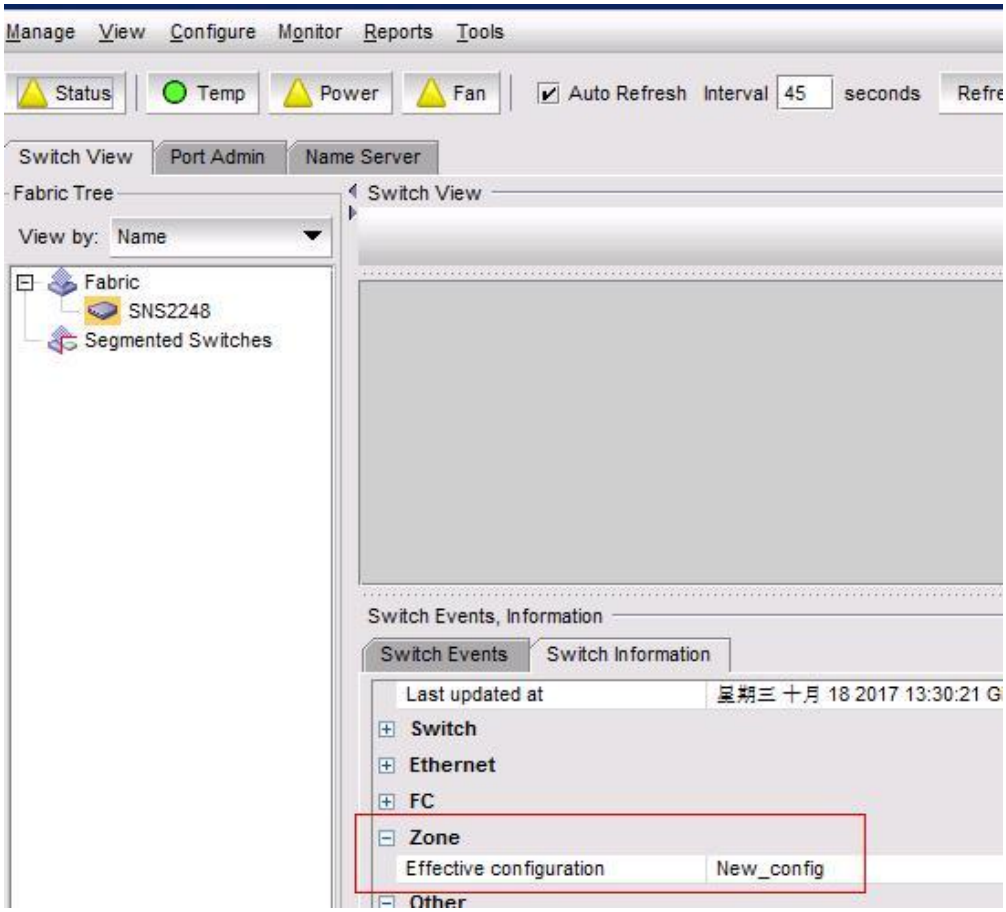


Use the same method to create **Zone002** to **Zone004**. Add ports 1 and 5 to **Zone0002**, ports 0 and 5 to **Zone003**, and ports 1 and 4 to **Zone004**.

Step 6 Add the new zones to the configuration file and activate them.

On the **Switch View** tab page, identify the effective configuration file, as shown in **Figure 5-7**.

Figure 5-7 Effective configuration file



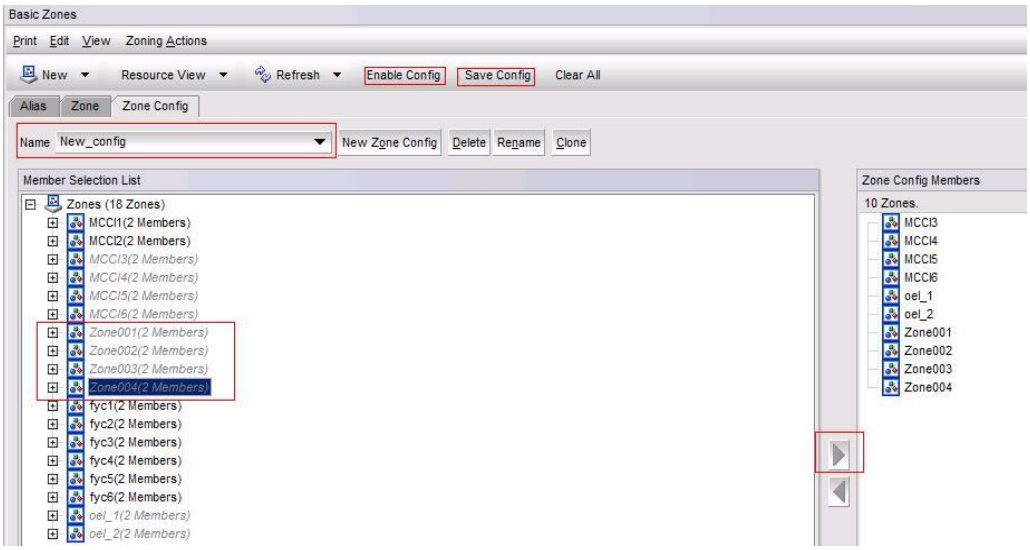
On the **Zone Admin** page, click the **Zone Config** tab. In the **Name** drop-down list, choose the effective configuration file **New_config**.

In **Member Selection List**, select **Zone001** to **Zone004** and add them to the configuration file.

Click **Save Config** to save the configuration and then click **Enable Config** for the configuration to take effect.

Figure 5-8 shows the configuration on the GUI.

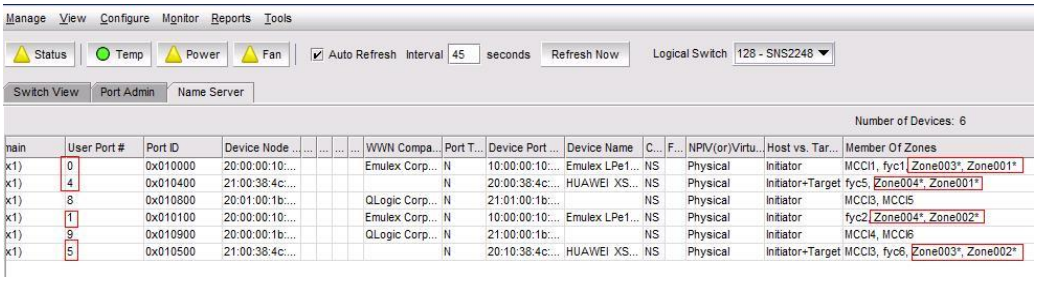
Figure 5-8 Adding zones to the configuration file



Step 7 Verify that the configuration has taken effect.

On the **Name Server** tab page, verify that the ports have been added to the zones and the zones have taken effect (marked * in the upper right corner), as shown in **Figure 5-9**.

Figure 5-9 Verifying the configuration



----End

5.1.3 Storage System Configuration

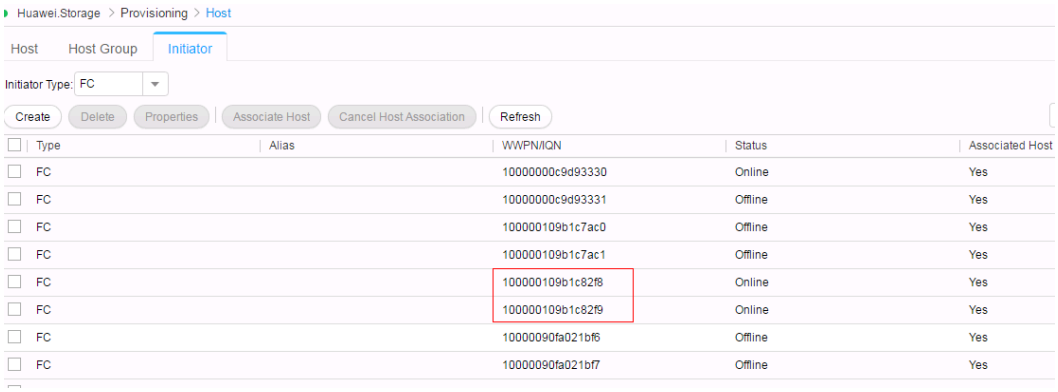
This section details how to add initiators to the hosts on the storage system. For other storage configurations, see the *Basic Storage Service Configuration Guide* corresponding to your storage system.

Step 1 Log in to the storage system on a web browser.

After you have configured the zones on the switch, log in to DeviceManager of the storage system and choose **Provisioning** > **Host** > **Initiator**. On the page that is displayed, select **FC** from the **Initiator Type** drop-down list. Check whether the host initiators have been discovered.

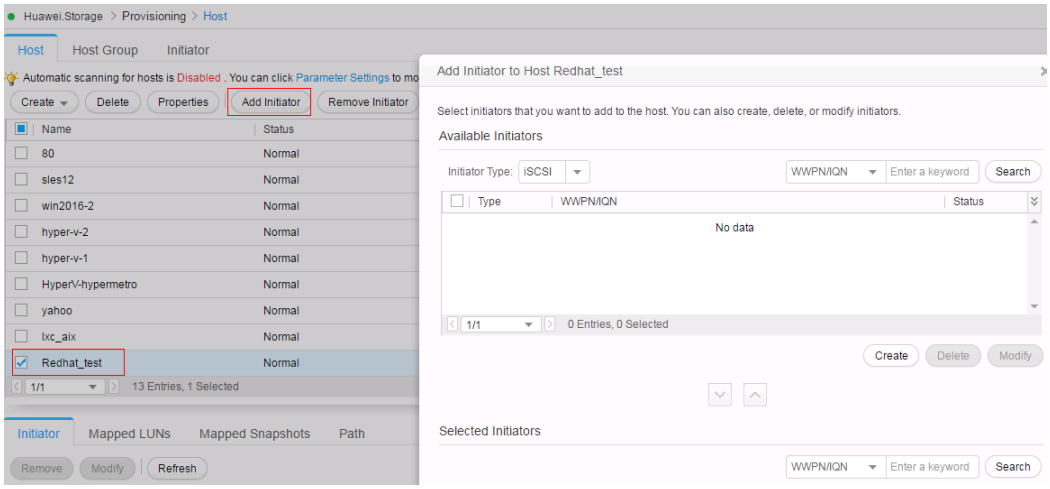
As shown in **Figure 5-10**, the host initiators have been discovered and are online.

Figure 5-10 Viewing initiators



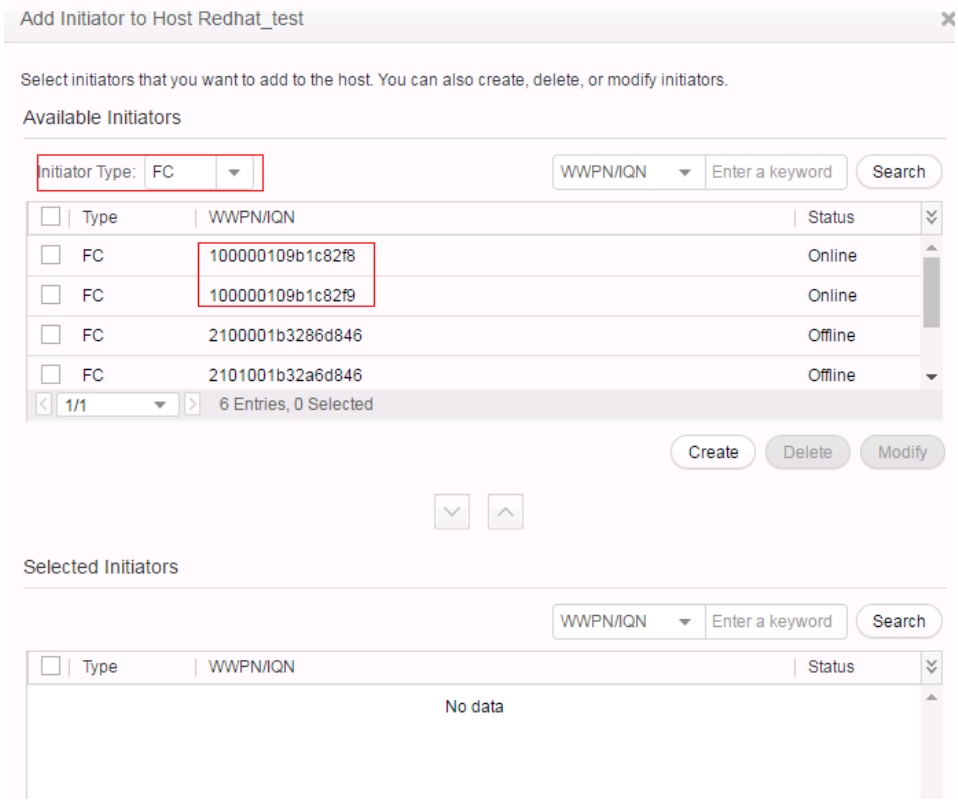
Step 2 Click the **Host** tab, select the host that was created on the storage system, and click **Add Initiator**.

Figure 5-11 Add Initiator dialog box



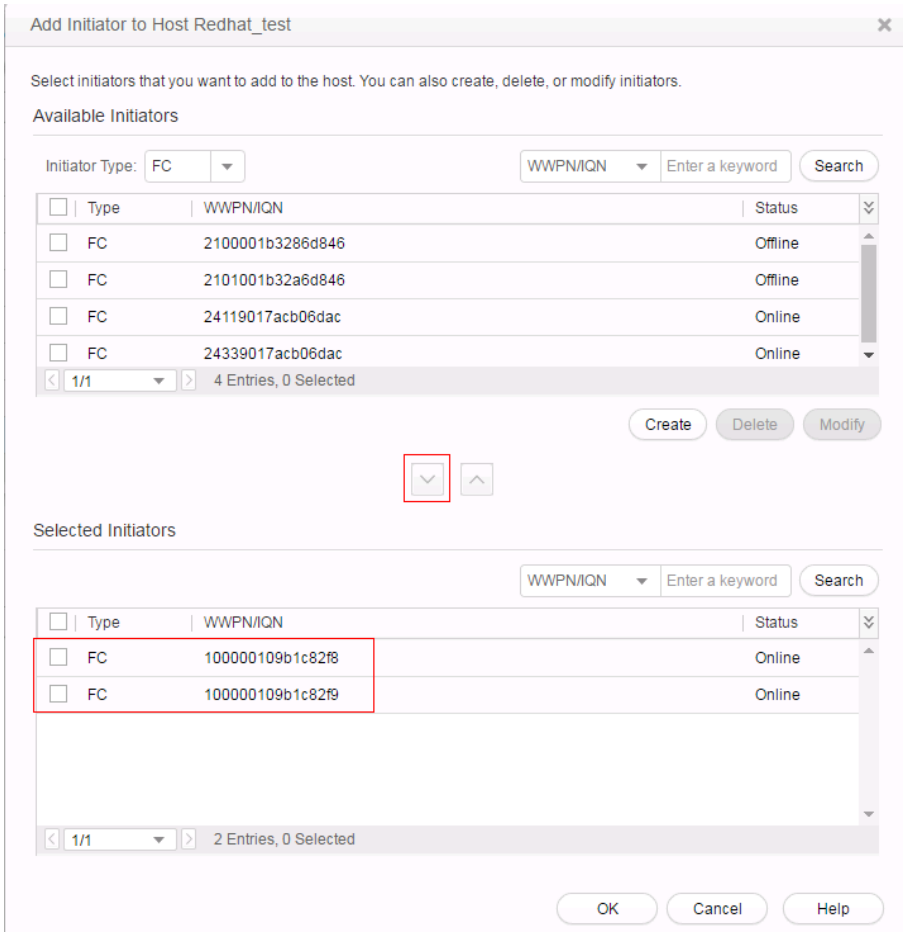
Step 3 Select **FC** from the **Initiator Type** drop-down list and find the host initiators' WWNs.

Figure 5-12 Selecting initiators



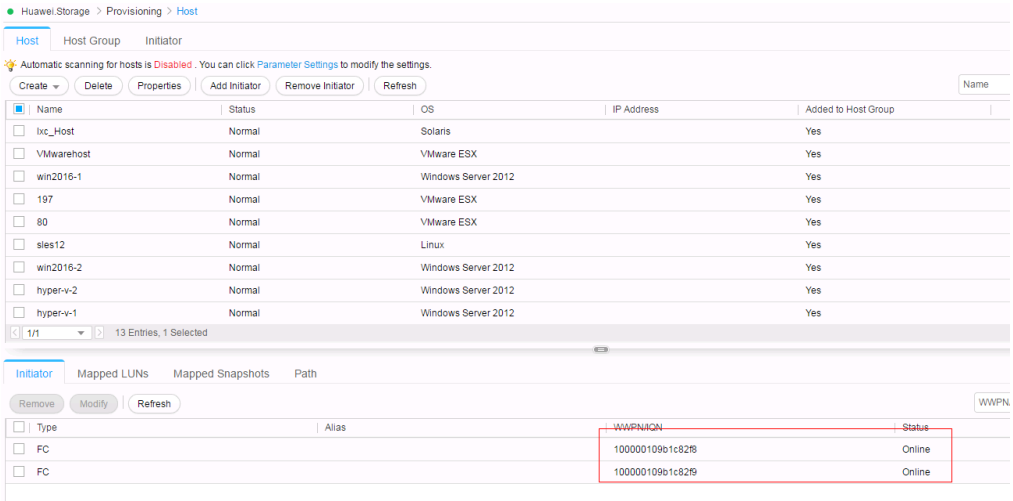
Step 4 Select the host initiators and add them to **Selected Initiators**.

Figure 5-13 Adding initiators



Step 5 Verify that the initiators have been added to the host correctly.

Figure 5-14 Verifying the configuration



As shown in **Figure 5-14**, the initiators have been added to the host successfully. The initiator properties depend on the operating system and multipathing software used by the hosts. For

details, see the storage-side configuration in the multipathing configuration section. After the initiators have been configured, you can scan for LUNs on the hosts to discover storage resources.

----End

5.2 Establishing iSCSI Connections

5.2.1 Host Configuration

Configuring Host IP Addresses

Step 1 Modify the parameters of host ports in their respective configuration files. **Figure 5-15** provides an example configuration for port eth1.

Figure 5-15 Configuring the IP address

```
[root@localhost ~]# cat /etc/sysconfig/network-scripts/ifcfg-eth1
DEVICE=eth1
HWADDR=00:46:4B:AE:04:D0
TYPE=Ethernet
UUID=b659005d-6934-4f61-842b-81b6e64bed73
ONBOOT=yes
NM_CONTROLLED=yes
BOOTPROTO=static
IPADDR=192.168.5.5
NETMASK=255.255.255.0
```

Step 2 Restart the network service.

Figure 5-16 Restarting the network

```
[root@localhost ~]# /etc/init.d/network restart
Shutting down interface eth0:                              [  OK  ]
Shutting down interface eth1:                              [  OK  ]
Shutting down loopback interface:                          [  OK  ]
Bringing up loopback interface:                            [  OK  ]
Bringing up interface eth0:  Determining if ip address [redacted] is already in use for device eth0...
                                                           [  OK  ]
Bringing up interface eth1:  Determining if ip address 192.168.5.5 is already in use for device eth1...
                                                           [  OK  ]
[root@localhost ~]#
[root@localhost ~]# ifconfig -a
eth0      Link encap:Ethernet  HWaddr 00:46:4B:AE:04:CF
          inet addr:[redacted]  Bcast:[redacted]  Mask:255.255.0.0
          inet6 addr: fe80::246:4bff:feae:4cf/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:1853 errors:0 dropped:0 overruns:0 frame:0
          TX packets:22 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:130910 (127.8 KiB)  TX bytes:3692 (3.6 KiB)

eth1      Link encap:Ethernet  HWaddr 00:46:4B:AE:04:D0
          inet addr:192.168.5.5  Bcast:192.168.5.255  Mask:255.255.255.0
          inet6 addr: fe80::246:4bff:feae:4d0/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:1391 errors:0 dropped:0 overruns:0 frame:0
          TX packets:10 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:97603 (95.3 KiB)  TX bytes:876 (876.0 b)
```

CAUTION

In some Red Hat versions, the network service and NetworkManager are both enabled by default, as shown in the following output:

```
[root@localhost ~]# chkconfig|grep -i network
NetworkManager  0:off 1:off 2:on    3:on    4:on    5:on    6:off
network         0:off 1:off 2:on    3:on    4:on    5:on    6:off
```

In this condition, network configurations may conflict and cannot take effect. Run the following command to disable NetworkManager.

```
service NetworkManager stop
```

To prevent NetworkManager from starting automatically after the system restarts, run the following command:

```
chkconfig NetworkManager off
```

----End

Checking iSCSI Software on the Host

After Red Hat is installed, run the **rpm -qa |grep iscsi** command to check the iSCSI software installation.

```
[root@root ~]# rpm -qa |grep iscsi
iscsi-initiator-utils-6.2.0.872-41.el6.x86_64
```

The output shows that iSCSI software is installed. If iSCSI software is not installed or the installed iSCSI software is of an early version, perform the following steps to install or upgrade the software:

Step 1 Obtain the iSCSI software package.

Step 2 Upload the software package to the host.

Step 3 Install the iSCSI software on the host. In this example, the software package is stored under the **/root** directory. Run the following command:

```
[root@root ~]# rpm -ivh /root/iscsi-initiator-utils-6.2.0.872-41.el6.x86_64.rpm
warning: /root/iscsi-initiator-utils-6.2.0.872-41.el6.x86_64.rpm: Header V3 RSA/
SHA256 Signature, key ID fd431d51: NOKEY
Preparing...                ########################################### [100%]
1:iscsi-initiator-utils  ########################################### [100%]
```

----End

Configuring Initiators Using the iscsiadm Command

Step 1 Start the iSCSI service.

```
[root@root ~]# /etc/init.d/iscsi start
```

Step 2 View the host initiator information.

```
[root@root ~]#cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1994-05.com.redhat:d0104b56adc6
```

The output shows that the host initiator name is **iqn.1994-05.com.redhat:d0104b56adc6**.

NOTE

An iSCSI initiator name must comply with the following format:

iqn.domaindate.reverse.domain.name:optional name

An iSCSI initiator name contains only:

- Special characters: hyphens (-), periods (.), and semicolons (:)
- Lower-case letters
- Digits

An iSCSI initiator name can contain a maximum of 223 characters.

Step 3 Query targets.

For example, if the service IP address of the storage port is 192.168.5.6, run the following command on the host to query targets:

```
[root@root ~]# iscsiadm -m discovery -t st -p 192.168.5.6
Starting iscsid:                                           [  OK  ]
192.168.5.6:3260,257 iqn.2006-08.com.huawei:oceanstor:
21000022a10b7bb2::192.168.5.6-20100
```

Step 4 Log in to the target.

```
[root@root ~]# iscsiadm -m node -p 192.168.5.6 -l
Logging in to [iface: default, target: iqn.2006-08.com.huawei:oceanstor:
21000022a10b7bb2::192.168.5.6-20100, portal: 192.168.5.6,3260] (multiple)
Login to [iface: default, target: iqn.2006-08.com.huawei:oceanstor:
21000022a10b7bb2::192.168.5.6-20100, portal: 192.168.5.6,3260] successful.
```

Step 5 Configure the **iscsi** service to run upon system startup.

```
[root@root ~]# chkconfig iscsi on
```

Step 6 Configure the host to automatically log in to the target upon startup.

```
[root@root ~ ]# iscsiadm -m node -o update -n node.startup -v automatic
```

Step 7 If CHAP authentication is not required between the storage system and host, the host initiator configuration is completed. If CHAP authentication is required, run the **iscsiadm** command.

```
[root@root ~]#iscsiadm –m node -o update -p 192.168.5.6 –n
node.session.auth.authmethod -v CHAP
[root@root ~]# iscsiadm -m node -o update -p 192.168.5.6 -n
node.session.auth.username -v root
[root@root ~]# iscsiadm -m node -o update -p 192.168.5.6 -n
node.session.auth.password -v huawei123456
[root@root ~]# /etc/init.d/iscsi restart
Stopping iSCSI daemon:
iscsid dead but pid file exists                            [  OK  ]
Turning off network shutdown. Starting iSCSI daemon:       [  OK  ]
                                                           [  OK  ]
Setting up iSCSI targets: Logging in to [iface: default, target: iqn.
2006-08.com.huawei:oceanstor:21000022a10b7bb2::192.168.5.6-20100, portal:
192.168.5.6,3260]
Login to [iface: default, target: iqn.2006-08.com.huawei:oceanstor:
21000022a10b7bb2::192.168.5.6-20100, portal: 192.168.5.6,3260]: successful
                                                           [  OK  ]
```

NOTE

When the **iscsiadm** command is executed, the user name and password of the initiator that was added to the storage system are required. If the user name and password used in this command are different from the initiator's, you will fail to establish connections between the storage system and the host.

The command syntax is as follows:

```
iscsiadm -m node -o update -p targetip -n node.session.auth.authmethod -v CHAP
iscsiadm -m node -o update -p targetip -n node.session.auth.username -v username
iscsiadm -m node -o update -p targetip -n node.session.auth.password -v password
```

You are advised to run the **iscsiadm** command to modify related parameters. Do not modify the parameters using the configuration file.

----End

Configuring Initiators using the iscsi.conf Configuration File

On some Red Hat operating systems of an early version, the **iscsiadm** command is unavailable. You need to use the configuration file to configure initiators.

Step 1 Start the iSCSI service.

```
[root@localhost ~]# /etc/init.d/iscsi start
```

Step 2 View the host initiator information.

```
[root@localhost ~]#cat /etc/iscsi/initiatorname.iscsi
InitiatorName=iqn.1994-05.com.redhat:d0104b56adc6
```

The output shows that the name of the host initiator is **iqn.1994-05.com.redhat:d0104b56adc6**.

NOTE

An iSCSI initiator name must comply with the following format:

iqn.domaindate.reverse.domain.name:optional name

An iSCSI initiator name contains only:

- Special characters: hyphens (-), periods (.), and semicolons (:)
- Lower-case letters, for example, a to z
- Digits, for example, 0 to 9

An iSCSI initiator name can contain a maximum of 223 characters.

Step 3 Add the port address of the storage system.

```
[root@localhost ~]# vi /etc/iscsi.conf
```

In the configuration file, add the IP address of the storage system port connected to the host in **DiscoveryAddress=xx.xx.xx.xx**. Ensure that no space is left in front of **DiscoveryAddress=xx.xx.xx.xx**.

Step 4 Restart the iSCSI service.

```
[root@localhost ~]# /etc/init.d/iscsi restart
Searching for iscsi-based multipath maps
Found 0 maps
Stopping iscsid: iscsid not running

Checking iscsi config:                                     [  OK  ]
Loading iscsi driver:                                      [  OK  ]
Starting iscsid:                                           [  OK  ]
```

Step 5 Modify the initiator startup configuration.

Run the **chkconfig iscsi** on command to configure the **iscsi** service to run upon system startup.

Step 6 Configure the host to automatically log in to the target upon startup.

```
[root@root ~ ]# iscsiadm -m node -o update -n node.startup -v automatic
```

Step 7 If CHAP authentication is not required between the storage system and host, skip this step.

In the **/etc/iscsi.conf** file, add the target IP address for CHAP authentication and specify the user name and password of the initiator used for target authentication. Related parameters are as follows:

DiscoveryAddress=xx.xx.xx.xx

OutgoingUsername=xxx

OutgoingPassword=xxxx

CAUTION

Remove only pounds (#) in front of **OutgoingUsername/OutgoingPassword** and keep the space in front of them unchanged.

----End

5.2.2 (Optional) Switch Configuration

This section describes how to configure Ethernet switches, including configuring VLANs and binding ports. Skip this section if you use direct connections.

Configuring VLANs

On an Ethernet network to which many hosts are connected, a large number of broadcast packets are generated during the host communication. Broadcast packets sent from one host will be received by all other hosts on the network, consuming more bandwidth. Moreover, all hosts on the network can access each other, resulting data security risks.

To save bandwidth and prevent security risks, hosts on an Ethernet network are divided into multiple logical groups. Each logical group is a VLAN. The following uses Huawei Quidway 2700 Ethernet switch as an example to explain how to configure VLANs.

In the following example, two VLANs (VLAN 1000 and VLAN 2000) are created. VLAN 1000 contains ports GE 1/0/1 to 1/0/16. VLAN 2000 contains ports GE 1/0/20 to 1/0/24.

Step 1 Go to the system view.

```
<Quidway>system-view
System View: return to User View with Ctrl+Z.
```

Step 2 Create VLAN 1000 and add ports to it.

```
[Quidway]VLAN 1000
[Quidway-vlan1000]port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/16
```

Step 3 Configure an IP address for VLAN 1000.

```
[Quidway-vlan1000]interface VLAN 1000
[Quidway-Vlan-interface1000]ip address 192.168.5.1 255.255.255.0
```

Step 4 Create VLAN 2000, add ports, and configure an IP address.

```
[Quidway]VLAN 2000
[Quidway-vlan2000]port GigabitEthernet 1/0/20 to GigabitEthernet 1/0/24
[Quidway-vlan2000]interface VLAN 2000
[Quidway-Vlan-interface2000]ip address 192.168.6.1 255.255.255.0
```

----End

Binding Ports

When storage systems and hosts are connected in point-to-point mode, existing bandwidth may be insufficient for storage data transmission. Moreover, devices cannot be redundantly

connected in point-to-point mode. To address these problems, ports are bound (link aggregation) to improve bandwidth and balance load among multiple links.

Three Ethernet link aggregation modes are available:

- Manual aggregation

  Ports are added to an aggregation group by running a command manually. Ports added to the aggregation group must have the same link type.

- Static aggregation

  Ports are added to an aggregation group by running a command manually. Ports added to the aggregation group must have the same link type and LACP enabled.

- Dynamic aggregation

  The protocol dynamically adds ports to an aggregation group. Ports added in this way must have LACP enabled and the same speed, duplex mode, and link type.

Table 5-2 compares these aggregation modes.

Table 5-2 Comparison among link aggregation modes

Link Aggregation Mode	Packet Exchange	Port Detection	CPU Usage
Manual aggregation	No	No	Low
Static aggregation	Yes	Yes	High
Dynamic aggregation	Yes	Yes	High

Huawei OceanStor storage devices support 802.3ad link aggregation (dynamic aggregation). In this link aggregation mode, multiple network ports are in an active aggregation group and work in duplex mode and at the same speed. After binding iSCSI host ports on a storage device, enable aggregation for their peer ports on the switch. Otherwise, links are unavailable between the storage device and the switch.

This section uses switch ports GE 1/0/1 and GE 1/0/2 and the storage system's ports P2 and P3 as an example to explain how to bind ports.

The port binding method differs with the OceanStor system version. For details, refer to the specific storage product documentation. The following steps use OceanStor V3 V300R003 as an example.

Step 1 Log in to DeviceManager and choose **Provisioning** > **Port**.

Step 2 Bind ports.

1. Select the ports that you want to bind and choose **More** > **Bond Ports**.

   The **Bond Port** dialog box is displayed.

2. Enter a **Bond Name**, select the target ports, and click **OK**.
3. In the security alert dialog box that is displayed, select **I have read and understand the consequences associated with performing this operation** and click **OK**.

After the storage system ports are bound, configure link aggregation on the switch using the following command:

```
<Quidway>system-view
System View: return to User View with Ctrl+Z.
[Quidway-Switch]interface GigabitEthernet 1/0/1
[Quidway-Switch-GigabitEthernet1/0/1]lacp enable
LACP is already enabled on the port!
[Quidway-Switch-GigabitEthernet1/0/1]quit
[Quidway-Switch]interface GigabitEthernet 1/0/2
[Quidway-Switch-GigabitEthernet1/0/2]lacp enable
LACP is already enabled on the port!
[Quidway-Switch-GigabitEthernet1/0/2]quit
```

After the command is executed, LACP is enabled for ports GE 1/0/1 and GE 1/0/2. Then the ports can be automatically detected and added to an aggregation group.

----End

5.2.3 Storage System Configuration

Configuring Port IP Addresses

Configure Ethernet port parameters to ensure proper communication between the storage system and application server.

Note the following items when setting the properties of an Ethernet port:

- The default internal heartbeat IP addresses of a dual-controller storage system are **127.127.127.10** and **127.127.127.11**, and those of a four-controller storage system are **127.127.127.10**, **127.127.127.11**, **127.127.127.12**, and **127.127.127.13**. Therefore, the IP address of the router must not be in the 127.127.127.XXX segment and the gateway address must not be **127.127.127.10**, **127.127.127.11**, **127.127.127.12**, or **127.127.127.13**. Otherwise, routing will fail. Internal heartbeat links are established between controllers for these controllers to detect each other's working status. You do not need to separately connect cables. In addition, internal heartbeat IP addresses have been assigned before delivery, and you cannot change these IP addresses.
- The IP address of the Ethernet port cannot be in the same network segment as that of a management network port.
- The IP address of the Ethernet port cannot be in the same network segment as that of a maintenance network port.
- If the Ethernet port connects to an application server, the IP address of the Ethernet port must be in the same network segment as that of the service network port on the application server. If the Ethernet port connects to another storage device, the IP address of the Ethernet port must be in the same network segment as that of the Ethernet port on the other storage device. Add routes if available IP addresses in the desired segment are insufficient.

To configure an IP address for an Ethernet port, perform the following steps:

Step 1 Go to the **Ethernet Port** dialog box.

1. In the basic information area of the function pane, click the device icon.
2. In the middle function pane, click the cabinet whose Ethernet ports you want to view.
3. Click the controller enclosure where the desired Ethernet host ports reside. The controller enclosure view is displayed.

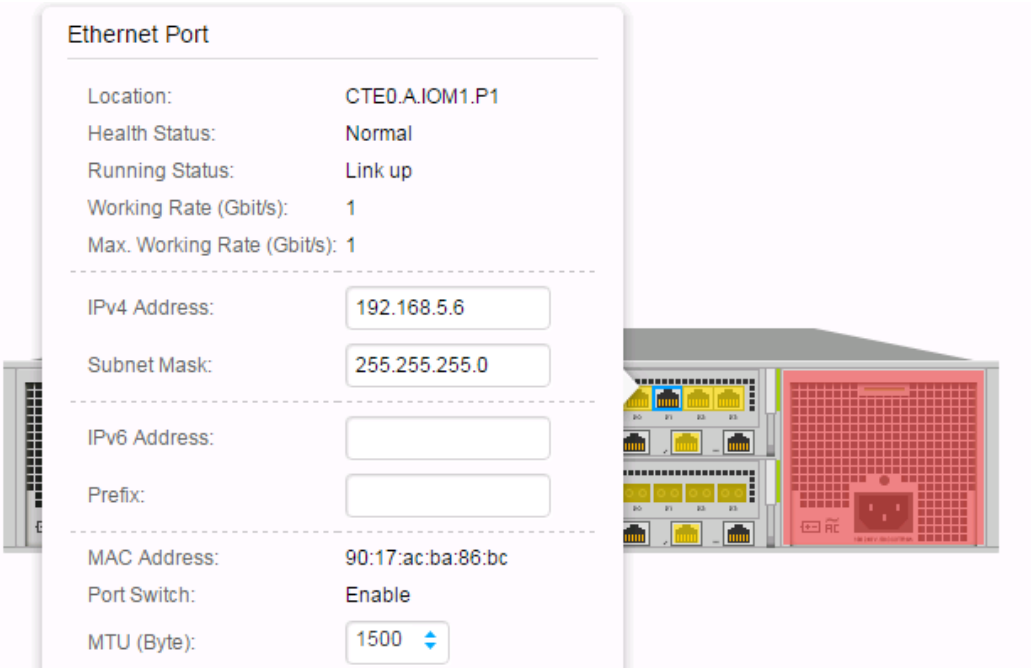
4. Click  to switch to the rear view.
5. Click the Ethernet port whose information you want to modify.

   The **Ethernet Port** dialog box is displayed.
6. Click **Modify**.

Step 2 Modify the Ethernet port, as shown in **Figure 5-17**.

1. In **IPv4 Address** or **IPv6 Address**, enter an IP address for the Ethernet port.
2. In **Subnet Mask** or **Prefix**, enter a subnet mask or prefix for the Ethernet port.
3. In **MTU (Byte)**, enter the maximum size of data packet that can be transferred between the Ethernet port and the host. The value is an integer ranging from 1500 to 9216.

Figure 5-17 Configuring an IP address



Step 3 Confirm the Ethernet port modification.

1. Click **Apply**.

   The **Danger** dialog box is displayed.
2. Confirm the information in the dialog box and select **I have read and understand the consequences associated with performing this operation.**
3. Click **OK**.

   The **Success** dialog box is displayed, indicating that the operation is successful.
4. Click **OK**.

----End

(Optional) Adding Routes

If iSCSI networking is used and data needs to be transmitted across network segments, you need to configure routes.

Step 1 Log in to DeviceManager.

Step 2 Choose **Provisioning** > **Port** > **Ethernet Ports**.

Step 3 Select the Ethernet port for which you want to add a route and click **Route Management**.

The **Route Management** dialog box is displayed.

Step 4 Configure the route information for the Ethernet port.

1. In **IP Address**, select the IP address of the Ethernet port.
2. Click **Add**.

   The **Add Route** dialog box is displayed.
3. In **Type**, select the type of the route to be added

   There are three route options:
   - Default route

     Data is forwarded through this route by default if no preferred route is available. The destination address field and the target mask field (IPv4) or prefix (IPv6) of the default route are automatically set to 0. To use this option, you only need to add a gateway.
   - Host route

     A route to an individual host. The destination mask (IPv4: 255.255.255.255) or prefix (IPv6: 128) of the host route is automatically set. To use this option, add the destination address and a gateway.
   - Network segment route

     A route to a network segment. You need to add the destination address, destination address mask (IPv4) or prefix (IPv6), and gateway. For example, the destination address is 172.17.0.0, destination address mask is 255.255.0.0, and gateway is 172.16.0.1.
4. Set **Destination Address.**

   Set **Destination Address** to the IPv4 or IPv6 (depending on which one you use) address or network segment of the application server's service network port or that of the other storage system's Ethernet port.
5. Set **Destination Mask** (IPv4) or **Prefix** (IPv6).
   - If an IPv4 address is used, this parameter specifies the subnet mask of the IP address for the service network port on the application server or the other storage device.
   - If an IPv6 address is used, this parameter specifies the prefix of the IPv6 address for the application server's service network port or that of the other storage system's Ethernet port.
6. In **Gateway**, enter the gateway of the local storage system's Ethernet port IP address.

Step 5 Click **OK**. The route information is added to the route list.

A security alert dialog box is displayed.

Step 6 Confirm the information in the dialog box and select **I have read and understand the consequences associated with performing this operation.**

Step 7 Click **OK**.

The **Success** dialog box is displayed, indicating that the operation is successful.

NOTE

To remove a route, select it and click **Remove**.

Step 8 Click **Close**.

----End

Adding an Initiator to a Host

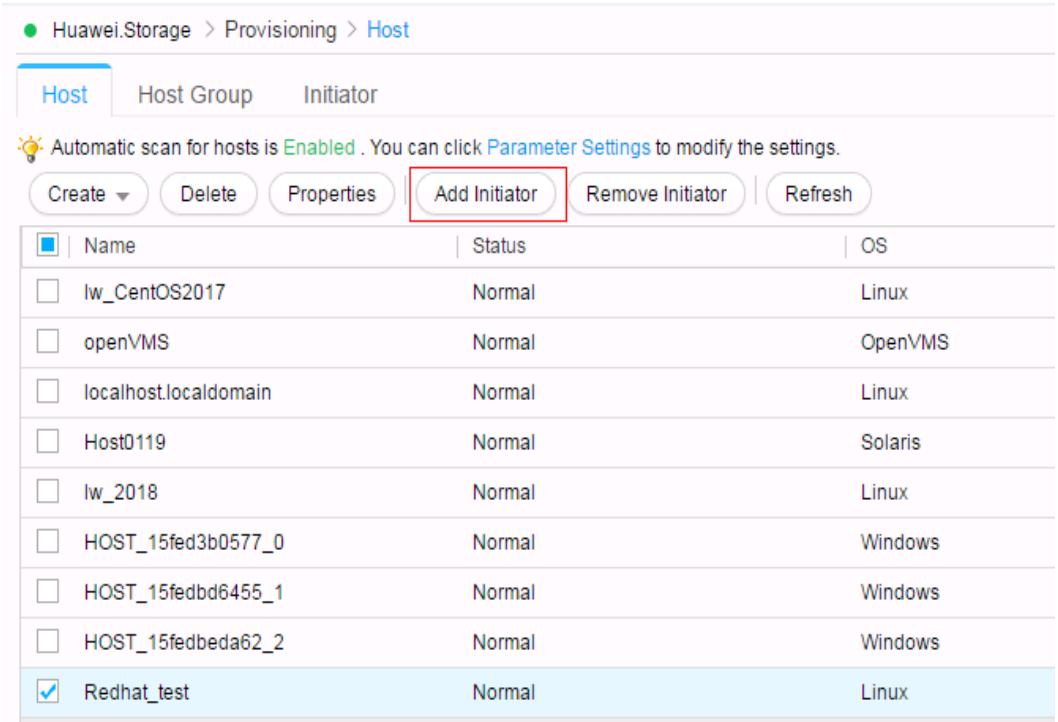
This section details how to add initiators to the hosts on the storage system. For other storage configurations, see the *Basic Storage Service Configuration Guide* corresponding to your storage system.

Step 1 Log in to DeviceManager.

Step 2 Choose **Provisioning** > **Host**.

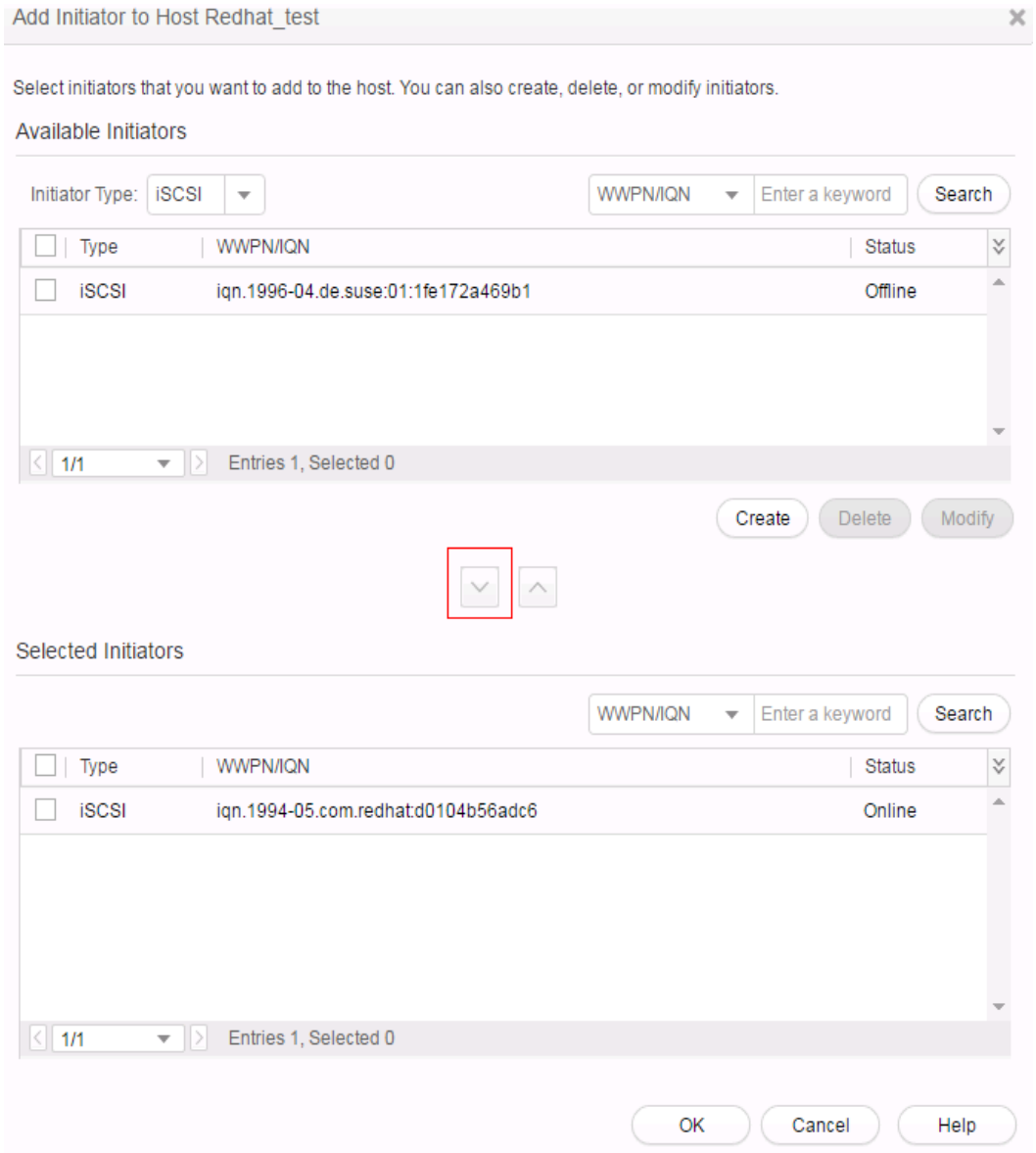
Step 3 Select the target host and click **Add Initiator**.

Figure 5-18 Selecting a host



Step 4 Select the initiator and click to add it to **Selected Initiators**.

Figure 5-19 Adding an initiator



Step 5 Click **OK**.

----End

The initiator properties depend on the operating system and multipathing software used by the hosts. For details, see the storage-side configuration in the multipathing configuration section. After the initiators have been configured, you can scan for LUNs on the hosts to discover storage resources.

6 Configuring Multipathing in HyperMetro Scenarios

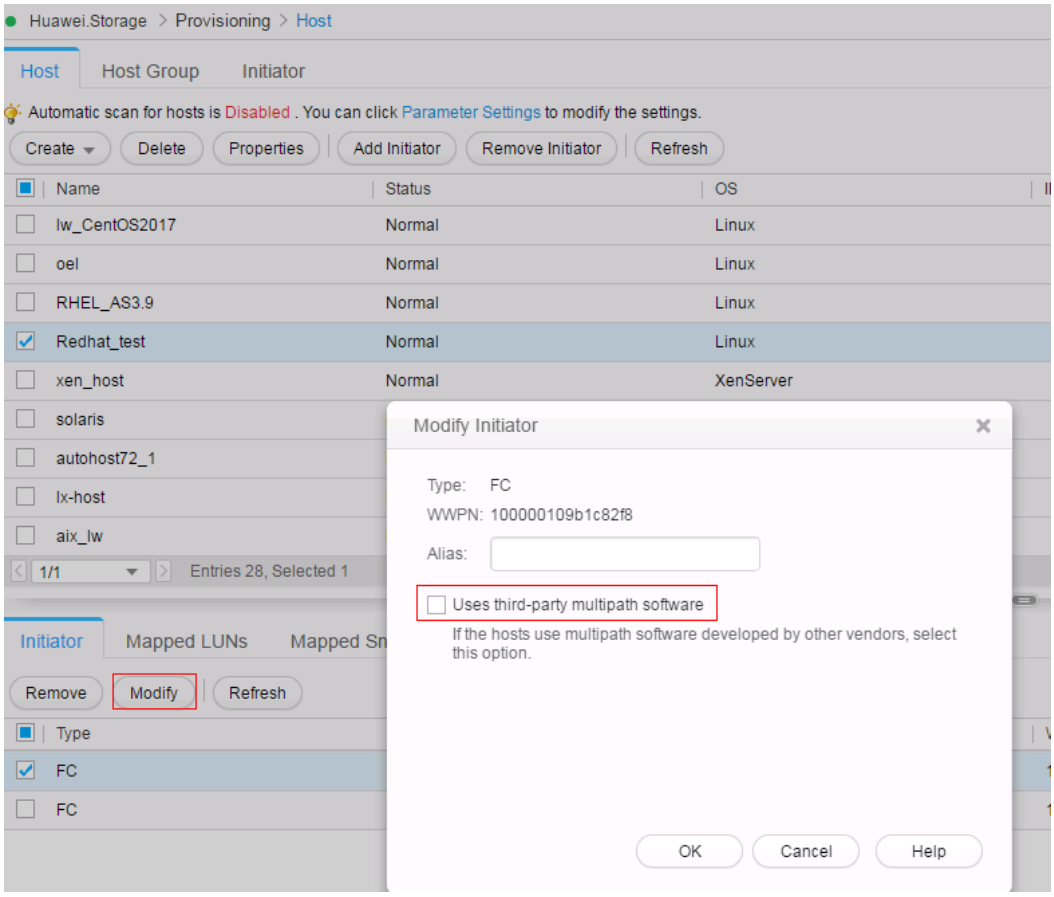
This chapter describes the multipathing software configurations on the hosts and storage systems. For details about how to configure HyperMetro services, see the *HyperMetro Feature Guide*.

6.1 UltraPath

6.1.1 Storage System Configuration

If you use UltraPath, retain the default initiator settings. Do not select **Uses third-party multipath software**.

Figure 6-1 Initiator setting when UltraPath is used



6.1.2 Host Configuration

Install UltraPath by following instructions in the *OceanStor UltraPath for Linux User Guide*.

In UltraPath, set the HyperMetro working mode to preferred storage array mode. In this mode, the local storage array is preferred in processing host services. The remote storage array is used only when the local array is faulty. This improves the service response speed and reduces the access latency.

Table 6-1 lists the command for setting the HyperMetro working mode.

Table 6-1 Setting the HyperMetro working mode

Operating System	Command	Example
Linux	**set hypermetroworking-mode=**{*priority*|*balance*}**primary_array_id=***ID*	**upadmin set hypermetro workingmode=priority primary_array_id=0**

Table 6-2 describes the parameters.

Table 6-2 Parameter description

Parameter	Description	Default Value
workingmode={ *priority* | *balance* }	HyperMetro working mode. ● **priority**: preferred storage array mode ● **balance**: load balancing mode **NOTE** If you set the HyperMetro working mode for a specific virtual LUN first and then the global HyperMetro working mode for the storage system, the working mode for the virtual LUN remains unchanged.	**priority** **priority** is recommended. **balance** is applicable when two active-active data centers are in the same building.
primary_array_id =*ID*	ID of the preferred storage array. The ID is allocated by UltraPath. The storage array that is in the same data center as the application hosts is preferred. Run the command to obtain the storage array ID: ● Linux: **upadmin show array** ● Windows/AIX/Solaris: **upadm show array** **NOTE** ● In **priority** mode, this parameter indicates the storage array to which I/Os are preferentially delivered. ● In **balance** mode, this parameter indicates the storage array where the first slice section resides.	None **NOTE** Mapping relationship between application hosts and storage arrays: ● Storage array A is the preferred array for all application hosts in data center A. ● Storage array B is the preferred array for all application hosts in data center B.

6.1.3 Verification

Run the **upadmin show upconfig** command. If the command output contains the following information, the configuration is successful.

```
HyperMetro WorkingMode : read write within primary array
```

Figure 6-2 provides an example.

Figure 6-2 Verifying the HyperMetro working mode

```
[root@localhost ~]# upadmin show upconfig
=========================================================
UltraPath Configuration
=========================================================
Basic Configuration
    Working Mode : load balancing within controller
    LoadBalance Mode : min-queue-depth
    Loadbanlance io threshold : 100
    LUN Trespass : on

Advanced Configuration
    Io Retry Times : 10
    Io Retry Delay : 0
    Faulty path check interval : 10
    Idle path check interval : 60
    Failback Delay Time : 600
    Io Suspension Time : 60
    Max io retry timeout : 1800
    Performance Record : off

Path reliability configuration
    Timeout degraded statistical time : 600
    Timeout degraded threshold : 1
    Timeout degraded path recovery time : 1800
    Intermittent IO error degraded statistical time : 300
    Min. I/Os for intermittent IO error degraded statistical : 5000
    Intermittent IO error degraded threshold : 20
    Intermittent IO error degraded path recovery time : 1800
    Intermittent fault degraded statistical time : 1800
    Intermittent fault degraded threshold : 3
    Intermittent fault degraded path recovery time : 3600
    High latency degraded statistical time : 300
    High latency degraded threshold : 1000
    High latency degraded path recovery time : 3600
    Sensitive delayed degraded threshold : 30000
    Sensitive delayed degraded recovery time : 120

HyperMetro configuration
    HyperMetro Primary Array SN : 210235982610F4000017
    HyperMetro WorkingMode : read write within primary array
    HyperMetro Split Size : 128MB
```

6.2 OS Native Multipathing Software

This section describes the concepts that may be used in configuring OS native multipathing software.

HyperMetro Working Modes

Typically, HyperMetro works in load balancing mode or local preferred mode. The typical working modes are valid only when both the storage system and host use ALUA. It is advised to set the host's path selection policy to round-robin. If HyperMetro works in load balancing mode, the host's path selection policy must be round-robin.

HyperMetro storage arrays can be classified into a local and a remote array by their distance to the host. The one closer to the host is the local array and the other one is the remote array.

Table 6-3 describes the configuration methods and application scenarios of the typical working modes.

Table 6-3 HyperMetro working modes

Working Mode	Configuration Method	Application Scenario
Load balancing mode	Enable ALUA on the host and set the path selection policy to round-robin. Configure a switchover mode that supports ALUA for both HyperMetro storage arrays' initiators that are added to the host. Set the path type for both storage arrays' initiators to the optimal path.	The distance between both HyperMetro storage arrays is less than 1 km. For example, they are in the same equipment room or on the same floor.
Local preferred mode	Enable ALUA on the host. It is advised to set the path selection policy to round-robin. Configure a switchover mode that supports ALUA for both HyperMetro storage arrays' initiators that are added to the host. Set the path type for the local storage array's initiators to the optimal path and that for the remote storage array's initiators to the non-optimal path.	The distance between both HyperMetro storage arrays is greater than 1 km. For example, they are in different locations or data centers.

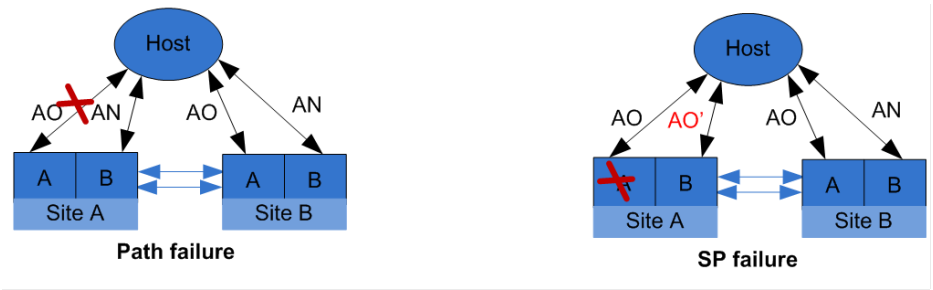
Working Principles and Failover

When ALUA works, the host multipathing software divides the physical paths to disks into Active Optimized (AO) and Active Non-optimized (AN) paths. The host delivers services to the storage system via the AO paths preferentially.

- An AO path is the optimal I/O access path and is between the host and a working controller.
- An AN path is the suboptimal I/O access path and is between the host and a non-working controller.

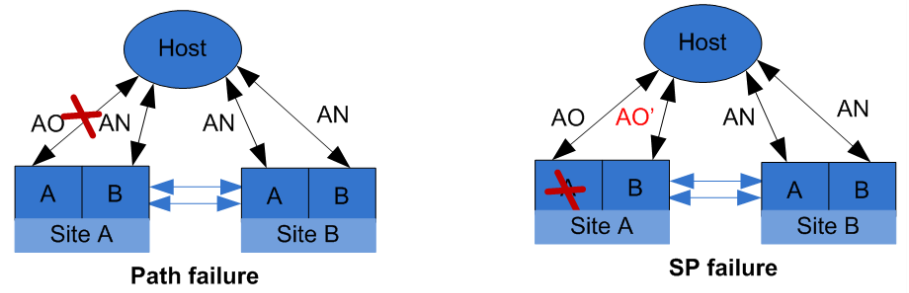
When HyperMetro works in load balancing mode, the host multipathing software selects the paths to the working controllers on both HyperMetro storage arrays as the AO paths, and those to the other controllers as the AN paths. The host accesses the storage arrays via the AO paths. If an AO path fails, the host delivers I/Os to another AO path. If the working controller of a storage array fails, the system switches the other controller to the working mode and maintains load balancing.

Figure 6-3 Load balancing mode



When HyperMetro works in local preferred mode, the host multipathing software selects the paths to the working controller on the local storage array as the AO paths. This ensures that the host delivers I/Os only to the working controller on the local storage array, reducing link consumption. If all AO paths fail, the host delivers I/Os to the AN paths on the non-working controller. If the working controller of the local storage array fails, the system switches the other controller to the working mode and maintains the local preferred mode.

Figure 6-4 Local preferred mode



Initiator

Table 6-4 describes the initiator parameters.

Table 6-4 Initiator parameters

Parameter	Description	Example
Uses third-party multipath software	This parameter is displayed only after an initiator has been added to the host. If LUNs have been mapped to the host before you enable or disable this parameter, restart the host after you configure this parameter. You do not need to enable this parameter on a host with UltraPath.	Enabled

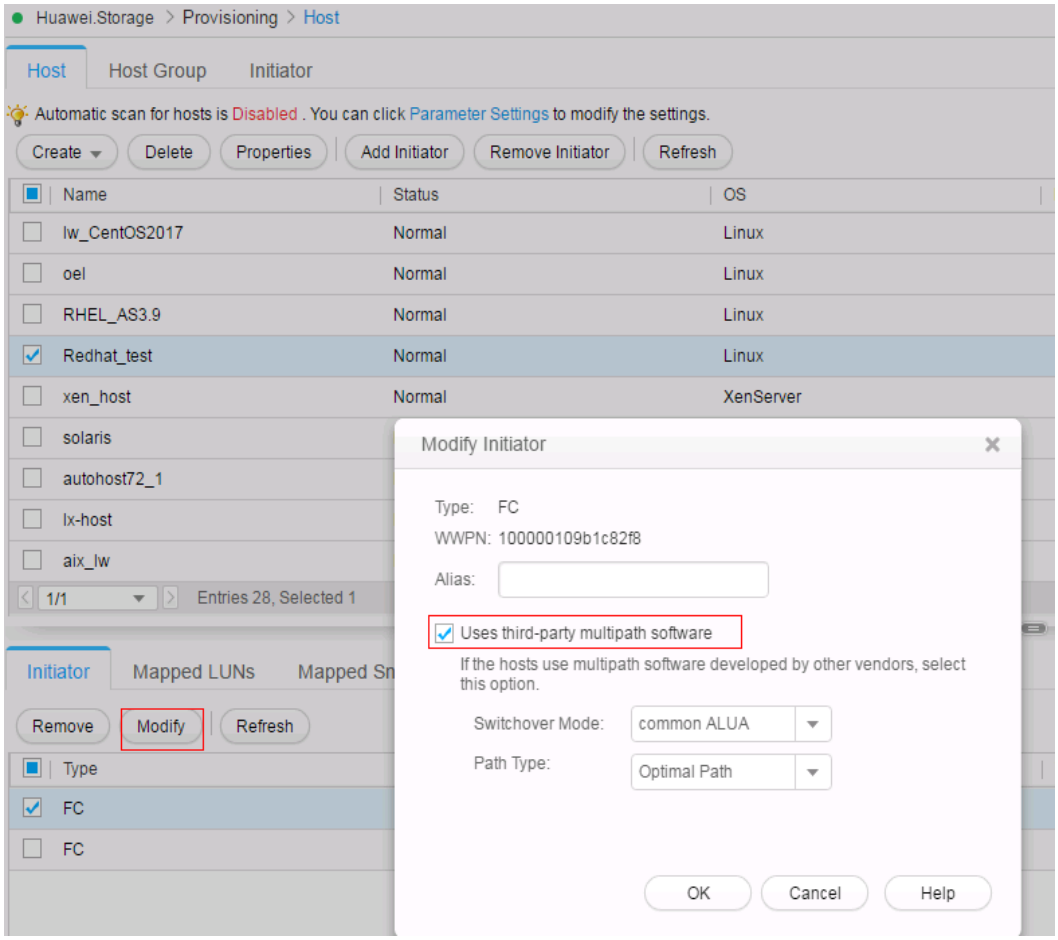
Parameter	Description	Example
Switchover Mode	Path switchover mode The system supports the following modes: ● **early-version ALUA**: default value of **Switchover Mode** for an upgrade from an earlier version to the current version. Detailed requirements are as follows: – The storage system must be upgraded from V300R003C10 and earlier to V300R003C20 or V300R006C00SPC100 and later; from V300R005 to V300R006C00SPC100 and later; from Dorado V300R001C00 to Dorado V300R001C01SPC100 and later. – Before the upgrade, the storage system must have a single or dual controllers and has enabled ALUA. ● **common ALUA**: Detailed requirements are as follows: – The storage system version must be V500R007C00 and later, V300R003C20 and later, V300R006C00SPC100 and later, or Dorado V300R001C01SPC100 and later. – The OS of the host that connects to the storage system must be SUSE, Red Hat 6.X, Windows Server 2012 (using Emulex HBAs), Windows Server 2008 (using Emulex HBAs), or HP-UX 11i V3. ● **ALUA not used**: does not support ALUA or HyperMetro. This mode is used when a host such as HP-UX 11i V2 does not support ALUA or ALUA is not needed. ● **Special mode**: supports ALUA and has multiple values. It is used by host operating systems that are not supported by the **common ALUA** mode. Detailed requirements are as follows: – The storage system version must be V500R007C00 and later, V300R003C20 and later, V300R006C00SPC100 and later, or Dorado V300R001C01SPC100 and later. – The OS of the host that connects to the storage system must be VMware, AIX, Red Hat 7.X, Windows Server 2012 (using QLogic HBAs), or Windows Server 2008 (using QLogic HBAs).	

Parameter	Description	Example
Special mode type	Special modes support ALUA and apply to V500R007C00 and later, V300R003C20 and later, V300R006C00SPC100 and later, or Dorado V300R001C01SPC100 and later. The detailed requirements are as follows: ● **Mode 0**: – The host and storage system must be connected using a Fibre Channel network. – The OS of the host that connects to the storage system is Red Hat 7.X, Windows Server 2012 (using QLogic HBAs), or Windows Server 2008 (using QLogic HBAs). ● **Mode 1**: – The OS of the host that connects to the storage system is AIX or VMware. – HyperMetro works in load balancing mode. ● **Mode 2**: – The OS of the host that connects to the storage system is AIX or VMware. – HyperMetro works in local preferred mode.	Mode 0
Path Type	The value can be either **Optimal Path** or **Non-Optimal Path**. ● When HyperMetro works in load balancing mode, set the **Path Type** for the initiators of both the local and remote storage arrays to **Optimal Path**. Enable ALUA on both the host and storage arrays. If the host uses the round-robin multipathing policy, it delivers I/Os to both storage arrays in round-robin mode. ● When HyperMetro works in local preferred mode, set the **Path Type** for the initiator of the local storage array to **Optimal Path**, and that of the remote storage array to **Non-Optimal Path**. Enable ALUA on both the host and storage arrays. The host delivers I/Os to the local storage array preferentially.	Optimal Path

6.2.1 Storage System Configuration

If you use OS native multipathing software, you must select **Uses third-party multipath software** for the initiator, as shown in **Figure 6-5**.

Figure 6-5 Using OS native multipathing software



The **Switchover Mode** and **Path Type** depend on the actual services, as described in **Table 6-5**.

Table 6-5 Initiator parameter settings

Server OS	Storage Array Configuration						
	HyperMetro Working Mode	Storage	OS Setting	Third-Party Multipathing Software	Switchover Mode	Special Mode	Path Type
Red Hat 6.x	Load balancing	Local storage array	Linux	Enabled	Common ALUA		Optimal path
		Remote storage array	Linux	Enabled	Common ALUA		Optimal path

	Local preferred	Local storage array	Linux	Enabled	Common ALUA		Optimal path
		Remote storage array	Linux	Enabled	Common ALUA		Non-optimal path
Red Hat 7.x	Load balancing	Local storage array	Linux	Enabled	Special mode	Mode 0	Optimal path
		Remote storage array	Linux	Enabled	Special mode	Mode 0	Optimal path
	Local preferred	Local storage array	Linux	Enabled	Special mode	Mode 0	Optimal path
		Remote storage array	Linux	Enabled	Special mode	Mode 0	Non-optimal path

For details about the Red Hat versions, see the compatibility list:

http://support-open.huawei.com/ready/pages/user/compatibility/support-matrix.jsf

CAUTION

If a LUN has been mapped to the host, you must restart the host for the configuration to take effect after you modify the initiator parameters. If you configure the initiator for the first time, restart is not needed.

6.2.2 Host Configuration

Installing Multipathing Software

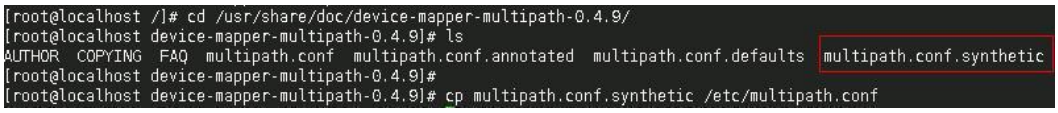
Generally, multipathing software packages in Red Hat are rpm packages starting with **device-mapper-multipath**. If you did not install the multipathing software when installing the operating system, you can obtain the software package from the system image and use the **rpm** command to install it.

Modifying the Configuration File

DM-Multipath's most important configuration file is **/etc/multipath.conf**.

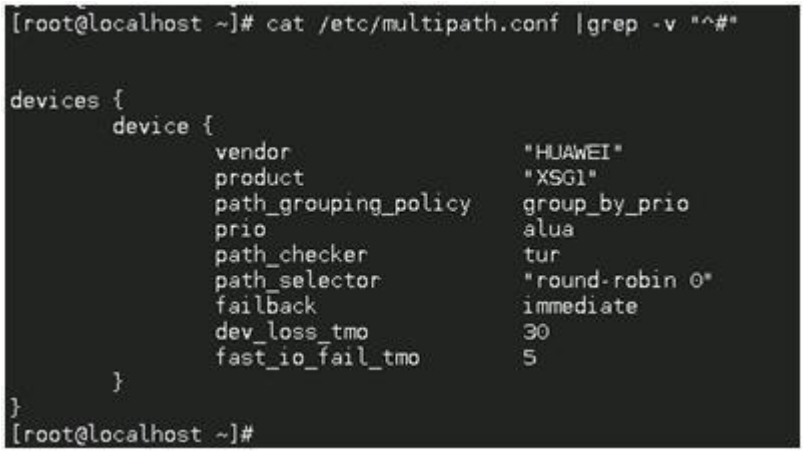
Some operating systems have this file by default. If your operating system does not have this file, you can copy the **multipath.conf.synthetic** file to the **/etc** directory to generate one, as shown in **Figure 6-6**.

Figure 6-6 Generating the multipathing configuration file



For Red Hat 6.x, add the contents in **Figure 6-7** to the **/etc/multipath.conf** file.

Figure 6-7 Configuration file for Red Hat 6.x



For Red Hat 7.x, add the contents in **Figure 6-8** to the **/etc/multipath.conf** file.

Figure 6-8 Configuration file for Red Hat 7.x

Starting the Multipathing Software

After configuring the configuration file, run the following command on the host to start the DM-Multipath process:

```
/etc/init.d/multipathd start
```

For Red Hat 7 and later versions, run the following command to start the DM-Multipath process:

```
systemctl start multipathd.service
```

Setting the Multipathing Software to Run at System Startup

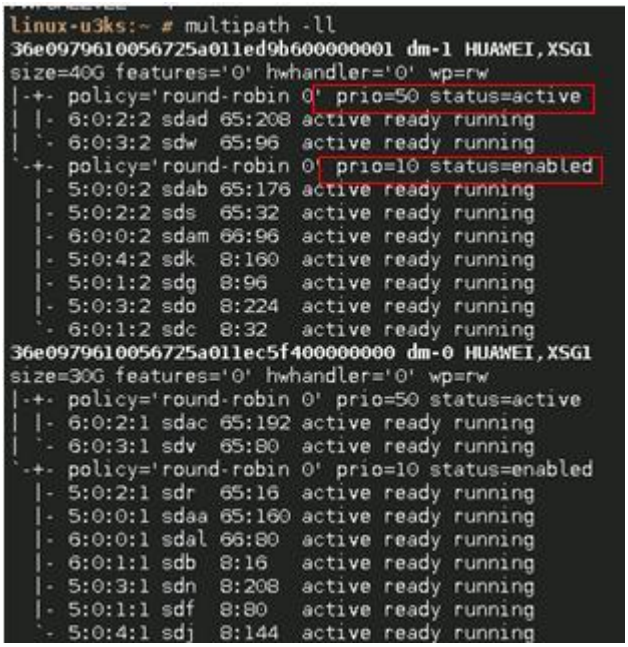
After enabling the software, you can run the following command to run the software at system startup:

```
chkconfig multipathd on
```

6.2.3 Verification

Run the **multipath -ll** command to verify that the configuration has taken effect.

Figure 6-9 Verifying the multipathing configuration



As shown in Figure 6-9, **status=active** corresponds to the AO path and **status=enabled** corresponds to the AN path. This indicates that the ALUA configuration has taken effect. Generally, the **prio** value of an AO path on a Linux system is **50**, and that of an AN path is **10**.

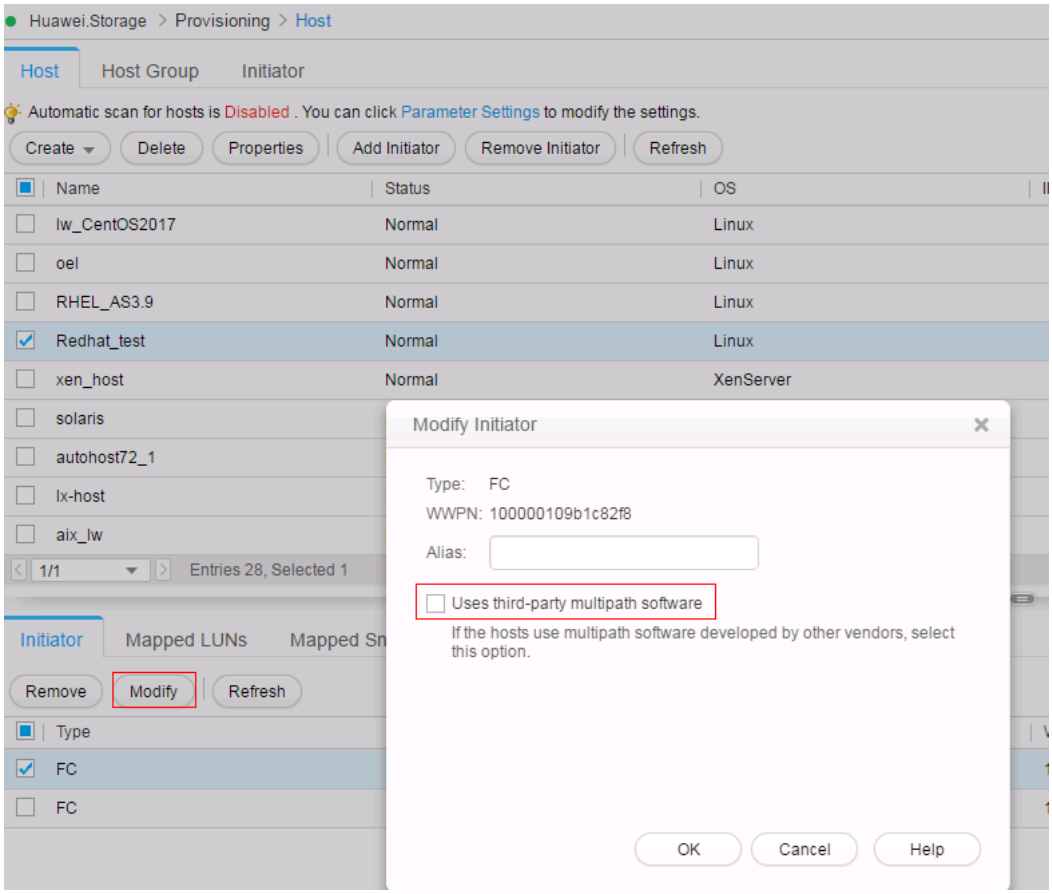
7 Configuring Multipathing in Non-HyperMetro Scenarios

7.1 UltraPath

7.1.1 Storage System Configuration

If you use UltraPath, retain the default initiator settings. Do not select **Uses third-party multipath software**.

Figure 7-1 Initiator setting when UltraPath is used



7.1.2 Host Configuration

Install and configure UltraPath by following instructions in the *OceanStor UltraPath for Linux User Guide*.

7.2 OS Native Multipathing Software

Huawei storage firmware's support for the OS native multipathing software is as follows:

- Old storage version (does not support multi-controller ALUA or ALUA HyperMetro)

  T V1, T V2, 18000 V1, V300R001, V300R002, V300R003C00, V300R003C10, V300R005, and Dorado V300R001C00
- New storage version (supports multi-controller ALUA and ALUA HyperMetro)

  V500R007C00, V300R003C20, V300R006C00, and Dorado V300R001C01

NOTE

For V300R003C20, only V300R003C20SPC200 and later versions are supported.

For V300R006C00, only V300R006C00SPC100 and later versions are supported.

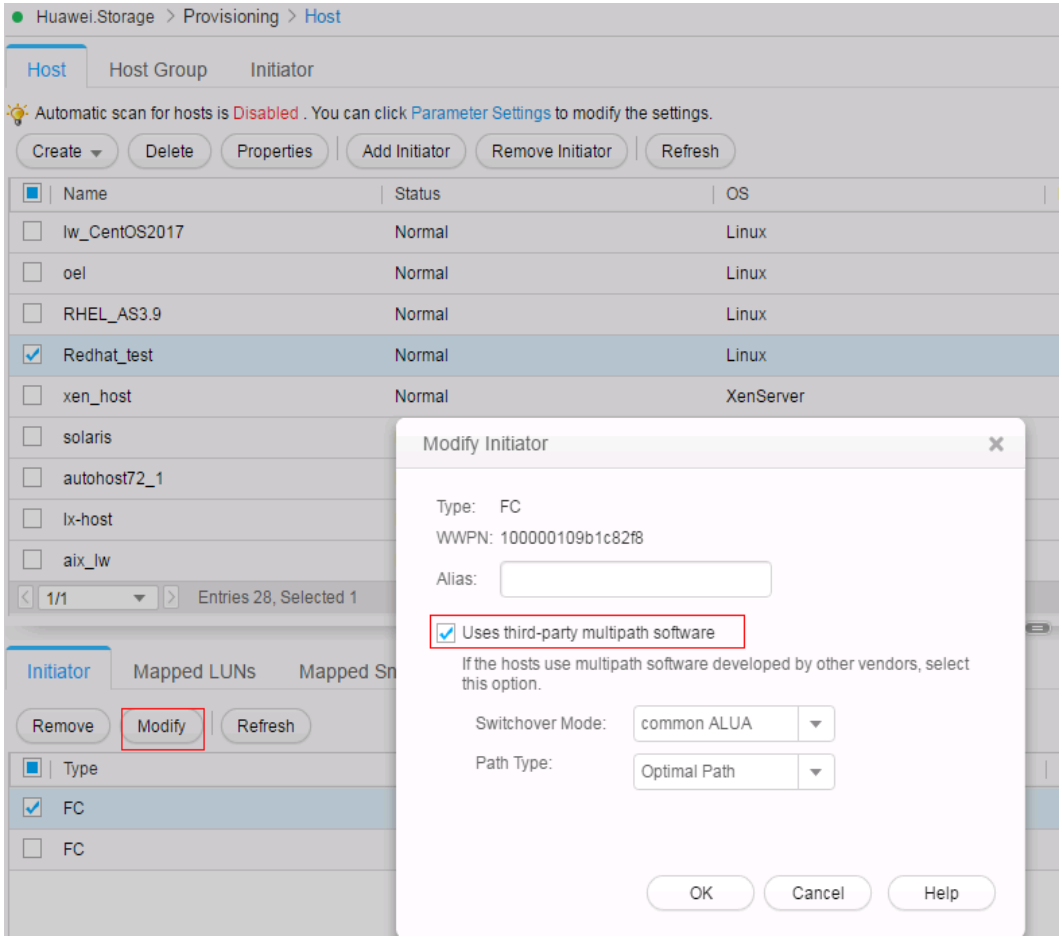
For Dorado V300R001C01, only V300R001C01SPC100 and later versions are supported.

7.2.1 New-Version Huawei Storage

Storage System Configuration

If you use OS native multipathing software, you must select **Uses third-party multipath software** for the initiator, as shown in **Figure 7-2**.

Figure 7-2 Using OS native multipathing software



The **Switchover Mode** and **Path Type** depend on the actual services, as described in **Table 7-1**.

Table 7-1 Initiator parameter settings

Server OS	Storage Array Configuration					
	Storage	OS Setting	Third-Party Multipathing Software	Switchover Mode	Special Mode	Path Type

Red Hat 6.x	Dual-controller, multi-controller	Linux	Enabled	Common ALUA		Optimal Path
Red Hat 7.x	Dual-controller, multi-controller	Linux	Enabled	Special mode	Mode 0	Optimal Path
Other Red Hat versions	Dual-controller, multi-controller	Linux	Enabled	ALUA not used		Optimal Path

For details about the Red Hat versions, see the compatibility list:

http://support-open.huawei.com/ready/pages/user/compatibility/support-matrix.jsf

CAUTION

If a LUN has been mapped to the host, you must restart the host for the configuration to take effect after you modify the initiator parameters. If you configure the initiator for the first time, restart is not needed.

7.2.1.1 Host Configuration

Installing Multipathing Software

Generally, multipathing software packages in Red Hat are rpm packages starting with **device-mapper-multipath**. If you did not install the multipathing software when installing the operating system, you can obtain the software package from the system image and use the **rpm** command to install it.

Modifying the Configuration File

DM-Multipath's most important configuration file is **/etc/multipath.conf**.

Some operating systems have this file by default. If your operating system does not have this file, you can copy the **multipath.conf.synthetic** file to the **/etc** directory to generate one, as shown in **Figure 7-3**.

Figure 7-3 Generating the multipathing configuration file

```
[root@localhost /]# cd /usr/share/doc/device-mapper-multipath-0.4.9/
[root@localhost device-mapper-multipath-0.4.9]# ls
AUTHOR  COPYING  FAQ  multipath.conf  multipath.conf.annotated  multipath.conf.defaults  multipath.conf.synthetic
[root@localhost device-mapper-multipath-0.4.9]#
[root@localhost device-mapper-multipath-0.4.9]# cp multipath.conf.synthetic /etc/multipath.conf
```

- If ALUA is enabled on the storage system:
  For Red Hat 6.x, add the contents in **Figure 7-4** to the **/etc/multipath.conf** file.

Figure 7-4 Configuration file for Red Hat 6.x

```
[root@localhost ~]# cat /etc/multipath.conf |grep -v "^#"


devices {
        device {
                vendor                  "HUAWEI"
                product                 "XSG1"
                path_grouping_policy    group_by_prio
                prio                    alua
                path_checker            tur
                path_selector           "round-robin 0"
                failback                immediate
                dev_loss_tmo            30
                fast_io_fail_tmo        5
        }
}
[root@localhost ~]#
```

For Red Hat 7.x, add the contents in **Figure 7-5** to the **/etc/multipath.conf** file.

Figure 7-5 Configuration file for Red Hat 7.x

```
devices {
        device {
                vendor                  "HUAWEI"
                product                 "XSG1"
                path_grouping_policy    group_by_prio
                prio                    alua
                path_checker            tur
                path_selector           "round-robin 0"
                failback                immediate
                fast_io_fail_tmo        15
                dev_loss_tmo            30
        }
}
```

- If ALUA is not enabled on the storage system:
  For versions earlier than Red Hat 6, add the contents in **Figure 7-6** to the **/etc/multipath.conf** file.

Figure 7-6 Configuration file for versions earlier than Red Hat 6

```
device {
                vendor                  "HUAWEI"
                product                 "XSG1"
                path_grouping_policy    multibus
                getuid_callout          "/sbin/scsi_id -g -u -s /block/%n"
                path_checker            tur
                path_selector           "round-robin 0"
                failback                immediate
        }
```

For Red Hat 6.x, add the contents in **Figure 7-7** to the **/etc/multipath.conf** file.

Figure 7-7 Configuration file for Red Hat 6.x

```
device {
                vendor                  "HUAWEI"
                product                 "XSG1"
                path_grouping_policy    multibus
                getuid_callout          "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
                path_checker            tur
                path_selector           "round-robin 0"
                failback                immediate
        }
```

For Red Hat 7.x, add the contents in **Figure 7-8** to the **/etc/multipath.conf** file.

Figure 7-8 Configuration file for Red Hat 7.x

```
device {
                vendor                  "HUAWEI"
                 product                 "XSG1"
                 path_grouping_policy     multibus
                 path_checker            tur
                 path_selector           "round-robin 0"
                 failback                immediate
        }
```

NOTE

Red Hat 6.9 and Red Hat 7.3 have added Huawei's multipathing configurations (when ALUA is disabled) to their kernel. When Huawei storage system is connected to Red Hat 6.9 or Red Hat 7.3 servers, you can use the default host settings and do not need to modify the **/etc/multipath.conf** configuration file if ALUA is disabled. However, if ALUA is enabled, you still need to modify the configuration file as required.

Starting the Multipathing Software

After configuring the configuration file, run the following command on the host to start the DM-Multipath process:

```
/etc/init.d/multipathd start
```

For Red Hat 7 and later versions, run the following command to start the DM-Multipath process:

```
systemctl start multipathd.service
```

Setting the Multipathing Software to Run at System Startup

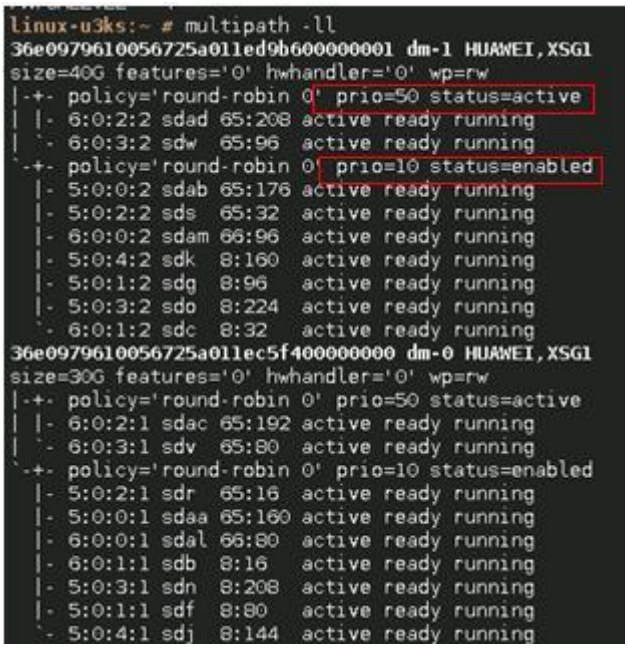
After enabling the software, you can run the following command to run the software at system startup:

```
chkconfig multipathd on
```

7.2.1.2 Verification

Run the **multipath -ll** command to verify that the configuration has taken effect.

Figure 7-9 Verifying the multipathing configuration



As shown in Figure 7-9, **status=active** corresponds to the AO path and **status=enabled** corresponds to the AN path. This indicates that the ALUA configuration has taken effect. Generally, the **prio** value of an AO path on a Linux system is **50**, and that of an AN path is **10**.

7.2.2 Old-Version Huawei Storage

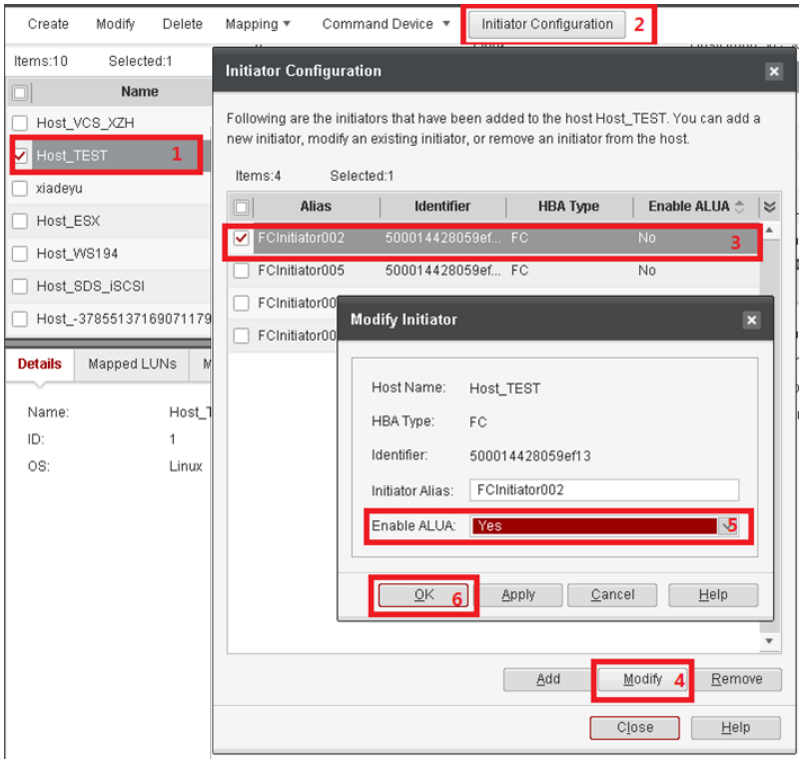
7.2.2.1 Storage System Configuration

For Huawei storage that does not support multi-controller ALUA or ALUA HyperMetro, it is advisable to retain the ALUA disabled state by default. To enable the ALUA function, do as follows:

- T series V100R005, Dorado2100, Dorado5100, and Dorado2100 G2

  Use the Huawei OceanStor ISM system to enable ALUA for all the host initiators, as shown in Figure 7-10.

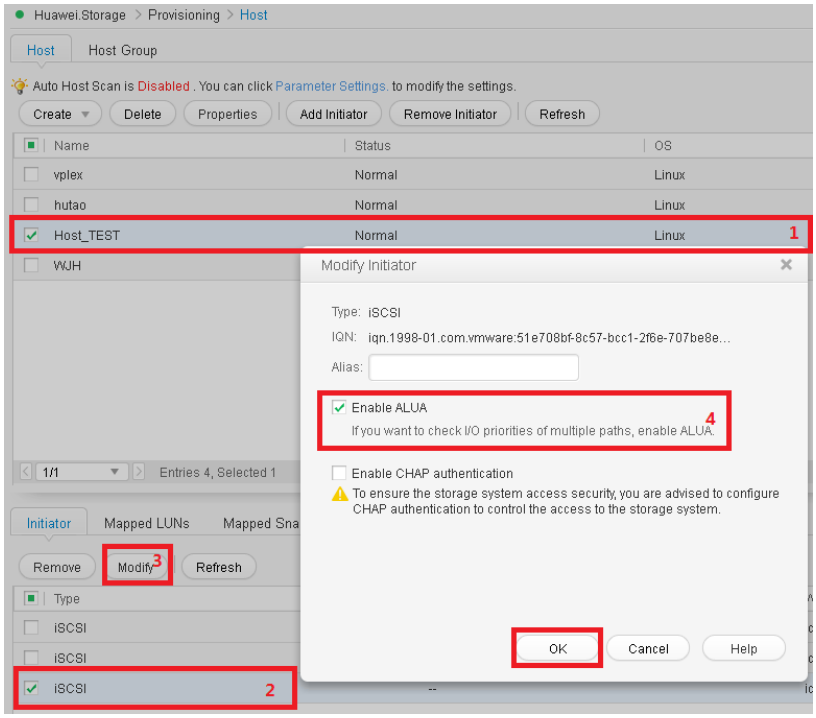
Figure 7-10 Enabling ALUA for T series V100R005/Dorado2100/Dorado5100/Dorado2100 G2



- T series V200R002, 18000 series, V3 series, and 18000 V3 series

  Use the Huawei OceanStor DeviceManager to enable ALUA for all the host initiators, as shown in **Figure 7-11**.

Figure 7-11 Enabling ALUA for T series V200R002/18000 series/V3 series/18000 V3 series



NOTE

Multi-controller ALUA is not supported. When there are more than two controllers, ALUA is disabled by default and the ALUA status cannot be changed.

7.2.2.2 Host Configuration

The host configurations are similar to those in section **7.2.1.1 Host Configuration**. Remember to change the vendor and product according to your site information; for example, a T series storage configuration file:

Figure 7-12 Multipathing configuration for Huawei T series storage

```
device {
                vendor                  "HUAWEI|HUASY"
                 product                 "S2600T"
                 path_grouping_policy     multibus
                 getuid_callout          "/sbin/scsi_id -g -u -s /block/%n"
                 path_checker            tur
                 path_selector           "round-robin 0"
                 failback                immediate
        }
```

8 FAQs

8.1 The Host Failed to Restart After iSCSI Connections Are Established

Symptom

The host failed to restart after iSCSI connections are set up between the host and the storage system.

Root Cause

The iSCSI service session is not terminated.

Solution

Stop the iSCSI service before the host restart.

8.2 LUN Information Failed to Be Updated After LUN Replacement

Symptom

After a LUN is replaced (the new LUN shares the same host ID as the original LUN), the information about the new LUN cannot be updated.

Root Cause

LUN information is not updated on the host.

Solution

Run the **echo 1 > /sys/block/sd*/device/rescan** command to update the LUN and then execute **/usr/bin/rescan-scsi-bus.sh**.

8.3 LUN Capacity Failed to Be Updated After Being Changed

Symptom

After the capacity of a LUN is changed, the new capacity of the LUN is not updated after system script **/usr/bin/rescan-scsi-bus.sh** is executed.

Root Cause

The LUN capacity is not updated on the host.

Solution

Run the **echo 1 > /sys/block/sd*/device/rescan** command to update the LUN and then execute **/usr/bin/rescan-scsi-bus.sh**.

8.4 Drive Letter Changes After Link Down for a Long Time

Symptom

After the link between the host and the storage system recovers from a long-time breakdown, the previously mounted drive letters are no longer available. The output of **lsscsi** shows that the drive letters change to new ones.

Root Cause

During the link recovery, the host deletes the original drive letters and generates new drive letters for the identified LUNs under the DEV directory. However, an error occurs and the original drive letters are not deleted. As a result, new drive letters are generated for the identified LUNs following the original drive letters. Therefore, drive letters shift backwards.

Solution

This problem can be resolved by mounting disks by UUID. Perform the following steps:

Step 1 Run the **fdisk -l** command to discover all disks.

Step 2 Partition and format the detected disks and create file systems for them.

Step 3 Run the following command to query the UUIDs of disks that you want to mount:

```
SMCDB-1:/# blkid
/dev/sdb1: UUID="894d76a6-b175-4eb1-89e5-3fd8d146eab7" SEC_TYPE="xfs" TYPE="ext2"
/dev/sdc1: UUID="ef285a94-2f34-4025-baa6-d35d8fbd0a86" SEC_TYPE="xfs" TYPE="ext2"
```

Step 4 Set files on disk partitions to be automatically mounted after the system restarts.

Modify the **/etc/fstab** file and add the following information to the end of this file (mount file systems **sdb1** and **sdc1** to directories **fs1** and **fs2** respectively):

```
UUID=894d76a6-b175-4eb1-89e5-3fd8d146eab7 /fs1 ext3 defaults 0 0
UUID=ef285a94-2f34-4025-baa6-d35d8fbd0a86 /fs2 ext3 defaults 0 0
```

Step 5 Run the following command to mount the file system:

```
SMCDB-1:/# mount -a
```

Step 6 Modify the system startup file to ensure the system mounts the file system after a restart as described in the **/etc/fstab** file.

```
SMCDB-1:/# vi /etc/rc.d/rc  ####Add /bin/ mount -a to the end of the file.
```

CAUTION

The **/bin/mount -a** command must be added in front of **exit0.**

After removing cables, unmount the directories and mount the disks.

----End

9 Acronyms and Abbreviations

C	
CHAP	Challenge Handshake Authentication Protocol
CLI	Command Line Interface
CDFS	CD-ROM File System
D	
DM-Multipath	Device Mapper-Multipath
E	
Ext2	The Second Extended File System
Ext3	Third extended file system
Ext4	The fourth extended file system
F	
FC	Fibre Channel
FHS	Filesystem Hierarchy Standard
G	
GE	Gigabit Ethernet
H	
HBA	Host Bus Adapter
I	
IP	Internet Protocol

ISM	Integrated Storage Manager
iSCSI	Internet Small Computer Systems Interface
L	
LACP	Link Aggregation Control Protocol
LE	Logical Extent
LUN	Logical Unit Number
LV	Logic Volume
LVM	Logical Volume Manager
LVS	Linux Virtual Server
LUCI	Linux Users of Central Illinois
M	
MB	MByte
N	
NFS	Network File System
R	
RAID	Redundant Array of Independent Disks
RHCS	Red Hat Cluster Suite
RHEL	Red Hat Enterprise Linux
S	
SAN	Storage Area Network
P	
PE	Physical Extent
PV	Physical Volume
V	
VLAN	Virtual Local Area Network
VG	Volume Group
W	
WWN	World Wide Name

10 Appendix A Volume Management

The most widely applied volume management software in Red Hat hosts is the Logical Volume Manager (LVM) built-in the operating systems.

This chapter details the LVM.

10.1 Overview

LVM can combine several disks (physical volumes) into a volume group and divide the volume group into logical volumes (LVM partitions).

LVM provides the following functions:

- Creating logical volumes across multiple disks
- Creating logical volumes on one disk
- Expanding and compressing logical volumes on demand

10.2 LVM Installation

By default, LVM is installed together with the host operating system. LVM requires no extra configuration.

10.3 Common Configuration Commands

Creating a Physical Volume

Step 1 Create primary and logical partitions.

Run the **fdisk –l** command to scan for the mapped LUNs. Suppose that the identified LUN is displayed as disk **sdb**. Run the **fdisk /dev/sdb** command to partition **sdb**.

```
[root@root ~]# fdisk /dev/sdb
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
```

```
disklabel
Building a new DOS disklabel. Changes will remain in memory only,
until you decide to write them. After that, of course, the previous
content won't be recoverable.


The number of cylinders for this disk is set to 13054.
There is nothing wrong with that, but this is larger than 1024,
and could in certain setups cause problems with:
1) software that runs at boot time (e.g., old versions of LILO)
2) booting and partitioning software from other OSs
 (e.g., DOS FDISK, OS/2 FDISK)
Warning: invalid flag 0x0000 of partition table 4 will be corrected by w(rite)

Command (m for help):
```

Type **n** to create new partitions and type **p** to create the primary partition. Specify the primary partition number to **1**. Keep the default value of **first cylinder** and specify a value to **last cylinder**.

```
Command (m for help): n
Command action
 e extended
 p primary partition (1-4)
p
Partition number (1-4): 1
First cylinder (1-13054, default 1):
Using default value 1
Last cylinder or +size or +sizeM or +sizeK (1-13054, default 13054): 200
```

Type **n** to create new partitions and type **e** to create expansion partitions. Then type **p** to view partitions.

```
Command (m for help): n
Command action
 e extended
 p primary partition (1-4)
e
Partition number (1-4): 4
First cylinder (201-13054, default 201):
Using default value 201
Last cylinder or +size or +sizeM or +sizeK (201-13054, default 13054): 1000

Command (m for help): p

Disk /dev/sdb: 107.3 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

 Device Boot      Start         End      Blocks Id  System
/dev/sdb1               1         200     1606468+  83  Linux
/dev/sdb4             201        1000     6426000    5  Extended
```

Type **n** to create new partitions and type **l** to create logical partitions. Type **p** to view partitions and type **w** to save partitions and exit from partition creation.

```
Command (m for help): n
Command action
 l logical (5 or over)
 p primary partition (1-4)
l
First cylinder (201-1000, default 201):
Using default value 201
Last cylinder or +size or +sizeM or +sizeK (201-1000, default 1000): 400

Command (m for help): p

Disk /dev/sdb: 107.3 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
```

```
Units = cylinders of 16065 * 512 = 8225280 bytes

 Device Boot      Start         End      Blocks Id  System
/dev/sdb1               1         200     1606468+  83  Linux
/dev/sdb4             201        1000     6426000    5  Extended
/dev/sdb5             201         400     1606468+  83  Linux
```

Step 2 Create LVM partitions.

Perform the following operation to convert partitions 5 and 6 of **sdb** to LVM partitions.

```
Command (m for help): p

Disk /dev/sdb: 107.3 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

 Device Boot      Start         End      Blocks Id  System
/dev/sdb1               1         200     1606468+  83  Linux
/dev/sdb4             201        1000     6426000    5  Extended
/dev/sdb5             201         400     1606468+  8e  Linux
/dev/sdb6             401         600     1606468+  83  Linux

Command (m for help): t
Partition number (1-6): 5
Hex code (type L to list codes): 8e
Changed system type of partition 6 to 8e (Linux LVM)

Command (m for help): t
Partition number (1-6): 6
Hex code (type L to list codes): 8e
Changed system type of partition 6 to 8e (Linux LVM)

Command (m for help): p

Disk /dev/sdb: 107.3 GB, 107374182400 bytes
255 heads, 63 sectors/track, 13054 cylinders
Units = cylinders of 16065 * 512 = 8225280 bytes

 Device Boot      Start         End      Blocks   Id  System
/dev/sdb1               1         200     1606468+  83  Linux
/dev/sdb4             201        1000     6426000    5  Extended
/dev/sdb5             201         400     1606468+  8e  Linux LVM
/dev/sdb6             401         600     1606468+  8e  Linux LVM
```

Step 3 Run the **pvcreate** command to create physical volumes.

```
[root@root ~]# pvcreate /dev/sdb5
  Physical volume "/dev/sdb5" successfully created
[root@root ~]# pvcreate /dev/sdb6
  Physical volume "/dev/sdb6" successfully created
```

Step 4 Run the **pvdisplay -v** command to verify the physical volume creation.

```
[root@root ~]# pvdisplay -v
    Scanning for physical volume names
    Wiping cache of LVM-capable devices
  --- Physical volume ---
  PV Name               /dev/sda2
  VG Name               VolGroup00
  PV Size               557.65 GB / not usable 21.17 MB
  Allocatable           yes (but full)
  PE Size (KByte)       32768
  Total PE              17844
  Free PE               0
  Allocated PE          17844
  PV UUID               KyucjQ-9zte-1Zyr-0sZ0-Xxzt-HVjZ-2vQp8B

  "/dev/sdb5" is a new physical volume of "1.53 GB"
  --- NEW Physical volume ---
```

```
  PV Name               /dev/sdb5
  VG Name
  PV Size               1.53 GB
  Allocatable           NO
  PE Size (KByte)       0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               g60zN0-3sYn-qPbd-7y0M-dGfZ-hVs7-763Ywo

  "/dev/sdb6" is a new physical volume of "1.53 GB"
  --- NEW Physical volume ---
  PV Name               /dev/sdb6
  VG Name
  PV Size               1.53 GB
  Allocatable           NO
  PE Size (KByte)       0
  Total PE              0
  Free PE               0
  Allocated PE          0
  PV UUID               5UhmY2-fS4p-gdCo-OOgZ-nOa9-AV3H-LkvrNc
```

----End

Changing the Size of a Physical Volume

Run the **pvresize** command to change the size of a physical volume. The command syntax is as follows:

pvresize –setphysicalvolumesize *capacity size* (unit: m or g) *device name*

In the following example, the size of a physical volume is changed from 1.53 GB to 300 MB.

```
[root@root ~]# pvscan
  PV /dev/sda2 VG VolGroup00      lvm2 [557.62 GB / 0    free]
  PV /dev/sdb5                      lvm2 [1.53 GB]
  PV /dev/sdb6                      lvm2 [1.53 GB]
  Total: 3 [560.69 GB] / in use: 1 [557.62 GB] / in no VG: 2 [3.06 GB]
 [root@root ~]# pvresize --setphysicalvolumesize 300 /dev/sdb5
  Physical volume "/dev/sdb5" changed
  1 physical volume(s) resized / 0 physical volume(s) not resized
[root@root ~]# pvscan
  PV /dev/sda2 VG VolGroup00      lvm2 [557.62 GB / 0    free]
  PV /dev/sdb5                      lvm2 [300.00 MB]
  PV /dev/sdb6                      lvm2 [1.53 GB]
  Total: 3 [559.45 GB] / in use: 1 [557.62 GB] / in no VG: 2 [1.83 GB]
```

Creating a Volume Group

Run the **vgcreate** command to create a volume group:

```
[root@root ~]# vgcreate vg0 /dev/sdb5 /dev/sdb6
  Volume group "vg0" successfully created
```

Expanding a Volume Group

Run the following command to expand a volume group:

```
vgextend vgname pvname
```

The following is an example:

```
[root@root ~]# vgdisplay -v /dev/vg0
    Using volume group(s) on command line
    Finding volume group "vg0"
  --- Volume group ---
```

```
  VG Name               vg0
  System ID
  Format                lvm2
  Metadata Areas        2
  Metadata Sequence No  1
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                2
  Act PV                2
  VG Size               1.82 GB
  PE Size               4.00 MB
  Total PE              466
  Alloc PE / Size       0 / 0
  Free  PE / Size       466 / 1.82 GB
  VG UUID               ARkbdL-9ID6-5HCy-DSQG-Aj5z-dQap-9VkM5X
  --- Physical volumes ---
  PV Name               /dev/sdb5
  PV UUID               g60zN0-3sYn-qPbd-7y0M-dGfZ-hVs7-763Ywo
  PV Status             allocatable
  Total PE / Free PE    74 / 74
  PV Name               /dev/sdb6
  PV UUID               5UhmY2-fS4p-gdCo-OOgZ-nOa9-AV3H-LkvrNc
  PV Status             allocatable
  Total PE / Free PE    392 / 392
[root@root ~]# vgextend /dev/vg0 /dev/sdb7
  Volume group "vg0" successfully extended
[root@root ~]# vgdisplay -v /dev/vg0
    Using volume group(s) on command line
    Finding volume group "vg0"
  --- Volume group ---
  VG Name               vg0
  System ID
  Format                lvm2
  Metadata Areas        3
  Metadata Sequence No  2
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                3
  Act PV                3
  VG Size               3.35 GB
  PE Size               4.00 MB
  Total PE              858
  Alloc PE / Size       0 / 0
  Free  PE / Size       858 / 3.35 GB
  VG UUID               ARkbdL-9ID6-5HCy-DSQG-Aj5z-dQap-9VkM5X
  --- Physical volumes ---
  PV Name               /dev/sdb5
  PV UUID               g60zN0-3sYn-qPbd-7y0M-dGfZ-hVs7-763Ywo
  PV Status             allocatable
  Total PE / Free PE    74 / 74
  PV Name               /dev/sdb6
  PV UUID               5UhmY2-fS4p-gdCo-OOgZ-nOa9-AV3H-LkvrNc
  PV Status             allocatable
  Total PE / Free PE    392 / 392
  PV Name               /dev/sdb7
  PV UUID               iF5Att-fVIj-9dOy-5055-rJlq-pOrS-aW8g2P
  PV Status             allocatable
  Total PE / Free PE    392 / 392
```

In this example, volume group **/dev/vg0** originally contains physical volume **/dev/sdb5** and **/dev/sdb6**. After the command is run, **/dev/sdb7** is added to this volume group.

Creating a Logical Volume

Step 1 Run the **lvcreate** command to create a logical volume. The following is an example:

```
[root@root ~]# lvcreate -L 10m -n lv0 vg0
  Rounding up size to full physical extent 12.00 MB
  Logical volume "lv0" created
```

The parameters in this output are describes as follows:

- -L: indicates the size of a logical volume, expressed in MB.
- -n: indicates the name of a logical volume.

Step 2 View and confirm that the information about the newly created logical volume is correct.

```
[root@root ~]# vgdisplay -v vg0
    Using volume group(s) on command line
    Finding volume group "vg0"
  --- Volume group ---
  VG Name               vg0
  System ID
  Format                lvm2
  Metadata Areas        3
  Metadata Sequence No  3
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                1
  Open LV               0
  Max PV                0
  Cur PV                3
  Act PV                3
  VG Size               3.35 GB
  PE Size               4.00 MB
  Total PE              858
  Alloc PE / Size       3 / 12.00 MB
  Free  PE / Size       855 / 3.34 GB
  VG UUID               ARkbdL-9ID6-5HCy-DSQG-Aj5z-dQap-9VkM5X

  --- Logical volume ---
  LV Name                /dev/vg0/lv0
  VG Name                vg0
  LV UUID                H6uskM-6clf-NVh2-KMiO-1Gk2-0iBz-nXOav2
  LV Write Access        read/write
  LV Status              available
  # open                 0
  LV Size                12.00 MB
  Current LE             3
  Segments               1
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:2

  --- Physical volumes ---
  PV Name               /dev/sdb5
  PV UUID               g60zN0-3sYn-qPbd-7y0M-dGfZ-hVs7-763Ywo
  PV Status             allocatable
  Total PE / Free PE    74 / 74

  PV Name               /dev/sdb6
  PV UUID               5UhmY2-fS4p-gdCo-OOgZ-nOa9-AV3H-LkvrNc
  PV Status             allocatable
  Total PE / Free PE    392 / 389

  PV Name               /dev/sdb7
  PV UUID               iF5Att-fVIj-9dOy-5055-rJlq-pOrS-aW8g2P
  PV Status             allocatable
  Total PE / Free PE    392 / 392
```

```
[root@root ~]# lvdisplay -v /dev/vg0/lv0
    Using logical volume(s) on command line
  --- Logical volume ---
  LV Name                /dev/vg0/lv0
  VG Name                vg0
  LV UUID                H6uskM-6clf-NVh2-KMiO-1Gk2-0iBz-nXOav2
  LV Write Access        read/write
  LV Status              available
  # open                 0
  LV Size                12.00 MB
  Current LE             3
  Segments               1
  Allocation             inherit
  Read ahead sectors     auto
  - currently set to     256
  Block device           253:2
```

----End

Creating a File System

Step 1 Run the **mkfs.xx** command to create a file system. The following is an example:

```
[root@root ~]# mkfs.ext3 /dev/vg0/rlv0
mke2fs 1.39 (29-May-2006)
Filesystem label=
OS type: Linux
Block size=1024 (log=0)
Fragment size=1024 (log=0)
3072 inodes, 12288 blocks
614 blocks (5.00%) reserved for the super user
First data block=1
Maximum filesystem blocks=12582912
2 block groups
8192 blocks per group, 8192 fragments per group
1536 inodes per group
Superblock backups stored on blocks:
    8193

Writing inode tables: done
Creating journal (1024 blocks): done
Writing superblocks and filesystem accounting information: done

This filesystem will be automatically checked every 20 mounts or
180 days, whichever comes first.  Use tune2fs -c or -i to override.
```

Step 2 Create a mount point and mount the logical volume.

```
[root@root ~]# mkdir /test/mnt1
[root@root ~]# mount /dev/vg0/lv0 /test/mnt1/
Display the mounting information.
[root@root ~]# df -l
Filesystem           1K-blocks      Used Available Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                     548527904 3105828 517108888 1% /
/dev/sda1               101086     15667     80200  17% /boot
tmpfs                  8137904         0 8137904 0% /dev/shm
/dev/mapper/vg0-lv0      11895      1138     10143  11% /test/mnt1
```

The output shows that the logical volume is mounted correctly and can be used for subsequent data read and write.

Step 3 (Optional) You can run the following command to unmount the logical volume:

```
[root@root ~]# umount /dev/vg0/lv0
[root@root ~]# df -l
Filesystem           1K-blocks      Used Available Use% Mounted on
/dev/mapper/VolGroup00-LogVol00
                     548527904 3105828 517108888 1% /
```

```
/dev/sda1               101086     15667     80200  17% /boot
tmpfs                  8137904         0 8137904 0% /dev/shm
```

----End

Expanding a Logical Volume

Run the **lvextend** command to expand a logical volume. The command syntax is as follows:

```
lvextend  -L +target capacity  logical volume path
The following is an example:
[root@root ~]# lvscan
  ACTIVE            '/dev/vg0/lv0' [12.00 MB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
[root@root ~]# pvscan
  PV /dev/sdb5 VG vg0          lvm2 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6 VG vg0          lvm2 [1.53 GB / 1.52 GB free]
  PV /dev/sdb7 VG vg0          lvm2 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
[root@root ~]# lvextend -L +100m /dev/vg0/lv0
  Extending logical volume lv0 to 112.00 MB
  Logical volume lv0 successfully resized
[root@root ~]# lvscan
  ACTIVE            '/dev/vg0/lv0' [112.00 MB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
[root@root ~]# pvscan
  PV /dev/sdb5 VG vg0          lvm2 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6 VG vg0          lvm2 [1.53 GB / 1.42 GB free]
  PV /dev/sdb7 VG vg0          lvm2 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
```

The output shows that the logical volume capacity is expanded.

Compressing a Logical Volume

Run the **lvreduce** command to compress a logical volume. The command syntax is as follows:

```
lvreduce  -L +target capacity  logical volume path
The following is an example:
[root@root ~]# lvscan
  ACTIVE            '/dev/vg0/lv0' [112.00 MB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
[root@root ~]# pvscan
  PV /dev/sdb5 VG vg0          lvm2 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6 VG vg0          lvm2 [1.53 GB / 1.42 GB free]
  PV /dev/sdb7 VG vg0          lvm2 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
[root@root ~]# lvreduce -L -100m /dev/vg0/lv0
  WARNING: Reducing active logical volume to 12.00 MB
  THIS MAY DESTROY YOUR DATA (filesystem etc.)
Do you really want to reduce lv0? [y/n]: y
  Reducing logical volume lv0 to 12.00 MB
  Logical volume lv0 successfully resized
[root@root ~]# lvscan
  ACTIVE            '/dev/vg0/lv0' [12.00 MB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
[root@root ~]# pvscan
  PV /dev/sdb5 VG vg0          lvm2 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6 VG vg0          lvm2 [1.53 GB / 1.52 GB free]
```

```
  PV /dev/sdb7 VG vg0          lvm2 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
```

The output shows that the logical volume capacity is compressed.

Activating a Volume Group

Run the following command to activate a volume group:

```
vgchange -a  y volume group name
```

The following is an example:

```
[root@root ~]# vgchange -a y /dev/vg0
  1 logical volume(s) in volume group "vg0" now active
```

Deactivating a Volume Group

Run the following command to deactivate a volume group:

```
vgchange –a n y volume group name
```

The following is an example:

```
[root@root ~]# vgchange -a n /dev/vg0
  0 logical volume(s) in volume group "vg0" now active
```

Exporting a Volume Group

A volume group needs to be imported or exported in clusters, data backup, or recovery.

Run the following command to export a volume group:

```
vgexport volume group name
```

The following is an example:

```
[root@root ~]# vgexport vg0
  Volume group "vg0" successfully exported
[root@root ~]# pvscan
  PV /dev/sdb5    is in exported VG vg0 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6    is in exported VG vg0 [1.53 GB / 1.52 GB free]
  PV /dev/sdb7    is in exported VG vg0 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
```

Importing a Volume Group

Run the following command to import a volume group:

```
vgimport volume group name
```

The following is an example (importing a volume group on a local computer):

```
[root@root ~]# vgimport vg0
  Volume group "vg0" successfully imported
[root@root ~]# pvscan
  PV /dev/sdb5 VG vg0          lvm2 [296.00 MB / 296.00 MB free]
  PV /dev/sdb6 VG vg0          lvm2 [1.53 GB / 1.52 GB free]
  PV /dev/sdb7 VG vg0          lvm2 [1.53 GB / 1.53 GB free]
  PV /dev/sda2 VG VolGroup00 lvm2 [557.62 GB / 0    free]
  Total: 4 [560.98 GB] / in use: 4 [560.98 GB] / in no VG: 0 [0 ]
```

Deleting a Logical Volume

Run the following command to delete a logical volume:

```
lvremove lvname
```

The following is an example:

```
[root@root ~]# lvscan
  inactive          '/dev/vg0/lv0' [12.00 MB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
[root@root ~]# lvremove /dev/vg0/lv0
  Logical volume "lv0" successfully removed
[root@root ~]# lvscan
  ACTIVE            '/dev/VolGroup00/LogVol00' [540.03 GB] inherit
  ACTIVE            '/dev/VolGroup00/LogVol01' [17.59 GB] inherit
```

Deleting a Volume Group

Run the following command to delete a volume group:

```
vgremove vgname
```

Perform the following steps:

Step 1 Ensure that all logical volumes are deleted from the volume group.

```
[root@root ~]# vgdisplay -v /dev/vg0
    Using volume group(s) on command line
    Finding volume group "vg0"
  --- Volume group ---
  VG Name               vg0
  System ID
  Format                lvm2
  Metadata Areas        3
  Metadata Sequence No  8
  VG Access             read/write
  VG Status             resizable
  MAX LV                0
  Cur LV                0
  Open LV               0
  Max PV                0
  Cur PV                3
  Act PV                3
  VG Size               3.35 GB
  PE Size               4.00 MB
  Total PE              858
  Alloc PE / Size       0 / 0
  Free  PE / Size       858 / 3.35 GB
  VG UUID               ARkbdL-9ID6-5HCy-DSQG-Aj5z-dQap-9VkM5X

  --- Physical volumes ---
  PV Name               /dev/sdb5
  PV UUID               g60zN0-3sYn-qPbd-7y0M-dGfZ-hVs7-763Ywo
  PV Status             allocatable
  Total PE / Free PE    74 / 74

  PV Name               /dev/sdb6
  PV UUID               5UhmY2-fS4p-gdCo-OOgZ-nOa9-AV3H-LkvrNc
  PV Status             allocatable
  Total PE / Free PE    392 / 392

  PV Name               /dev/sdb7
  PV UUID               iF5Att-fVIj-9dOy-5055-rJlq-pOrS-aW8g2P
  PV Status             allocatable
  Total PE / Free PE    392 / 392
```

Step 2 Delete the volume group.

```
[root@root ~]# vgremove /dev/vg0
  Volume group "vg0" successfully removed
[root@root ~]# vgdisplay -v /dev/vg0
    Using volume group(s) on command line
    Finding volume group "vg0"
    Wiping cache of LVM-capable devices
  Volume group "vg0" not found
```

----End

Deleting a Physical Volume

Run the following command to delete a physical volume:

```
Pvremove raw device name
```

The following is an example:

```
[root@root ~]# pvremove /dev/sdb5
  Labels on physical volume "/dev/sdb5" successfully wiped
[root@root ~]# pvremove /dev/sdb6
  Labels on physical volume "/dev/sdb6" successfully wiped
[root@root ~]# pvremove /dev/sdb7
  Labels on physical volume "/dev/sdb7" successfully wiped
```

11 Appendix B High Availability Technology

As services grow, key applications must be available 24/7 and systems must have the fault tolerance capability. However, the systems with fault tolerance capability are costly. To lower the system costs, economical applications that provide the fault tolerance capacity are required.

A high availability (HA) solution ensures the availability of applications and data in an event of any system component fault. This solution aims at eliminating single points of failure and minimizing the impact of expected or unexpected system downtimes.

The most widely applied cluster management software in Red Hat hosts is Red Hat Cluster Suite (RHCS).

This chapter details the RHCS software.

11.1 Overview

RHCS is a software suit that can be flexibly deployed to meet requirements of high availability, load balancing, scalability, file sharing, and cost saving.

With the high availability deployment that makes applications to run all the time, RHCS enables enterprises to have the capability of Linux deployment scalability. RHCS also provides an always available failover solution to the open source applications such as the network file system (NFS), Samba, and Apache. For other applications, you can use the templates provided by RHCS to customize failover scripts. Additionally, RHCS provides customized deployment services completed by professionals from Red Hat.

RHCS has the following technical advantages:

- Up to 128 nodes (Red Hat 3 and Red Hat 4 support 16 nodes each)

- High availability of multiple applications
- NFS/CIFS failover: high availability files in both Windows and Unix
- Fully shared storage subsystems: All cluster members can access the same storage subsystem.
- Data integrity: the state-of-art I/O shied technology, such as programmable embedded/ external power switches
- Service failover: timely detection of nonfunctional hardware, failover upon faults for automatic recovery, and application monitoring that enables automatic application restart upon faults

11.2 Working Principle

RHCS is a collection of clustering tools. Its major components are as follows:

- Cluster architecture manager

  This is a basic component of RHCS. This component provides basic clustering functions, including distributed cluster management, membership management, lock management, configuration file management, and fence devices.
- High availability service manager

  This component provides node service monitoring and failover. It automatically switches services from a faulty node to a normal one.
- Cluster configuration management tool

  RHCS of the latest version employs Linux Users of Central Illinois (LUCI) to configure and manage clusters. LUCI is a web-based configuration method that can construct a powerful cluster system at ease.
- LVS

  Linux Virtual Server (LVS) is an open-source load balancing software that evenly distributes client requests to service nodes based on specified load balancing policies and algorithms, achieving dynamic and intelligent load balancing.

11.3 Installation and Configuration

For details, visit:

https://access.redhat.com/knowledge/docs/en-US/Red_Hat_Enterprise_Linux/5/html/Cluster_Suite_Overview/index.html

Huawei also provides RHCS configuration guides. You can obtain the guides from Huawei Customer Service Center.

11.4 Cluster Maintenance

Starting a Cluster

Run the following command to start a cluster:

```
# service cman start
# service clvmd start
# service gfs start
```

```
# service rgmanager start
```

Stopping a Cluster

Run the following command to stop a cluster:

```
# service rgmanager stop
# service gfs stop
# service clvmd stop
# service cman stop
```

Checking Cluster Status

Run the following command to check the cluster status:

```
# clustat -l
```

Packet Service Switchover

Run the following command to switch a service from one node to another. This command switches the **trssvc** service to node **webdb2**.

```
# clusvcadm –r trssvc –m webdb2
```