Lenovo

ThinkSystem DE Series
Linux express configuration

ThinkSystem DE Series 11.70.3

Second edition (June 2022)

Contents

Chapter 1. Linux express configuration overview

The Linux express method for installing your storage array and accessing ThinkSystem System Manager is appropriate for setting up a standalone Linux host to a DE Series storage system. It is designed to get the storage system up and running as quickly as possible with minimal decision points.

1.1. Procedure overview

The Linux express method includes the following steps.

1. Set up one of the following communication environments:
    - Fibre Channel (FC)
    - iSCSI
    - SAS
    - NVMe over RoCE
    - NVMe over Fibre Channel
2. Create logical volumes on the storage array.
3. Make the volumes available to the data host.

1.2. Find more information

- Online help — Describes how to use ThinkSystem System Manager to complete configuration and storage management tasks. It is available within the product.
- Lenovo ThinkSystem Storage Documentation Center DE Series (a database of articles) — Provides troubleshooting information, FAQs, and instructions for a wide range of Lenovo products and technologies.
- Lenovo Storage Interoperation Center (LSIC) — Enables you to search for configurations of Lenovo products and components that meet the standards and requirements specified by Lenovo.

Chapter 2. Assumptions

The Linux express method is based on the following assumptions:

Component	Assumptions
Hardware	• You have used the Installation and Setup Instructions included with the controller shelves to install the hardware. • You have connected cables between the optional drive shelves and the controllers. • You have applied power to the storage system. • You have installed all other hardware (for example, management station, switches) and made the necessary connections. • If you are using NVMe over Fabrics, each DE6000F or DE6000H controller contains at least 64 GB of RAM.
Host	• You have made a connection between the storage system and the data host. • You have installed the host operating system. • You are not using Linux as a virtualized guest. • You are not configuring the data (I/O attached) host to boot from SAN. • You have installed any OS updates as listed under Lenovo Storage Interoperation Center (LSIC).
Storage management station	• You are using a 1 Gbps or faster management network. • You are using a separate station for management rather than the data (I/O attached) host. • You are using out-of-band management, in which a storage management station sends commands to the storage system through the Ethernet connections to the controller. • You have attached the management station to the same subnet as the storage management ports.
IP addressing	• You have installed and configured a DHCP server. • You have **not** yet made an Ethernet connection between the management station and the storage system.

Component	Assumptions
Storage provisioning	• You will not use shared volumes. • You will create pools rather than volume groups.
Protocol: FC	• You have made all host-side FC connections and activated switch zoning. • You are using Lenovo-supported FC HBAs and switches. • You are using FC HBA driver and firmware versions as listed in Lenovo Storage Interoperation Center (LSIC).
Protocol: iSCSI	• You are using Ethernet switches capable of transporting iSCSI traffic. • You have configured the Ethernet switches according to the vendor's recommendation for iSCSI.
Protocol: SAS	• You are using Lenovo-supported SAS HBAs. • You are using SAS HBA driver and firmware versions as listed in Lenovo Storage Interoperation Center (LSIC).
Protocol: NVMe over RoCE	• You have received the 100G host interface cards in a DE6000F, or DE6000H storage system pre-configured with the NVMe over RoCE protocol or the controllers were ordered with standard IB ports and need to be converted to NVMe-oF ports. • You are using NVMe/RoCE driver and firmware versions as listed in Lenovo Storage Interoperation Center (LSIC).
Protocol: NVMe over Fibre Channel	• You have received the 32G host interface cards in a DE6000F, or DE6000H storage system pre-configured with the NVMe over Fibre Channel protocol or the controllers were ordered with standard FC ports and need to be converted to NVMe-oF ports. • You are using NVMe/FC driver and firmware versions as listed in Lenovo Storage Interoperation Center (LSIC).

Note: These express method instructions include examples for SUSE Linux Enterprise Server (SLES) and for Red Hat Enterprise Linux (RHEL). Examples for RHEL are specific to RHEL7.

Chapter 3. Fibre Channel Express Setup

3.1. Verify the Linux configuration is supported

To ensure reliable operation, you create an implementation plan and then use the Lenovo Storage Interoperation Center (LSIC) to verify that the entire configuration is supported.

Steps

1. Go to Lenovo Storage Interoperation Center (LSIC) for interop support configuration.
2. Choose your storage model, firmware, protocol, HBA, and operating system and click here for guidance on how to use LSIC to search the products support configuration.

3.2. Configure IP addresses using DHCP

To configure communications between the management station and the storage array, use Dynamic Host Configuration Protocol (DHCP) to provide IP addresses.

What you'll need

A DHCP server installed and configured on the same subnet as the storage management ports.

About this task

Each storage array has two storage management ports. Each management port will be assigned an IP address.

The following instructions refer to a storage array with two controllers.

Steps

1. If you have not already done so, connect an Ethernet cable to the management station and to management port 1 on each controller (A and B).

   The DHCP server assigns an IP address to port 1 of each controller.

   > Do not use management port 2 on either controller. Port 2 is reserved for use by Lenovo technical personnel.

   > If you disconnect and reconnect the Ethernet cable, or if the storage array is power-cycled, DHCP assigns IP addresses again. This process occurs until static IP addresses are configured. It is recommended that you avoid disconnecting the cable or power-cycling the array.

   If the storage array cannot get DHCP-assigned IP addresses within 30 seconds, the following default IP addresses are set:

   - Controller A, port 1: 169.254.128.101
   - Controller B, port 1: 169.254.128.102
   - Subnet mask: 255.255.0.0

2. Locate the MAC address label on the back of each controller, and then provide your network administrator with the MAC address for port 1 of each controller.

   Your network administrator needs the MAC addresses to determine the IP address for each controller. You will need the IP addresses to connect to your storage system through your browser.

3.3. Install and configure Lenovo ThinkSystem Host Utilities for Linux

The Lenovo ThinkSystem Host Utilities for Linux for DE Series tools help you manage Lenovo storage.

> Alternatively, you can use the `SMdevices` utility to perform the same functions as the Host Utilities for Linux. The `SMdevices` utility is included as part of the `SMutils` package. The `SMutils` package is a collection of utilities to verify what the host sees from the storage array. It is included as part of the Lenovo ThinkSystem Host Utilities.

About this task

This section describes how to install ThinkSystem Host Utilities on both the Windows and Linux OS platforms, because both Windows and Linux are common management station platforms when

Linux is used for the data host.

Steps

1. Download the Lenovo ThinkSystem Host Utilities from Lenovo Data Center Support by searching "Lenovo ThinkSystem Host Utilities".
2. Run the Lenovo ThinkSystem Host Utilities.

Windows	Linux
Double-click the `lnvgy_utl_hostutilities_*.exe` installation package to start the installation.	a. Go to the directory where the `lnvgy_utl_hostutilities_*.bin` installation package is located. b. If the temp mount point does not have execute permissions, set the `IATEMPDIR` variable. Example: `IATEMPDIR=/root ./Lnvgy_utl_hostutilities-11.70.0A00.0002.bin` c. Run the `chmod +x Lnvgy_utl_hostutilities_*.bin` command to grant execute permission to the file. d. Run the `./Lnvgy_utl_hostutilities_*.bin` command to start the installer.

3. Use the installation wizard to install the software on the management station.

3.4. Access ThinkSystem System Manager and use the Setup wizard

To configure your storage array, you can use the Setup wizard in ThinkSystem System Manager.

ThinkSystem System Manager is a web-based interface embedded on each controller. To access the user interface, you point a browser to the controller's IP address. A setup wizard helps you get started with system configuration.

What you'll need

- Out-of-band management.
- A management station for accessing ThinkSystem System Manager that includes one of the following browsers:

Browser	Minimum version
Google Chrome	79
Microsoft Internet Explorer	11
Microsoft Edge	79
Mozilla Firefox	70
Safari	12

About this task

The wizard automatically relaunches when you open System Manager or refresh your browser and *at least one* of the following conditions is met:

- No pools and volume groups are detected.
- No workloads are detected.
- No notifications are configured.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been configured, the Set Administrator Password prompt appears. Role-based access management configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Use the Setup wizard to perform the following tasks:
   - **Verify hardware (controllers and drives)** — Verify the number of controllers and drives in the storage array. Assign a name to the array.
   - **Verify hosts and operating systems** — Verify the host and operating system types that the

storage array can access.

- **Accept pools** — Accept the recommended pool configuration for the express installation method. A pool is a logical group of drives.
- **Configure alerts** — Allow System Manager to receive automatic notifications when a problem occurs with the storage array.
- **Enable AutoSupport** — Automatically monitor the health of your storage array and have dispatches sent to technical support.

4. If you have not already created a volume, create one by going to **Storage › Volumes › Create › Volume**.

   For more information, see the online help for ThinkSystem System Manager.

3.5. Configure the multipath software

To provide a redundant path to the storage array, you can configure multipath software.

What you'll need

You must install the required packages on your system.

- For Red Hat (RHEL) hosts, verify the packages are installed by running `rpm -q device-mapper-multipath`.
- For SLES hosts, verify the packages are installed by running `rpm -q multipath-tools`.

If you have not already installed the operating system, use the media supplied by your operating system vendor.

About this task

Multipath software provides a redundant path to the storage array in case one of the physical paths is disrupted. The multipath software presents the operating system with a single virtual device that represents the active physical paths to the storage. The multipath software also manages the failover process that updates the virtual device.

You use the device mapper multipath (DM-MP) tool for Linux installations. By default, DM-MP is disabled in RHEL and SLES. Complete the following steps to enable DM-MP components on the host.

Steps

1. If a multipath.conf file is not already created, run the `# touch /etc/multipath.conf` command.
2. Use the default multipath settings by leaving the multipath.conf file blank.
3. Start the multipath service.

```
# systemctl start multipathd
```

4. Save your kernel version by running the `uname -r` command.

```
# uname -r
3.10.0-327.el7.x86_64
```

You will use this information when you assign volumes to the host.

5. Do one of the following to enable the multipathd daemon on boot.

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`systemctl enable multipathd`
SLES 12.x and 15.x systems:	`systemctl enable multipathd`

6. Rebuild the `initramfs` image or the `initrd` image under /boot directory:

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`dracut --force --add multipath`
SLES 12.x and 15.x systems:	`dracut --force --add multipath`

7. Make sure that the newly created /boot/initrams-* image or /boot/initrd-* image is selected in the boot configuration file.

   For example, for grub it is `/boot/grub/menu.lst` and for grub2 it is `/boot/grub2/menu.cfg`.

8. Use the "Create host manually" procedure in the online help to check whether the hosts are defined. Verify that each host type is either **Linux DM-MP (Kernel 3.10 or later)** if you enable the Automatic Load Balancing feature, or **Linux DM-MP (Kernel 3.9 or earlier)** if you disable the Automatic Load Balancing feature. If necessary, change the selected host type to the appropriate setting.
9. Reboot the host.

3.6. Set up the multipath.conf file

The multipath.conf file is the configuration file for the multipath daemon, multipathd. The multipath.conf file overrides the built-in configuration table for multipathd. Any line in the file with a first non-white-space character of # is considered a comment line. Empty lines are ignored.

For ThinkSystem SAN OS 11.50 and newer, Lenovo recommends using the default settings as provided.

The multipath.conf files are generated as:

- For SLES, run the command:

  ```
  multipath -T > /etc/multipath.conf
  ```

- For RHEL, run the command:

  ```
  mpathconf --enable --with_multipathd y
  ```

Alternatively, edit `/etc/multipath.conf` content manually and add the section below:

```
devices {
      device {
        vendor "LENOVO"
        product "DE_Series"
        product_blacklist "Universal Xport"
        path_grouping_policy "group_by_prio"
        path_checker "rdac"
        features "2 pg_init_retries 50"
        hardware_handler "1 rdac"
        prio "rdac"
        failback immediate
        rr_weight "uniform"
        no_path_retry 30
        retain_attached_hw_handler yes
        detect_prio yes
      }
}
```

3.7. Configure the FC switches

Configuring (zoning) the Fibre Channel (FC) switches enables the hosts to connect to the storage array and limits the number of paths. You zone the switches using the management interface for the switches.

What you'll need

- Administrator credentials for the switches.
- The WWPN of each host initiator port and of each controller target port connected to the switch. (Use your HBA utility for discovery.)

About this task

Each initiator port must be in a separate zone with all of its corresponding target ports. For details about zoning your switches, see the switch vendor's documentation.

Steps

1. Log in to the FC switch administration program, and then select the zoning configuration option.
2. Create a new zone that includes the first host initiator port and that also includes all of the target ports that connect to the same FC switch as the initiator.
3. Create additional zones for each FC host initiator port in the switch.
4. Save the zones, and then activate the new zoning configuration.

3.8. Determine host WWPNs and make the recommended settings

You install an FC HBA utility so you can view the worldwide port name (WWPN) of each host port.

Additionally, you can use the HBA utility to change any settings recommended in the Notes column of the Lenovo Storage Interoperation Center (LSIC) for the supported configuration.

About this task

Guidelines for HBA utilities:

- Most HBA vendors offer an HBA utility. You will need the correct version of HBA for your host operating system and CPU. Examples of FC HBA utilities include:
  - Emulex OneCommand Manager for Emulex HBAs
  - QLogic QConverge Console for QLogic HBAs
- Host I/O ports might automatically register if the host context agent is installed.

Steps

1. Download the appropriate utility from your HBA vendor's web site.
2. Install the utility.
3. Select the appropriate settings in the HBA utility.

   Appropriate settings for your configuration are listed in Lenovo Storage Interoperation Center (LSIC).

3.9. Create partitions and filesystems

Because a new LUN has no partition or file system when the Linux host first discovers it, you must format the LUN before it can be used. Optionally, you can create a file system on the LUN.

What you'll need

- A LUN that is discovered by the host.
- A list of available disks. (To see available disks, run the `ls` command in the /dev/mapper folder.)

About this task

You can initialize the disk as a basic disk with a GUID partition table (GPT) or Master boot record (MBR).

Format the LUN with a file system such as ext4. Some applications do not require this step.

Steps

1. Retrieve the SCSI ID of the mapped disk by issuing the `multipath -ll` command.

   The SCSI ID is a 33-character string of hexadecimal digits, beginning with the number 3. If user-friendly names are enabled, Device Mapper reports disks as mpath instead of by a SCSI ID.

   ```
   # multipath -ll
   mpathd(360080e5000321bb8000092b1535f887a) dm-2 Lenovo DE_Series
   size=1.0T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 alua' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- 16:0:4:4 sde     69:144 active ready running
   | `- 15:0:5:4 sdf     65:176 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- 16:0:5:4 sdg     70:80  active ready running
     `- 15:0:1:4 sdh     66:0   active ready running
   ```

2. Create a new partition according to the method appropriate for your Linux OS release.

   Typically, characters identifying the partition of a disk are appended to the SCSI ID (the number 1 or p3 for instance).

   ```
   # parted -a optimal -s -- /dev/mapper/360080e5000321bb8000092b1535f887a mklabel
   gpt mkpart primary ext4 0% 100%
   ```

3. Create a file system on the partition.

   The method for creating a file system varies depending on the file system chosen.

```
# mkfs.ext4 /dev/mapper/360080e5000321bb8000092b1535f887a1
```

4. Create a folder to mount the new partition.

```
# mkdir /mnt/ext4
```

5. Mount the partition.

```
# mount /dev/mapper/360080e5000321bb8000092b1535f887a1 /mnt/ext4
```

3.10. Verify storage access on the host

Before using the volume, verify that the host can write data to the volume and read it back.

What you'll need

An initialized volume that is formatted with a file system.

Steps

1. On the host, copy one or more files to the mount point of the disk.
2. Copy the files back to a different folder on the original disk.
3. Run the `diff` command to compare the copied files to the originals.

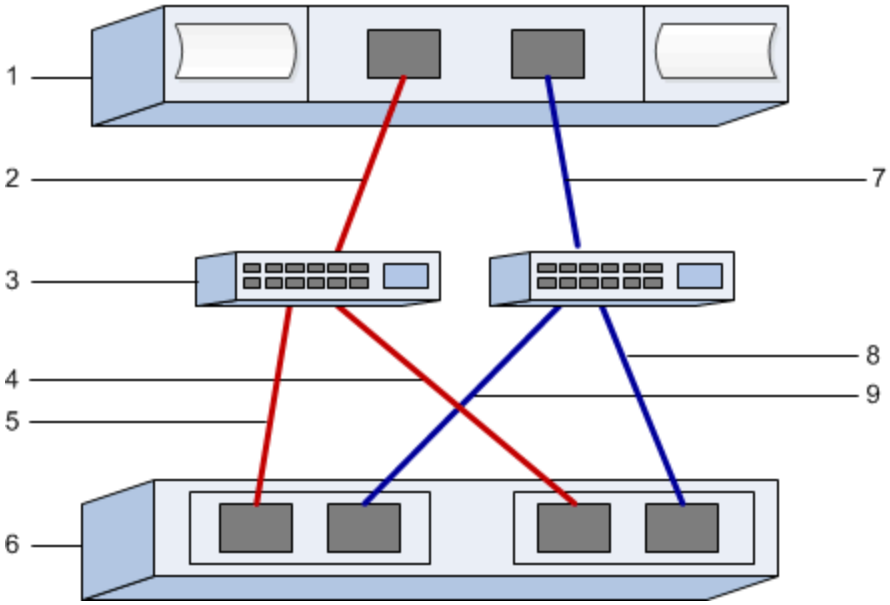
After you finish

Remove the file and folder that you copied.

3.11. Record your FC configuration

You can generate and print a PDF of this page, and then use the following worksheet to record FC storage configuration information. You need this information to perform provisioning tasks.

The illustration shows a host connected to a DE Series storage array in two zones. One zone is indicated by the blue line; the other zone is indicated by the red line. Any single port has two paths to the storage (one to each controller).



3.11.1. Host identifiers

Callout No.	Host (initiator) port connections	WWPN
1	Host	*not applicable*
2	Host port 0 to FC switch zone 0	
7	Host port 1 to FC switch zone 1	

3.11.2. Target identifiers

Callout No.	Array controller (target) port connections	WWPN
3	Switch	*not applicable*
6	Array controller (target)	*not applicable*
5	Controller A, port 1 to FC switch 1	
9	Controller A, port 2 to FC switch 2	
4	Controller B, port 1 to FC switch 1	

Callout No.	Array controller (target) port connections	WWPN
8	Controller B, port 2 to FC switch 2	

3.11.3. Mapping host

Mapping host name	
Host OS type	

Chapter 4. SAS Setup

4.1. Verify the Linux configuration is supported

To ensure reliable operation, you create an implementation plan and then use the Lenovo Storage Interoperation Center (LSIC) to verify that the entire configuration is supported.

Steps

1. Go to Lenovo Storage Interoperation Center (LSIC) for interop support configuration.
2. Choose your storage model, firmware, protocol, HBA, and operating system and click here for guidance on how to use LSIC to search the products support configuration.

4.2. Configure IP addresses using DHCP

To configure communications between the management station and the storage array, use Dynamic Host Configuration Protocol (DHCP) to provide IP addresses.

What you'll need

A DHCP server installed and configured on the same subnet as the storage management ports.

About this task

Each storage array has two storage management ports. Each management port will be assigned an IP address.

The following instructions refer to a storage array with two controllers.

Steps

1. If you have not already done so, connect an Ethernet cable to the management station and to management port 1 on each controller (A and B).

   The DHCP server assigns an IP address to port 1 of each controller.

   > Do not use management port 2 on either controller. Port 2 is reserved for use by Lenovo technical personnel.

   > If you disconnect and reconnect the Ethernet cable, or if the storage array is power-cycled, DHCP assigns IP addresses again. This process occurs until static IP addresses are configured. It is recommended that you avoid disconnecting the cable or power-cycling the array.

   If the storage array cannot get DHCP-assigned IP addresses within 30 seconds, the following default IP addresses are set:

   - Controller A, port 1: 169.254.128.101
   - Controller B, port 1: 169.254.128.102
   - Subnet mask: 255.255.0.0

2. Locate the MAC address label on the back of each controller, and then provide your network administrator with the MAC address for port 1 of each controller.

   Your network administrator needs the MAC addresses to determine the IP address for each controller. You will need the IP addresses to connect to your storage system through your browser.

4.3. Install and configure Lenovo ThinkSystem Host Utilities for Linux

The Lenovo ThinkSystem Host Utilities for Linux tools help you manage Lenovo storage.

> Alternatively, you can use the `SMdevices` utility to perform the same functions as the Host Utilities for Linux. The `SMdevices` utility is included as part of the `SMutils` package. The `SMutils` package is a collection of utilities to verify what the host sees from the storage array. It is included as part of the Lenovo ThinkSystem Host Utilities.

About this task

This section describes how to install ThinkSystem Host Utilities on both the Windows and Linux OS platforms, because both Windows and Linux are common management station platforms when

Linux is used for the data host.

Steps

1. Download the Lenovo ThinkSystem Host Utilities from Lenovo Data Center Support by searching "Lenovo ThinkSystem Host Utilities".
2. Run the Lenovo ThinkSystem Host Utilities.

Windows	Linux
Double-click the `lnvgy_utl_hostutilities_*.exe` installation package to start the installation.	a. Go to the directory where the `lnvgy_utl_hostutilities_*.bin` installation package is located. b. If the temp mount point does not have execute permissions, set the `IATEMPDIR` variable. Example: `IATEMPDIR=/root ./Lnvgy_utl_hostutilities-11.70.0A00.0002.bin` c. Run the `chmod +x Lnvgy_utl_hostutilities_*.bin` command to grant execute permission to the file. d. Run the `./Lnvgy_utl_hostutilities_*.bin` command to start the installer.

3. Use the installation wizard to install the software on the management station.

4.4. Access ThinkSystem System Manager and use the Setup wizard

To configure your storage array, you can use the Setup wizard in ThinkSystem System Manager.

ThinkSystem System Manager is a web-based interface embedded on each controller. To access the user interface, you point a browser to the controller's IP address. A setup wizard helps you get started with system configuration.

What you'll need

- Out-of-band management.
- A management station for accessing ThinkSystem System Manager that includes one of the following browsers:

Browser	Minimum version
Google Chrome	79
Microsoft Internet Explorer	11
Microsoft Edge	79
Mozilla Firefox	70
Safari	12

- You are using out-of-band management.

About this task

The wizard automatically relaunches when you open System Manager or refresh your browser and *at least one* of the following conditions is met:

- No pools and volume groups are detected.
- No workloads are detected.
- No notifications are configured.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been configured, the Set Administrator Password prompt appears. Role-based access management configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Use the Setup wizard to perform the following tasks:

   - **Verify hardware (controllers and drives)** — Verify the number of controllers and drives in the storage array. Assign a name to the array.

- **Verify hosts and operating systems** — Verify the host and operating system types that the storage array can access.
- **Accept pools** — Accept the recommended pool configuration for the express installation method. A pool is a logical group of drives.
- **Configure alerts** — Allow System Manager to receive automatic notifications when a problem occurs with the storage array.
- **Enable AutoSupport** — Automatically monitor the health of your storage array and have dispatches sent to technical support.

4. If you have not already created a volume, create one by going to **Storage › Volumes › Create › Volume**.

   For more information, see the online help for ThinkSystem System Manager.

4.5. Configure the multipath software

To provide a redundant path to the storage array, you can configure multipath software.

What you'll need

You must install the required packages on your system.

- For Red Hat (RHEL) hosts, verify the packages are installed by running `rpm -q device-mapper-multipath`.
- For SLES hosts, verify the packages are installed by running `rpm -q multipath-tools`.

If you have not already installed the operating system, use the media supplied by your operating system vendor.

About this task

Multipath software provides a redundant path to the storage array in case one of the physical paths is disrupted. The multipath software presents the operating system with a single virtual device that represents the active physical paths to the storage. The multipath software also manages the failover process that updates the virtual device.

You use the device mapper multipath (DM-MP) tool for Linux installations. By default, DM-MP is disabled in RHEL and SLES. Complete the following steps to enable DM-MP components on the host.

Steps

1. If a multipath.conf file is not already created, run the `# touch /etc/multipath.conf` command.
2. Use the default multipath settings by leaving the multipath.conf file blank.

3. Start the multipath service.

```
# systemctl start multipathd
```

4. Save your kernel version by running the `uname -r` command.

```
# uname -r
3.10.0-327.el7.x86_64
```

You will use this information when you assign volumes to the host.

5. Do one of the following to enable the `multipathd` daemon on boot.

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`systemctl enable multipathd`
SLES 12.x and 15.x systems:	`systemctl enable multipathd`

6. Rebuild the `initramfs` image or the `initrd` image under /boot directory:

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`dracut --force --add multipath`
SLES 12.x and 15.x systems:	`dracut --force --add multipath`

7. Make sure that the newly created /boot/initrams-* image or /boot/initrd-* image is selected in the boot configuration file.

   For example, for grub it is `/boot/grub/menu.lst` and for grub2 it is `/boot/grub2/menu.cfg`.

8. Use the "Create host manually" procedure in the online help to check whether the hosts are defined. Verify that each host type is either **Linux DM-MP (Kernel 3.10 or later)** if you enable the Automatic Load Balancing feature, or **Linux DM-MP (Kernel 3.9 or earlier)** if you disable the Automatic Load Balancing feature. If necessary, change the selected host type to the appropriate setting.
9. Reboot the host.

4.6. Set up the multipath.conf file

The multipath.conf file is the configuration file for the multipath daemon, multipathd.

The multipath.conf file overrides the built-in configuration table for multipathd. Any line in the file with a first non-white-space character of # is considered a comment line. Empty lines are ignored.

For ThinkSystem SAN OS 11.50 and newer, Lenovo recommends using the default settings as provided.

The multipath.conf files are generated as:

- For SLES, run the command:

  ```
  multipath -T > /etc/multipath.conf
  ```

- For RHEL, run the command:

  ```
  mpathconf --enable --with_multipathd y
  ```

Alternatively, edit `/etc/multipath.conf` content manually and add the section below:

```
devices {
      device {
        vendor "LENOVO"
        product "DE_Series"
        product_blacklist "Universal Xport"
        path_grouping_policy "group_by_prio"
        path_checker "rdac"
        features "2 pg_init_retries 50"
        hardware_handler "1 rdac"
        prio "rdac"
        failback immediate
        rr_weight "uniform"
        no_path_retry 30
        retain_attached_hw_handler yes
        detect_prio yes
      }
}
```

4.7. Determine SAS host identifiers - Linux

For the SAS protocol, you find the SAS addresses using the HBA utility, then use the HBA BIOS to make the appropriate configuration settings.

Before you begin this procedure, review these guidelines for HBA utilities:

- Most HBA vendors offer an HBA utility. Depending on your host operating system and CPU, use either the LSI-sas2flash(6G) or sas3flash(12G) utility.
- Host I/O ports might automatically register if the host context agent is installed.

Steps

1. Download the HBA utility from your HBA vendor's web site.
2. Install the utility.
3. Use the HBA BIOS to select the appropriate settings for your configuration.

   See the Notes column of Lenovo Storage Interoperation Center (LSIC) for recommendations.

4.8. Create partitions and filesystems

A new LUN has no partition or file system when the Linux host first discovers it. You must format the LUN before it can be used. Optionally, you can create a file system on the LUN.

What you'll need

- A LUN that is discovered by the host.
- A list of available disks. (To see available disks, run the `ls` command in the /dev/mapper folder.)

About this task

You can initialize the disk as a basic disk with a GUID partition table (GPT) or Master boot record (MBR).

Format the LUN with a file system such as ext4. Some applications do not require this step.

Steps

1. Retrieve the SCSI ID of the mapped disk by issuing the `multipath -ll` command.

   The SCSI ID is a 33-character string of hexadecimal digits, beginning with the number 3. If user-friendly names are enabled, Device Mapper reports disks as mpath instead of by a SCSI ID.

   ```
   # multipath -ll
   mpathd(360080e5000321bb8000092b1535f887a) dm-2 Lenovo DE_Series
   size=1.0T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 alua' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- 16:0:4:4 sde     69:144 active ready running
   | `- 15:0:5:4 sdf     65:176 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- 16:0:5:4 sdg     70:80  active ready running
     `- 15:0:1:4 sdh     66:0   active ready running
   ```

2. Create a new partition according to the method appropriate for your Linux OS release.

   Typically, characters identifying the partition of a disk are appended to the SCSI ID (the number 1 or p3 for instance).

```
# parted -a optimal -s -- /dev/mapper/360080e5000321bb8000092b1535f887a mklabel
gpt mkpart primary ext4 0% 100%
```

3. Create a file system on the partition.

   The method for creating a file system varies depending on the file system chosen.

```
# mkfs.ext4 /dev/mapper/360080e5000321bb8000092b1535f887a1
```

4. Create a folder to mount the new partition.

```
# mkdir /mnt/ext4
```

5. Mount the partition.

```
# mount /dev/mapper/360080e5000321bb8000092b1535f887a1 /mnt/ext4
```

4.9. Verify storage access on the host

Before using the volume, you verify that the host can write data to the volume and read it back.

What you'll need

An initialized volume that is formatted with a file system.

Steps

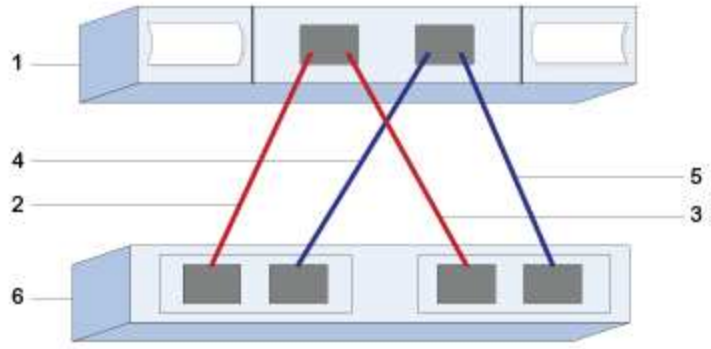
1. On the host, copy one or more files to the mount point of the disk.
2. Copy the files back to a different folder on the original disk.
3. Run the `diff` command to compare the copied files to the originals.

After you finish

Remove the file and folder that you copied.

4.10. Record your SAS configuration

You can generate and print a PDF of this page, and then use the following worksheet to record SAS storage configuration information. You need this information to perform provisioning tasks.



4.10.1. Host identifiers

Callout No.	Host (initiator) port connections	SAS address
1	Host	*not applicable*
2	Host (initiator) port 1 connected to Controller A, port 1	
3	Host (initiator) port 1 connected to Controller B, port 1	
4	Host (initiator) port 2 connected to Controller A, port 1	
5	Host (initiator) port 2 connected to Controller B, port 1	

4.10.2. Target identifiers

Recommended configurations consist of two target ports.

4.10.3. Mapping host

Mapping Host Name	
Host OS Type	

Chapter 5. iSCSI Setup

5.1. Verify the Linux configuration is supported

To ensure reliable operation, you create an implementation plan and then use the Lenovo Storage Interoperation Center (LSIC) to verify that the entire configuration is supported.

Steps

1. Go to Lenovo Storage Interoperation Center (LSIC) for interop support configuration.
2. Choose your storage model, firmware, protocol, HBA, and operating system and click here for guidance on how to use LSIC to search the products support configuration.

5.2. Configure IP addresses using DHCP

To configure communications between the management station and the storage array, use Dynamic Host Configuration Protocol (DHCP) to provide IP addresses.

What you'll need

A DHCP server installed and configured on the same subnet as the storage management ports.

About this task

Each storage array has two storage management ports. Each management port will be assigned an IP address.

The following instructions refer to a storage array with two controllers.

Steps

1. If you have not already done so, connect an Ethernet cable to the management station and to management port 1 on each controller (A and B).

   The DHCP server assigns an IP address to port 1 of each controller.

   > Do not use management port 2 on either controller. Port 2 is reserved for use by Lenovo technical personnel.

   > If you disconnect and reconnect the Ethernet cable, or if the storage array is power-cycled, DHCP assigns IP addresses again. This process occurs until static IP addresses are configured. It is recommended that you avoid disconnecting the cable or power-cycling the array.

   If the storage array cannot get DHCP-assigned IP addresses within 30 seconds, the following default IP addresses are set:

   - Controller A, port 1: 169.254.128.101
   - Controller B, port 1: 169.254.128.102
   - Subnet mask: 255.255.0.0

2. Locate the MAC address label on the back of each controller, and then provide your network administrator with the MAC address for port 1 of each controller.

   Your network administrator needs the MAC addresses to determine the IP address for each controller. You will need the IP addresses to connect to your storage system through your browser.

5.3. Install and configure Lenovo ThinkSystem Host Utilities for Linux

The Lenovo ThinkSystem Host Utilities for Linux help you manage Lenovo storage.

> Alternatively, you can use the `SMdevices` utility to perform the same functions as the Host Utilities for Linux. The `SMdevices` utility is included as part of the `SMutils` package. The `SMutils` package is a collection of utilities to verify what the host sees from the storage array. It is included as part of the Lenovo ThinkSystem Host Utilities.

About this task

This section describes how to install ThinkSystem Host Utilities on both the Windows and Linux OS platforms, because both Windows and Linux are common management station platforms when

Linux is used for the data host.

Steps

1. Download the Lenovo ThinkSystem Host Utilities from Lenovo Data Center Support by searching "Lenovo ThinkSystem Host Utilities".
2. Run the Lenovo ThinkSystem Host Utilities.

Windows	Linux
Double-click the `lnvgy_utl_hostutilities_*.exe` installation package to start the installation.	a. Go to the directory where the `lnvgy_utl_hostutilities_*.bin` installation package is located. b. If the temp mount point does not have execute permissions, set the `IATEMPDIR` variable. Example: `IATEMPDIR=/root ./lnvgy_util_hostutilities-11.70.0A00.0002.bin` c. Run the `chmod +x Lnvgy_utl_hostutilities_*.bin` command to grant execute permission to the file. d. Run the `./Lnvgy_utl_hostutilities_*.bin` command to start the installer.

3. Use the installation wizard to install the software on the management station.

5.4. Access ThinkSystem System Manager and use the Setup wizard

To configure your storage array, you can use the Setup wizard in ThinkSystem System Manager.

ThinkSystem System Manager is a web-based interface embedded on each controller. To access the user interface, you point a browser to the controller's IP address. A setup wizard helps you get started with system configuration.

What you'll need

- Out-of-band management.
- A management station for accessing ThinkSystem System Manager that includes one of the following browsers:

Browser	Minimum version
Google Chrome	79
Microsoft Internet Explorer	11
Microsoft Edge	79
Mozilla Firefox	70
Safari	12

About this task

If you are an iSCSI user, you closed the Setup wizard while configuring iSCSI.

The wizard automatically relaunches when you open System Manager or refresh your browser and *at least one* of the following conditions is met:

- No pools and volume groups are detected.
- No workloads are detected.
- No notifications are configured.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been configured, the Set Administrator Password prompt appears. Role-based access management configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Use the Setup wizard to perform the following tasks:

   - **Verify hardware (controllers and drives)** — Verify the number of controllers and drives in the storage array. Assign a name to the array.

- **Verify hosts and operating systems** — Verify the host and operating system types that the storage array can access.
- **Accept pools** — Accept the recommended pool configuration for the express installation method. A pool is a logical group of drives.
- **Configure alerts** — Allow System Manager to receive automatic notifications when a problem occurs with the storage array.
- **Enable AutoSupport** — Automatically monitor the health of your storage array and have dispatches sent to technical support.

4. If you have not already created a volume, create one by going to **Storage › Volumes › Create › Volume**.

   For more information, see the online help for ThinkSystem System Manager.

5.5. Configure the multipath software

To provide a redundant path to the storage array, you can configure multipath software.

What you'll need

You must install the required packages on your system.

- For Red Hat (RHEL) hosts, verify the packages are installed by running `rpm -q device-mapper-multipath`.
- For SLES hosts, verify the packages are installed by running `rpm -q multipath-tools`.

If you have not already installed the operating system, use the media supplied by your operating system vendor.

About this task

Multipath software provides a redundant path to the storage array in case one of the physical paths is disrupted. The multipath software presents the operating system with a single virtual device that represents the active physical paths to the storage. The multipath software also manages the failover process that updates the virtual device.

You use the device mapper multipath (DM-MP) tool for Linux installations. By default, DM-MP is disabled in RHEL and SLES. Complete the following steps to enable DM-MP components on the host.

Steps

1. If a multipath.conf file is not already created, run the `# touch /etc/multipath.conf` command.
2. Use the default multipath settings by leaving the multipath.conf file blank.

3. Start the multipath service.

```
# systemctl start multipathd
```

4. Save your kernel version by running the `uname -r` command.

```
# uname -r
3.10.0-327.el7.x86_64
```

You will use this information when you assign volumes to the host.

5. Do one of the following to enable the `multipathd` daemon on boot.

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`systemctl enable multipathd`
SLES 12.x and 15.x systems:	`systemctl enable multipathd`

6. Rebuild the `initramfs` image or the `initrd` image under /boot directory:

If you are using….	Do this…
RHEL 7.x and 8.x systems:	`dracut --force --add multipath`
SLES 12.x and 15.x systems:	`dracut --force --add multipath`

7. Use the "Create host manually" procedure in the online help to check whether the hosts are defined. Verify that each host type is either **Linux DM-MP (Kernel 3.10 or later)** if you enable the Automatic Load Balancing feature, or **Linux DM-MP (Kernel 3.9 or earlier)** if you disable the Automatic Load Balancing feature. If necessary, change the selected host type to the appropriate setting.
8. Reboot the host.

5.6. Set up the multipath.conf file

The multipath.conf file is the configuration file for the multipath daemon, multipathd. The multipath.conf file overrides the built-in configuration table for multipathd. Any line in the file with a first non-white-space character of # is considered a comment line.

For ThinkSystem SAN OS 11.50 and newer, Lenovo recommends using the default settings as provided.

The multipath.conf files are generated as:

- For SLES, run the command:

  ```
  multipath -T > /etc/multipath.conf
  ```

- For RHEL, run the command:

  ```
  mpathconf --enable --with_multipathd y
  ```

Alternatively, edit `/etc/multipath.conf` content manually and add the section below:

```
devices {
      device {
        vendor "LENOVO"
        product "DE_Series"
        product_blacklist "Universal Xport"
        path_grouping_policy "group_by_prio"
        path_checker "rdac"
        features "2 pg_init_retries 50"
        hardware_handler "1 rdac"
        prio "rdac"
        failback immediate
        rr_weight "uniform"
        no_path_retry 30
        retain_attached_hw_handler yes
        detect_prio yes
      }
}
```

5.7. Configure the switches

You configure the switches according to the vendor's recommendations for iSCSI. These recommendations might include both configuration directives as well as code updates.

You must ensure the following:

- You have two separate networks for high availability. Make sure that you isolate your iSCSI traffic to separate network segments.
- You must enable flow control **end to end**.
- If appropriate, you have enabled jumbo frames.

Port channels/LACP is not supported on the controller's switch ports. Host-side LACP is not recommended; multipathing provides the same benefits, and in some cases, better benefits.

5.8. Configure networking

You can set up your iSCSI network in many ways, depending on your data storage requirements.

Consult your network administrator for tips on selecting the best configuration for your environment.

To configure an iSCSI network with basic redundancy, connect each host port and one port from each controller to separate switches, and partition each set of host ports and controller ports on separate network segments or VLANs.

You must enable send and receive hardware flow control **end to end**. You must disable priority flow control.

If you are using jumbo frames within the IP SAN for performance reasons, make sure to configure the array, switches, and hosts to use jumbo frames. Consult your operating system and switch documentation for information on how to enable jumbo frames on the hosts and on the switches. To enable jumbo frames on the array, complete the steps in Configure array-side networking.

Many network switches must be configured above 9,000 bytes for IP overhead. Consult your switch documentation for more information.

5.9. Configure array-side networking

You use the ThinkSystem System Manager GUI to configure iSCSI networking on the array side.

What you'll need

- The IP address or domain name for one of the storage array controllers.
- A password for the System Manager GUI, or Role-Based Access Control (RBAC) or LDAP and a directory service configured for the appropriate security access to the storage array. See the ThinkSystem System Manager online help for more information about Access Management.

About this task

This task describes how to access the iSCSI port configuration from System Manager's Hardware page. You can also access the configuration from **System › Settings › Configure iSCSI ports**.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been

configured, the Set Administrator Password prompt appears. Role-based access management configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Close the Setup wizard.

   You will use the wizard later to complete additional setup tasks.

4. Select **Hardware**.
5. If the graphic shows the drives, click **Show back of shelf**.

   The graphic changes to show the controllers instead of the drives.

6. Click the controller with the iSCSI ports you want to configure.

   The controller's context menu appears.

7. Select **Configure iSCSI ports**.

   The Configure iSCSI Ports dialog box opens.

8. In the drop-down list, select the port you want to configure, and then click **Next**.
9. Select the configuration port settings, and then click **Next**.

   To see all port settings, click the **Show more port settings** link on the right of the dialog box.

Port Setting	Description
Configured ethernet port speed	Select the desired speed. The options that appear in the drop-down list depend on the maximum speed that your network can support (for example, 10 Gbps).  The optional 25Gb iSCSI host interface cards available on the controllers do not auto-negotiate speeds. You must set the speed for each port to either 10 Gb or 25 Gb. All ports must be set to the same speed.
Enable IPv4 / Enable IPv6	Select one or both options to enable support for IPv4 and IPv6 networks.
TCP listening port (Available by clicking **Show more port settings**.)	If necessary, enter a new port number. The listening port is the TCP port number that the controller uses to listen for iSCSI logins from host iSCSI initiators. The default listening port is 3260. You must enter 3260 or a value between 49152 and 65535.
MTU size (Available by clicking **Show more port settings**.)	If necessary, enter a new size in bytes for the Maximum Transmission Unit (MTU). The default Maximum Transmission Unit (MTU) size is 1500 bytes per frame. You must enter a value between 1500 and 9000.
Enable ICMP PING responses	Select this option to enable the Internet Control Message Protocol (ICMP). The operating systems of networked computers use this protocol to send messages. These ICMP messages determine whether a host is reachable and how long it takes to get packets to and from that host.

If you selected **Enable IPv4**, a dialog box opens for selecting IPv4 settings after you click **Next**. If you selected **Enable IPv6**, a dialog box opens for selecting IPv6 settings after you click **Next**. If you selected both options, the dialog box for IPv4 settings opens first, and then after you click **Next**, the dialog box for IPv6 settings opens.

10. Configure the IPv4 and/or IPv6 settings, either automatically or manually. To see all port settings, click the **Show more settings** link on the right of the dialog box.

Port setting	Description
Automatically obtain configuration	Select this option to obtain the configuration automatically.
Manually specify static configuration	Select this option, and then enter a static address in the fields. For IPv4, include the network subnet mask and gateway. For IPv6, include the routable IP address and router IP address.

11. Click **Finish**.
12. Close System Manager.

5.10. Configure host-side networking

To configure host-side networking, you must perform several steps.

About this task

You configure iSCSI networking on the host side by setting the number of node sessions per physical path, turning on the appropriate iSCSI services, configuring the network for the iSCSI ports, creating iSCSI face bindings, and establishing the iSCSI sessions between initiators and targets.

In most cases, you can use the inbox software-initiator for iSCSI CNA/NIC. You do not need to download the latest driver, firmware, and BIOS. Refer to Lenovo Storage Interoperation Center (LSIC) to determine code requirements.

Steps

1. Check the `node.session.nr_sessions` variable in the /etc/iscsi/iscsid.conf file to see the default number of sessions per physical path. If necessary, change the default number of sessions to one session.

```
node.session.nr_sessions = 1
```

2. Change the `node.session.timeo.replacement_timeout` variable in the /etc/iscsi/iscsid.conf file to `20`, from a default value of `120`.

```
node.session.timeo.replacement_timeout=20
```

3. Make sure `iscsid` and `(open-)iscsi` services are on and enabled for boot.

   Red Hat Enterprise Linux 7 and 8 (RHEL 7 and RHEL 8)

   ```
   # systemctl start iscsi
   # systemctl start iscsid
   # systemctl enable iscsi
   # systemctl enable iscsid
   ```

   SUSE Linux Enterprise Server 12 and 15 (SLES 12 and SLES 15)

   ```
   # systemctl start iscsid.service
   # systemctl enable iscsid.service
   ```

   Optionally, you set `node.startup = automatic` in /etc/iscsi/iscsid.conf before running any `iscsiadm` commands to have sessions persist after reboot.

4. Get the host IQN initiator name, which will be used to configure the host to an array.

   ```
   # cat /etc/iscsi/initiatorname.iscsi
   ```

5. Configure the network for iSCSI ports:

   In addition to the public network port, iSCSI initiators should use two or more NICs on separate private segments or vLANs.

   a. Determine the iSCSI port names using the `# ifconfig -a` command.

   b. Set the IP address for the iSCSI initiator ports. The initiator ports should be present on the same subnet as the iSCSI target ports.

      On Red Hat Enterprise Linux 7 and 8 (RHEL 7 and RHEL 8)

      ```
      # vim /etc/sysconfig/network-scripts/ifcfg-<NIC port>
      Edit:
      BOOTPROTO=none
      ONBOOT=yes
      NM_CONTROLLED=no
      Add:
      IPADDR=192.168.xxx.xxx
      NETMASK=255.255.255.0
      ```

      On SUSE Linux Enterprise Server 12 and 15 (SLES 12 and SLES 15)

```
# vim /etc/sysconfig/network-scripts/ifcfg-<NIC port>
Edit:
BOOTPROTO=static
ONBOOT=yes
NM_CONTROLLED=no
Add:
IPADDR=192.168.xxx.xxx
NETMASK=255.255.255.0
```

Be sure to set the address for both iSCSI initiator ports.

c. Restart network services.

```
# systemctl restart network
```

d. Make sure the Linux server can ping *all* of the iSCSI target ports.

6. Configure the iSCSI interfaces by creating two iSCSI iface bindings.

```
iscsiadm -m iface -I iface0 -o new
iscsiadm -m iface -I iface0 -o update -n iface.net_ifacename -v <NIC port1>
```

```
iscsiadm -m iface -I iface1 -o new
iscsiadm -m iface -I iface1 -o update -n iface.net_ifacename -v <NIC port2>
```

To list the interfaces, use `iscsiadm -m iface`.

7. Establish the iSCSI sessions between initiators and targets (four total).

a. Discover iSCSI targets. Save the IQN (it will be the same with each discovery) in the worksheet for the next step.

```
iscsiadm -m discovery -t sendtargets -p 192.168.0.1:3260 -I iface0 -P 1
```

The IQN looks like the following:

```
iqn.1992-01.com.lenovo:2365.60080e50001bf1600000000531d7be3
```

b. Create the connection between the iSCSI initiators and iSCSI targets, using ifaces.

```
iscsiadm -m node -T iqn.1992-01.com.lenovo:2365.60080e50001bf1600000000531d7be3
-p 192.168.0.1:3260 -I iface0 -l
```

c. List the iSCSI sessions established on the host.

```
# iscsiadm -m session
```

5.11. Verify IP network connections

You verify Internet Protocol (IP) network connections by using ping tests to ensure the host and array are able to communicate.

Steps

1. On the host, run one of the following commands, depending on whether jumbo frames are enabled:
   - If jumbo frames are not enabled, run this command:

     ```
     ping -I <hostIP> <targetIP>
     ```

   - If jumbo frames are enabled, run the ping command with a payload size of 8,972 bytes. The IP and ICMP combined headers are 28 bytes, which when added to the payload, equals 9,000 bytes. The -s switch sets the `packet size` bit. The -d switch sets the debug option. These options allow jumbo frames of 9,000 bytes to be successfully transmitted between the iSCSI initiator and the target.

     ```
     ping -I <hostIP> -s 8972 -d <targetIP>
     ```

     In this example, the iSCSI target IP address is `192.0.2.8`.

     ```
     #ping -I 192.0.2.100 -s 8972 -d 192.0.2.8
     Pinging 192.0.2.8 with 8972 bytes of data:
     Reply from 192.0.2.8: bytes=8972 time=2ms TTL=64
     Reply from 192.0.2.8: bytes=8972 time=2ms TTL=64
     Reply from 192.0.2.8: bytes=8972 time=2ms TTL=64
     Reply from 192.0.2.8: bytes=8972 time=2ms TTL=64
     Ping statistics for 192.0.2.8:
       Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
     Approximate round trip times in milli-seconds:
       Minimum = 2ms, Maximum = 2ms, Average = 2ms
     ```

2. Issue a `ping` command from each host's initiator address (the IP address of the host Ethernet port used for iSCSI) to each controller iSCSI port. Perform this action from each host server in the configuration, changing the IP addresses as necessary.

If the command fails (for example, returns `Packet needs to be fragmented but DF set`), verify the MTU size (jumbo frame support) for the Ethernet interfaces on the host server, storage controller, and switch ports.

5.12. Create partitions and filesystems

Because a new LUN has no partition or file system when the Linux host first discovers it, you must format the LUN before it can be used. Optionally, you can create a file system on the LUN.

What you'll need

- A LUN that is discovered by the host.
- A list of available disks. (To see available disks, run the `ls` command in the /dev/mapper folder.)

About this task

You can initialize the disk as a basic disk with a GUID partition table (GPT) or Master boot record (MBR).

Format the LUN with a file system such as ext4. Some applications do not require this step.

Steps

1. Retrieve the SCSI ID of the mapped disk by issuing the `multipath -ll` command.

   The SCSI ID is a 33-character string of hexadecimal digits, beginning with the number 3. If user-friendly names are enabled, Device Mapper reports disks as mpath instead of by a SCSI ID.

   ```
   # multipath -ll
   mpathd(360080e5000321bb8000092b1535f887a) dm-2 Lenovo DE_Series
   size=1.0T features='3 queue_if_no_path pg_init_retries 50' hwhandler='1 alua' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- 16:0:4:4 sde     69:144 active ready running
   | `- 15:0:5:4 sdf     65:176 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- 16:0:5:4 sdg     70:80  active ready running
     `- 15:0:1:4 sdh     66:0   active ready running
   ```

2. Create a new partition according to the method appropriate for your Linux OS release.

   Typically, characters identifying the partition of a disk are appended to the SCSI ID (the number 1 or p3 for instance).

   ```
   # parted -a optimal -s -- /dev/mapper/360080e5000321bb8000092b1535f887a mklabel
   gpt mkpart primary ext4 0% 100%
   ```

3. Create a file system on the partition.

   The method for creating a file system varies depending on the file system chosen.

```
# mkfs.ext4 /dev/mapper/360080e5000321bb8000092b1535f887a1
```

4. Create a folder to mount the new partition.

```
# mkdir /mnt/ext4
```

5. Mount the partition.

```
# mount /dev/mapper/360080e5000321bb8000092b1535f887a1 /mnt/ext4
```

5.13. Verify storage access on the host

Before using the volume, you verify that the host can write data to the volume and read it back.

What you'll need

An initialized volume that is formatted with a file system.

Steps

1. On the host, copy one or more files to the mount point of the disk.
2. Copy the files back to a different folder on the original disk.
3. Run the `diff` command to compare the copied files to the originals.

After you finish

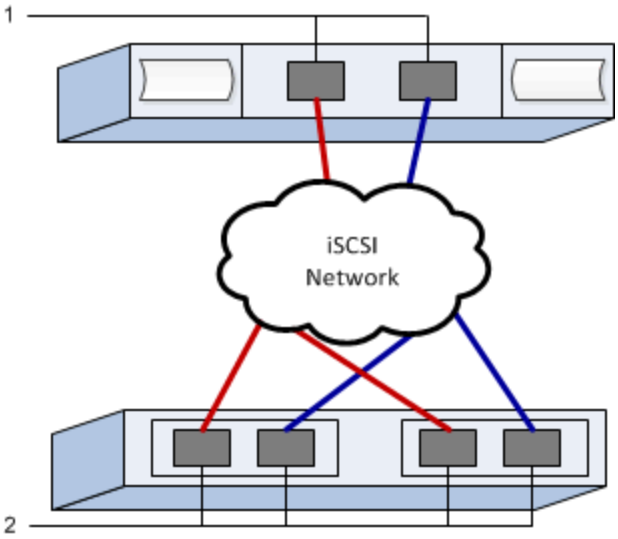
Remove the file and folder that you copied.

5.14. Record your iSCSI configuration

You can generate and print a PDF of this page, and then use the following worksheet to record iSCSI storage configuration information. You need this information to perform provisioning tasks.

5.14.1. Recommended configuration

Recommended configurations consist of two initiator ports and four target ports with one or more VLANs.



5.14.2. Target IQN

Callout No.	Target port connection	IQN
2	Target port	

5.14.3. Mapping host name

Callout No.	Host information	Name and type
1	Mapping host name	
	Host OS type	

Chapter 6. NVMe over RoCE Setup

6.1. Verify Linux support and review restrictions

As a first step, you should verify that your Linux configuration is supported and also review the controller, switch, host, and recovery restrictions.

6.1.1. Verify the Linux configuration is supported

To ensure reliable operation, you create an implementation plan and then use the Lenovo Storage Interoperation Center (LSIC) to verify that the entire configuration is supported.

Steps

1. Go to Lenovo Storage Interoperation Center (LSIC) for interop support configuration.
2. Choose your storage model, firmware, protocol, HBA, and operating system and click here for guidance on how to use LSIC to search the products support configuration.

3. Select the criteria you know you want for your configuration, and then see what compatible configuration elements apply.
4. As necessary, make the updates for your operating system and protocol that are prescribed in the tool.

   Detailed information for your chosen configuration is accessible on the View Supported Configurations page by clicking the right page arrow.

6.1.2. Verify NVMe over RoCE restrictions

Before using NVMe over RoCE, see the Lenovo Storage Interoperation Center (LSIC) to review the latest controller, host, and recovery restrictions.

Switch restrictions

RISK OF DATA LOSS. You must enable Priority Flow Control or Global Pause Control on the switch to eliminate the risk of data loss in an NVMe over RoCE environment.

Storage and disaster recovery restrictions

- Asynchronous and synchronous mirroring are not supported.
- Thin provisioning (the creation of thin volumes) is not supported.

6.2. Configure IP addresses using DHCP

To configure communications between the management station and the storage array, use Dynamic Host Configuration Protocol (DHCP) to provide IP addresses.

What you'll need

A DHCP server installed and configured on the same subnet as the storage management ports.

About this task

Each storage array has two storage management ports. Each management port will be assigned an IP address.

The following instructions refer to a storage array with two controllers.

Steps

1. If you have not already done so, connect an Ethernet cable to the management station and to management port 1 on each controller (A and B).

   The DHCP server assigns an IP address to port 1 of each controller.

   Do not use management port 2 on either controller. Port 2 is reserved for use by Lenovo technical personnel.

   If you disconnect and reconnect the Ethernet cable, or if the storage array is power-cycled, DHCP assigns IP addresses again. This process occurs until static IP addresses are configured. It is recommended that you avoid disconnecting the cable or power-cycling the array.

   If the storage array cannot get DHCP-assigned IP addresses within 30 seconds, the following default IP addresses are set:

   - Controller A, port 1: 169.254.128.101
   - Controller B, port 1: 169.254.128.102

- Subnet mask: 255.255.0.0

2. Locate the MAC address label on the back of each controller, and then provide your network administrator with the MAC address for port 1 of each controller.

   Your network administrator needs the MAC addresses to determine the IP address for each controller. You will need the IP addresses to connect to your storage system through your browser.

6.3. Access ThinkSystem System Manager and use Setup wizard

To configure your storage array, you can use the Setup wizard in ThinkSystem System Manager.

ThinkSystem System Manager is a web-based interface embedded on each controller. To access the user interface, you point a browser to the controller's IP address. A setup wizard helps you get started with system configuration.

What you'll need

- Out-of-band management.
- A management station for accessing ThinkSystem System Manager that includes one of the following browsers:

Browser	Minimum version
Google Chrome	79
Microsoft Internet Explorer	11
Microsoft Edge	79
Mozilla Firefox	70
Safari	12

About this task

The wizard automatically relaunches when you open System Manager or refresh your browser and *at least one* of the following conditions is met:

- No pools and volume groups are detected.
- No workloads are detected.
- No notifications are configured.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been configured, the Set Administrator Password prompt appears. Role-based access management configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Use the Setup wizard to perform the following tasks:

   - **Verify hardware (controllers and drives)** — Verify the number of controllers and drives in the storage array. Assign a name to the array.
   - **Verify hosts and operating systems** — Verify the host and operating system types that the storage array can access.
   - **Accept pools** — Accept the recommended pool configuration for the express installation method. A pool is a logical group of drives.
   - **Configure alerts** — Allow System Manager to receive automatic notifications when a problem occurs with the storage array.
   - **Enable AutoSupport** — Automatically monitor the health of your storage array and have dispatches sent to technical support.

4. If you have not already created a volume, create one by going to **Storage › Volumes › Create › Volume**.

   For more information, see the online help for ThinkSystem System Manager.

6.4. Configure the switch

You configure the switches according to the vendor's recommendations for NVMe over RoCE. These recommendations might include both configuration directives as well as code updates.

RISK OF DATA LOSS. You must enable Priority Flow Control or Global Pause Control on the switch to eliminate the risk of data loss in an NVMe over RoCE environment.

Steps

1. Enable Ethernet pause frame flow control **end to end** as the best practice configuration.
2. Consult your network administrator for tips on selecting the best configuration for your environment.

6.5. Set up NVMe over RoCE on the host side

NVMe initiator configuration in a RoCE environment includes installing and configuring the rdma-core and nvme-cli packages, configuring initiator IP addresses, and setting up the NVMe-oF layer on the host.

What you'll need

You must be running RHEL 8 and the latest compatible SUSE Linux Enterprise Server 12 and 15 service pack operating system. See the Lenovo Storage Interoperation Center (LSIC) for a complete list of the latest requirements.

Steps

1. Install the rdma and nvme-cli packages:

   SLES 12 or SLES 15

   ```
   # zypper install rdma-core
   # zypper install nvme-cli
   ```

   RHEL 8

   ```
   # yum install rdma-core
   # yum install nvme-cli
   ```

2. Set up IPv4 IP addresses on the ethernet ports used to connect NVMe over RoCE. For each network interface, create a configuration script that contains the different variables for that interface.

   The variables used in this step are based on server hardware and the network environment. The variables include the `IPADDR` and `GATEWAY`. These are example instructions for SLES and RHEL:

   SLES 12 and SLES 15

- Create the example file `/etc/sysconfig/network/ifcfg-eth4` as follows:

```
BOOTPROTO='static'
   BROADCAST=
   ETHTOOL_OPTIONS=
   IPADDR='192.168.1.87/24'
   GATEWAY='192.168.1.1'
   MTU=
   NAME='MT27800 Family [ConnectX-5]'
   NETWORK=
   REMOTE_IPADDR=
   STARTMODE='auto'
```

- Create the file `/etc/sysconfig/network/ifcfg-eth5` as follows:

```
BOOTPROTO='static'
   BROADCAST=
   ETHTOOL_OPTIONS=
   IPADDR='192.168.2.87/24'
   GATEWAY='192.168.2.1'
   MTU=
   NAME='MT27800 Family [ConnectX-5]'
   NETWORK=
   REMOTE_IPADDR=
   STARTMODE='auto'
```

RHEL 8

- Create the example file `/etc/sysconfig/network-scripts/ifcfg-eth4` as follows:

```
BOOTPROTO='static'
  BROADCAST=
  ETHTOOL_OPTIONS=
  IPADDR='192.168.1.87/24'
  GATEWAY='192.168.1.1'
  MTU=
  NAME='MT27800 Family [ConnectX-5]'
  NETWORK=
  REMOTE_IPADDR=
  STARTMODE='auto'
```

- Create the file `/etc/sysconfig/network-scripts/ifcfg-eth5` as follows:

```
BOOTPROTO='static'
   BROADCAST=
   ETHTOOL_OPTIONS=
   IPADDR='192.168.2.87/24'
   GATEWAY='192.168.2.1'
   MTU=
   NAME='MT27800 Family [ConnectX-5]'
   NETWORK=
   REMOTE_IPADDR=
   STARTMODE='auto'
```

3. Enable the network interfaces:

```
# ifup eth4
# ifup eth5
```

4. Set up the NVMe-oF layer on the host.
   a. Create the following file under `/etc/modules-load.d/` to load the `nvme-rdma` kernel module and make sure the kernel module will always be on, even after a reboot:

```
# cat /etc/modules-load.d/nvme-rdma.conf
  nvme-rdma
```

   To verify the `nvme-rdma` kernel module is loaded, run this command:

```
# lsmod | grep nvme
nvme_rdma              36864  0
nvme_fabrics           24576  1 nvme_rdma
nvme_core             114688  5 nvme_rdma,nvme_fabrics
rdma_cm               114688  7 rpcrdma,ib_srpt,ib_srp,nvme_rdma,ib_iser,ib_isert,rdma_ucm
ib_core               393216  15
rdma_cm,ib_ipoib,rpcrdma,ib_srpt,ib_srp,nvme_rdma,iw_cm,ib_iser,ib_umad,ib_isert,rdma_ucm,ib
_uverbs,mlx5_ib,qedr,ib_cm
t10_pi                 16384  2 sd_mod,nvme_core
```

6.6. Configure storage array NVMe over RoCE connections

If your controller includes a connection for NVMe over RoCE (RDMA over Converged Ethernet), you can configure the NVMe port settings from the Hardware page or the System page in ThinkSystem System Manager.

What you'll need

- An NVMe over RoCE host port on your controller; otherwise, the NVMe over RoCE settings are not available in System Manager.

- The IP address of the host connection.

About this task

You can access the NVMe over RoCE configuration from the **Hardware** page or from **Settings › System**. This task describes how to configure the ports from the Hardware page.

> The NVMe over RoCE settings and functions appear only if your storage array's controller includes an NVMe over RoCE port.

Steps

1. From the System Manager interface, select **Hardware**.
2. Click the controller with the NVMe over RoCE port you want to configure.

   The controller's context menu appears.

3. Select **Configure NVMe over RoCE ports**.

   The **Configure NVMe over RoCE ports** dialog box opens.

4. In the drop-down list, select the port you want to configure, and then click **Next**.
5. Select the port configuration settings you want to use, and then click **Next**.

   To see all port settings, click the **Show more port settings** link on the right of the dialog box.

Port Setting	Description
Configured ethernet port speed	Select the desired speed. The options that appear in the drop-down list depend on the maximum speed that your network can support (for example, 10 Gbps). Possible values include: • Auto-negotiate • 10 Gbps • 25 Gbps • 40 Gbps • 50 Gbps • 100 Gbps The configured NVMe over RoCE port speed should match the speed capability of the SFP on the selected port. All ports must be set to the same speed.
Enable IPv4 and/or Enable IPv6	Select one or both options to enable support for IPv4 and IPv6 networks.
MTU size (Available by clicking **Show more port settings**.)	If necessary, enter a new size in bytes for the maximum transmission unit (MTU). The default MTU size is 1500 bytes per frame. You must enter a value between 1500 and 4200.

If you selected **Enable IPv4**, a dialog box opens for selecting IPv4 settings after you click **Next**. If you selected **Enable IPv6**, a dialog box opens for selecting IPv6 settings after you click **Next**. If you selected both options, the dialog box for IPv4 settings opens first, and then after you click **Next**, the dialog box for IPv6 settings opens.

6. Configure the IPv4 and/or IPv6 settings, either automatically or manually. To see all port settings, click the **Show more settings** link on the right of the dialog box.

Port setting	Description
Automatically obtain configuration from DHCP server	Select this option to obtain the configuration automatically.

Port setting	Description
Manually specify static configuration	Select this option, and then enter a static address in the fields. For IPv4, include the network subnet mask and gateway. For IPv6, include the routable IP addresses and router IP address. If there is only one routable IP address, set the remaining address to 0:0:0:0:0:0:0:0.
Enable VLAN support (Available by clicking **Show more settings**.)	This option is only available in an iSCSI environment. It is not available in an NVMe over RoCE environment.
Enable ethernet priority (Available by clicking **Show more settings**.)	This option is only available in an iSCSI environment. It is not available in an NVMe over RoCE environment.

7. Click **Finish**.

6.7. Discover and connect to the storage from the host

Before making definitions of each host in ThinkSystem System Manager, you must discover the target controller ports from the host, and then establish NVMe connections.

Steps

1. Discover available subsystems on the NVMe-oF target for all paths using the following command:

```
nvme discover -t rdma -a target_ip_address
```

In this command, `target_ip_address` is the IP address of the target port.

The `nvme discover` command discovers all controller ports in the subsystem, regardless of host access.

```
# nvme discover -t rdma -a 192.168.1.77
Discovery Log Number of Records 2, Generation counter 0
=====Discovery Log Entry 0======
trtype:  rdma
adrfam:  ipv4
subtype: nvme subsystem
treq:    not specified
portid:  0
trsvcid: 4420
subnqn:  nqn.1992-08.com.netapp:5700.600a098000a527a7000000005ab3af94
traddr:  192.168.1.77
rdma_prtype: roce
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
=====Discovery Log Entry 1======
trtype:  rdma
adrfam:  ipv4
subtype: nvme subsystem
treq:    not specified
portid:  1
trsvcid: 4420
subnqn:  nqn.1992-08.com.netapp:5700.600a098000a527a7000000005ab3af94
traddr:  192.168.2.77
rdma_prtype: roce
rdma_qptype: connected
rdma_cms:    rdma-cm
rdma_pkey: 0x0000
```

2. Repeat step 1 for any other connections.
3. Connect to the discovered subsystem on the first path using the command: `nvme connect -t rdma -n discovered_sub_nqn-a target_ip_address -Q queue_depth_setting-l controller_loss_timeout_period`

The command listed above does not persist through reboot. The `NVMe connect` command will need to be executed after each reboot to re-establish the NVMe connections.

Connections are not established for any discovered port inaccessible by the host.

If you specify a port number using this command, the connection fails. The default port is the only port set up for connections.

The recommended queue depth setting is 1024. Override the default setting of 128 with 1024 using the `-Q 1024` command line option, as shown in the following example.

The recommended controller loss timeout period in seconds is 60 minutes (3600 seconds). Override the default setting of 600 seconds with 3600 seconds using the `-l 3600` command line option, as shown in the following example.

```
# nvme connect -t rdma -a 192.168.1.77 -n nqn.1992-
08.com.netapp:5700.600a098000a527a7000000005ab3af94 -Q 1024 -l 3600
# nvme connect -t rdma -a 192.168.2.77 -n nqn.1992-
08.com.netapp:5700.600a098000a527a7000000005ab3af94 -Q 1024 -l 3600
```

4. Repeat step 3 to connect the discovered subsystem on the second path.

6.8. Define a host

Using ThinkSystem System Manager, you define the hosts that send data to the storage array. Defining a host is one of the steps required to let the storage array know which hosts are attached to it and to allow I/O access to the volumes.

About this task

Keep these guidelines in mind when you define a host:

- You must define the host identifier ports that are associated with the host.
- Make sure that you provide the same name as the host's assigned system name.
- This operation does not succeed if the name you choose is already in use.
- The length of the name cannot exceed 30 characters.

Steps

1. Select **Storage › Hosts**.
2. Click **Create › Host**.

   The Create Host dialog box appears.

3. Select the settings for the host as appropriate.

Setting	Description
Name	Type a name for the new host.
Host operating system type	Select the following: `Linux DM-MP (Kernel 3.10 or later)`

Setting	Description
Host interface type	Select the host interface type that you want to use. If the array you configure only has one available host interface type, this setting may not be available to select.
Host ports	Do one of the following: • **Select I/O Interface** If the host ports have logged in, you can select host port identifiers from the list. This is the recommended method. • **Manual add** If the host ports have not logged in, look at /etc/nvme/hostnqn on the host to find the hostnqn identifiers and associate them with the host definition. You can manually enter the host port identifiers or copy/paste them from the /etc/nvme/hostnqn file (one at a time) into the **Host ports** field. You must add one host port identifier at a time to associate it with the host, but you can continue to select as many identifiers that are associated with the host. Each identifier is displayed in the **Host ports** field. If necessary, you also can remove an identifier by selecting the **X** next to it.

4. Click **Create**.

Result

After the host is successfully created, ThinkSystem System Manager creates a default name for each host port configured for the host.

The default alias is `<Hostname_Port Number>`. For example, the default alias for the first port created for `host IPT is IPT_1`.

6.9. Assign a volume

You must assign a volume (namespace) to a host or host cluster so it can be used for I/O operations. This assignment grants a host or host cluster access to one or more namespaces in a storage array.

About this task

Keep these guidelines in mind when you assign volumes:

- You can assign a volume to only one host or host cluster at a time.
- Assigned volumes are shared between controllers in the storage array.
- The same namespace ID (NSID) cannot be used twice by a host or a host cluster to access a volume. You must use a unique NSID.

Assigning a volume fails under these conditions:

- All volumes are assigned.
- The volume is already assigned to another host or host cluster.

The ability to assign a volume is unavailable under these conditions:

- No valid hosts or host clusters exist.
- All volume assignments have been defined.

All unassigned volumes are displayed, but functions for hosts with or without Data Assurance (DA) apply as follows:

- For a DA-capable host, you can select volumes that are either DA-enabled or not DA-enabled.
- For a host that is not DA-capable, if you select a volume that is DA-enabled, a warning states that the system must automatically turn off DA on the volume before assigning the volume to the host.

Steps

1. Select **Storage › Hosts**.
2. Select the host or host cluster to which you want to assign volumes, and then click **Assign Volumes**.

   A dialog box appears that lists all the volumes that can be assigned. You can sort any of the columns or type something in the **Filter** box to make it easier to find particular volumes.

3. Select the checkbox next to each volume that you want to assign or select the checkbox in the table header to select all volumes.
4. Click **Assign** to complete the operation.

Result

After successfully assigning a volume or volumes to a host or a host cluster, the system performs the following actions:

- The assigned volume receives the next available NSID. The host uses the NSID to access the

volume.

- The user-supplied volume name appears in volume listings associated to the host.

6.10. Display the volumes visible to the host

You can use the SMdevices tool to view volumes currently visible on the host. This tool is part of the nvme-cli package, and can be used as an alternative to the `nvme list` command.

To view information about each NVMe path to a DE-series volume, use the `nvme netapp smdevices [-o <format>]` command. The output <format> can be normal (the default if -o is not used), column, or json.

```
# nvme netapp smdevices
/dev/nvme1n1, Array Name ICTM0706SYS04, Volume Name NVMe2, NSID 1, Volume ID
000015bd5903df4a00a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n2, Array Name ICTM0706SYS04, Volume Name NVMe3, NSID 2, Volume ID
000015c05903e24000a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n3, Array Name ICTM0706SYS04, Volume Name NVMe4, NSID 4, Volume ID
00001bb0593a46f400a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n4, Array Name ICTM0706SYS04, Volume Name NVMe6, NSID 6, Volume ID
00001696593b424b00a0980000af4112, Controller A, Access State unknown, 2.15GB
/dev/nvme2n1, Array Name ICTM0706SYS04, Volume Name NVMe2, NSID 1, Volume ID
000015bd5903df4a00a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n2, Array Name ICTM0706SYS04, Volume Name NVMe3, NSID 2, Volume ID
000015c05903e24000a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n3, Array Name ICTM0706SYS04, Volume Name NVMe4, NSID 4, Volume ID
00001bb0593a46f400a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n4, Array Name ICTM0706SYS04, Volume Name NVMe6, NSID 6, Volume ID
00001696593b424b00a0980000af4112, Controller B, Access State unknown, 2.15GB
```

6.11. Set up failover on the host

To provide a redundant path to the storage array, you can configure the host to run failover.

What you'll need

You must install the required packages on your system.

- For Red Hat (RHEL) hosts, verify the packages are installed by running `rpm -q device-mapper-multipath`
- For SLES hosts, verify the packages are installed by running `rpm -q multipath-tools`

Refer to the Lenovo Storage Interoperation Center (LSIC) to ensure any required updates are installed, as multipathing might not work correctly with the GA versions of SLES or RHEL.

About this task

SLES 12 uses Device Mapper Multipath (DMMP) for multipathing for NVMe over RoCE. RHEL 8 and SLES 15 use a built-in Native NVMe Failover. Depending on which OS you are running, some additional configuration of multipath is required to get it running properly.

6.11.1. Enable Device Mapper Multipath (DMMP) for SLES 12

By default, DM-MP is disabled in RHEL and SLES. Complete the following steps to enable DM-MP components on the host.

Steps

1. Add the NVMe DE Series device entry to the devices section of the `/etc/multipath.conf` file, as shown in the following example:

```
devices {
        device {
                vendor "NVME"
                product "Lenovo DE_Series *"
                path_grouping_policy group_by_prio
                failback immediate
                no_path_retry 30
        }
}
```

2. Configure `multipathd` to start at system boot.

```
# systemctl enable multipathd
```

3. Start `multipathd` if it is not currently running.

```
# systemctl start multipathd
```

4. Verify the status of `multipathd` to make sure it is active and running:

```
# systemctl status multipathd
```

6.11.2. Set up RHEL 8 with Native NVMe Multipathing

Native NVMe Multipathing is disabled by default in RHEL 8 and must be enabled using the following procedure.

1. Set up the `modprobe` rule to turn on Native NVMe Multipathing.

```
# echo "options nvme_core multipath=y" >> /etc/modprobe.d/50-nvme_core.conf
```

2. Remake `initramfs` with the new `modprobe` parameter.

```
# dracut -f
```

3. Reboot the server to bring it up with the Native NVMe Multipathing enabled.

```
# reboot
```

4. Verify that Native NVMe Multipathing is enabled after the host boots back up.

```
# cat /sys/module/nvme_core/parameters/multipath
```

   a. If the command output is `N`, then Native NVMe Multipathing is still disabled.

   b. If the command output is `Y`, then Native NVMe Multipathing is enabled and any NVMe devices you discover will use it.

For SLES 15, Native NVMe Multipathing is enabled by default and no additional configuration is required.

6.12. Access NVMe volumes for virtual device targets

You can configure the I/O that is directed to the device target based on which OS (and by extension multipathing method) you are using.

For SLES 12, I/O is directed to virtual device targets by the Linux host. DM-MP manages the physical paths underlying these virtual targets.

6.12.1. Virtual devices are I/O targets

Make sure you are running I/O only to the virtual devices created by DM-MP and not to the physical device paths. If you are running I/O to the physical paths, DM-MP cannot manage a failover event and the I/O fails.

You can access these block devices through the `dm` device or the `symlink` in /dev/mapper. For example:

```
/dev/dm-1
/dev/mapper/eui.00001bc7593b7f5f00a0980000af4462
```

6.12.2. Example

The following example output from the `nvme list` command shows the host node name and its correlation with the namespace ID.

```
NODE         SN           MODEL           NAMESPACE

/dev/nvme1n1 021648023072 Lenovo DE_Series  10
/dev/nvme1n2 021648023072 Lenovo DE_Series  11
/dev/nvme1n3 021648023072 Lenovo DE_Series  12
/dev/nvme1n4 021648023072 Lenovo DE_Series  13
/dev/nvme2n1 021648023151 Lenovo DE_Series  10
/dev/nvme2n2 021648023151 Lenovo DE_Series  11
/dev/nvme2n3 021648023151 Lenovo DE_Series  12
/dev/nvme2n4 021648023151 Lenovo DE_Series  13
```

Column	Description
`Node`	The node name includes two parts: • The notation `nvme1` represents controller A and `nvme2` represents controller B. • The notation `n1`, `n2`, and so on represent the namespace identifier from the host perspective. These identifiers are repeated in the table, once for controller A and once for controller B.
`Namespace`	The Namespace column lists the namespace ID (NSID), which is the identifier from the storage array perspective.

In the following `multipath -ll` output, the optimized paths are shown with a `prio` value of 50, while the non-optimized paths are shown with a `prio` value of 10.

The Linux operating system routes I/O to the path group that is shown as `status=active`, while the path groups listed as `status=enabled` are available for failover.

```
eui.00001bc7593b7f500a0980000af4462 dm-0 NVME,Lenovo DE_Series
size=15G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- #:#:#:# nvme1n1 259:5 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- #:#:#:# nvme2n1 259:9  active ready running

eui.00001bc7593b7f5f00a0980000af4462 dm-0 NVME,Lenovo DE_Series
size=15G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
| `- #:#:#:# nvme1n1 259:5 failed faulty running
`-+- policy='service-time 0' prio=10 status=active
  `- #:#:#:# nvme2n1 259:9  active ready running
```

Line item	Description
`policy='service-time 0' prio=50 status=active`	This line and the following line show that `nvme1n1`, which is the namespace with an NSID of 10, is optimized on the path with a `prio` value of 50 and a `status` value of `active`. This namespace is owned by controller A.
`policy='service-time 0' prio=10 status=enabled`	This line shows the failover path for namespace 10, with a `prio` value of 10 and a `status` value of `enabled`. I/O is not being directed to the namespace on this path at the moment. This namespace is owned by controller B.
`policy='service-time 0' prio=0 status=enabled`	This example shows `multipath -l`output from a different point in time, while controller A is rebooting. The path to namespace 10 is shown as `failed faulty running` with a `prio` value of 0 and a `status` value of `enabled`.
`policy='service-time 0' prio=10 status=active`	Note that the `active` path refers to `nvme2`, so the I/O is being directed on this path to controller B.

6.13. Accessing NVMe volumes for physical NVMe device targets

You can configure the I/O directed to the device target based on which OS (and by extension multipathing method) you are using.

For RHEL 8 and SLES 15, I/O is directed to the physical NVMe device targets by the Linux host. A native NVMe multipathing solution manages the physical paths underlying the single apparent physical device displayed by the host.

It is best practice to use the links in /dev/disk/by-id/ rather than /dev/nvme0n1. For example:

```
# ls /dev/disk/by-id/ -l lrwxrwxrwx 1 root root 13 Oct 18 15:14
nvme-eui.0000320f5cad32cf00a0980000af4112 -> ../../nvme0n1
```

6.13.1. Physical NVMe devices are I/O targets

Run I/O to the physical nvme device path. There should only be one of these devices present for each namespace using the following format:

```
/dev/nvme[subsys#]n[id#]
```

All paths are virtualized using the native multipathing solution underneath this device.

You can view your paths by running:

```
# nvme list-subsys
```

Example output:

```
nvme-subsys0 - NQN=nqn.1992-08.com.netapp:5700.600a098000a522500000000589aa8a6
\
+- nvme0 rdma traddr=192.4.21.131 trsvcid=4420 live
+- nvme1 rdma traddr=192.4.22.141 trsvcid=4420 live
```

If you specify a namespace device when using the `nvme list-subsys` command, it provides additional information about the paths to that namespace:

```
# nvme list-subsys /dev/nvme0n1
nvme-subsys0 - NQN=nqn.1992-08.com.netapp:5700.600a098000af44620000000058d5dd96
\
 +- nvme0 rdma traddr=192.168.130.101 trsvcid=4420 live non-optimized
 +- nvme1 rdma traddr=192.168.131.101 trsvcid=4420 live non-optimized
 +- nvme2 rdma traddr=192.168.130.102 trsvcid=4420 live optimized
 +- nvme3 rdma traddr=192.168.131.102 trsvcid=4420 live optimized
```

There are also hooks into the multipath commands to allow you to view your path information for native failover through them as well:

```
#multipath -ll
```

To view the path information, the following must be set in /etc/multipath.conf:

```
defaults {
        enable_foreign nvme
}
```

Example output:

```
eui.0000a0335c05d57a00a0980000a5229d [nvme]:nvme0n9 NVMe,Lenovo DE_Series,08520001
size=4194304 features='n/a' hwhandler='ANA' wp=rw
|-+- policy='n/a' prio=50 status=optimized
| `- 0:0:1 nvme0c0n1 0:0 n/a optimized    live
`-+- policy='n/a' prio-10 status=non-optimized
`- 0:1:1 nvme0c1n1 0:0 n/a non-optimized    live
```

6.14. Create filesystems (SLES 12)

For SLES 12, you create a file system on the namespace and mount the filesystem.

Steps

1. Run the `multipath -ll` command to get a list of `/dev/mapper/dm` devices.

   ```
   # multipath -ll
   ```

   The result of this command shows two devices, `dm-19` and `dm-16`:

   ```
   eui.00001ffe5a94ff8500a0980000af4444 dm-19 NVME,Lenovo DE_Series
   size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- #:#:#:# nvme0n19 259:19  active ready running
   | `- #:#:#:# nvme1n19 259:115 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- #:#:#:# nvme2n19 259:51  active ready running
     `- #:#:#:# nvme3n19 259:83  active ready running
   eui.00001fd25a94fef000a0980000af4444 dm-16 NVME, Lenovo DE_Series
   size=16G features='1 queue_if_no_path' hwhandler='0' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- #:#:#:# nvme0n16 259:16  active ready running
   | `- #:#:#:# nvme1n16 259:112 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- #:#:#:# nvme2n16 259:48  active ready running
     `- #:#:#:# nvme3n16 259:80  active ready running
   ```

2. Create a file system on the partition for each `/dev/mapper/eui-` device.

   The method for creating a file system varies depending on the file system chosen. This example shows creating an `ext4` file system.

```
# mkfs.ext4 /dev/mapper/dm-19
mke2fs 1.42.11 (09-Jul-2014)
Creating filesystem with 2620928 4k blocks and 655360 inodes
Filesystem UUID: 97f987e9-47b8-47f7-b434-bf3ebbe826d0
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

3. Create a folder to mount the new device.

```
# mkdir /mnt/ext4
```

4. Mount the device.

```
# mount /dev/mapper/eui.00001ffe5a94ff8500a0980000af4444 /mnt/ext4
```

6.15. Create filesystems (RHEL 8 and SLES 15)

For RHEL 8 and SLES 15, you create a filesystem on the native nvme device and mount the filesystem.

Steps

1. Run the `multipath -ll` command to get a list of `/dev/nvme` devices.

```
# multipath -ll
```

The result of this command shows device `nvme0n6`.

```
eui.000082dd5c05d39300a0980000a52225 [nvme]:nvme0n6 NVMe,Lenovo DE_Series,08520000
size=4194304 features='n/a' hwhandler='ANA' wp=rw
|-+- policy='n/a' prio=50 status=optimized
| `- 0:0:1 nvme0c0n1 0:0 n/a optimized     live
|-+- policy='n/a' prio=50 status=optimized
| `- 0:1:1 nvme0c1n1 0:0 n/a optimized     live
|-+- policy='n/a' prio=10 status=non-optimized
| `- 0:2:1 nvme0c2n1 0:0 n/a non-optimized live
`-+- policy='n/a' prio=10 status=non-optimized
  `- 0:3:1 nvme0c3n1 0:0 n/a non-optimized live
```

2. Create a file system on the partition for each `/dev/nvme0n#` device.

The method for creating a file system varies depending on the file system chosen. This example shows creating an `ext4` file system.

```
# mkfs.ext4 /dev/disk/by-id/nvme-eui.000082dd5c05d39300a0980000a52225
mke2fs 1.42.11 (22-Oct-2019)
Creating filesystem with 2620928 4k blocks and 655360 inodes
Filesystem UUID: 97f987e9-47b8-47f7-b434-bf3ebbe826d0
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

3. Create a folder to mount the new device.

```
# mkdir /mnt/ext4
```

4. Mount the device.

```
# mount /dev/disk/by-id/nvme-eui.000082dd5c05d39300a0980000a52225 /mnt/ext4
```

6.16. Verify storage access on the host

Before using the namespace, verify that the host can write data to the namespace and read it back.

What you'll need

An initialized namespace that is formatted with a file system.

Steps

1. On the host, copy one or more files to the mount point of the disk.
2. Copy the files back to a different folder on the original disk.
3. Run the `diff` command to compare the copied files to the originals.

After you finish

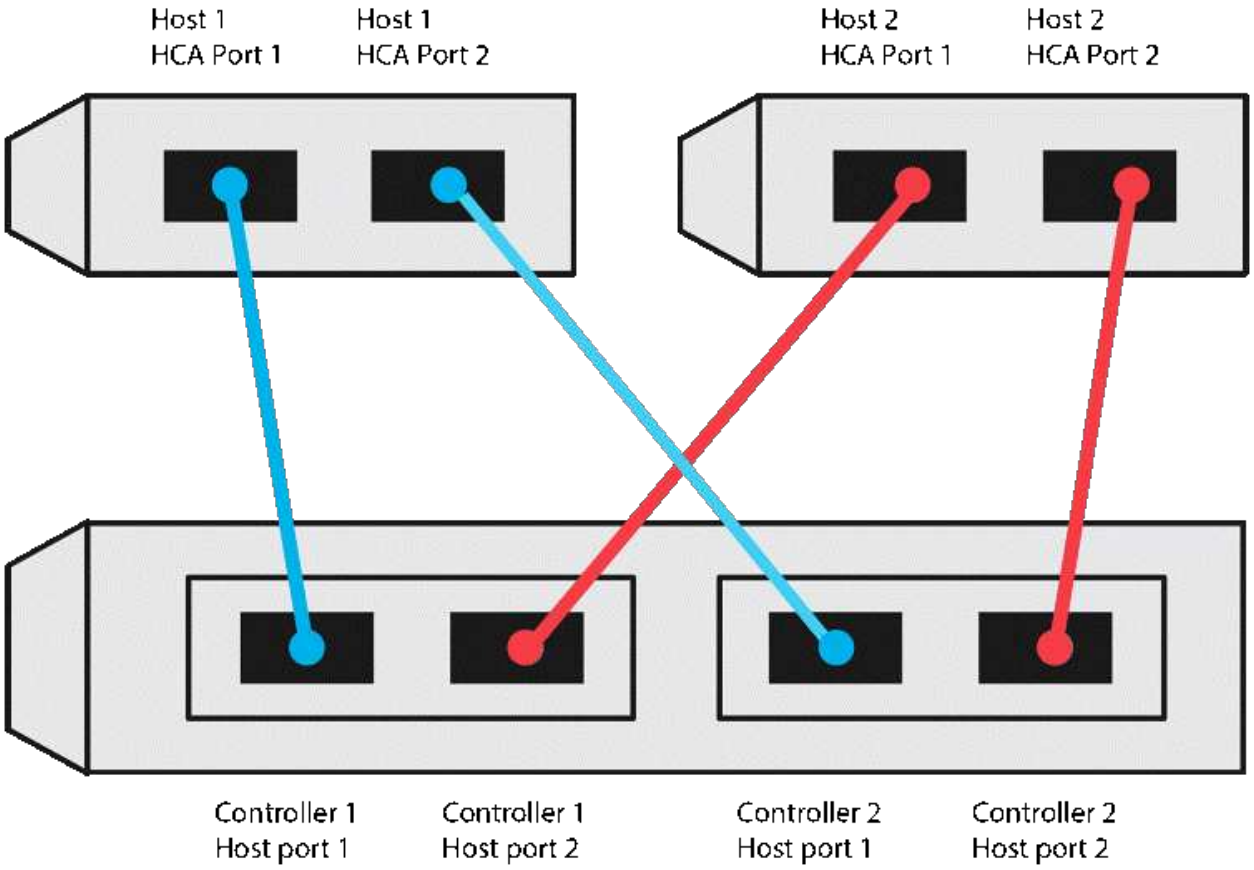
You remove the file and folder that you copied.

6.17. Record your NVMe over RoCE configuration

You can generate and print a PDF of this page, and then use the following worksheet to record NVMe over RoCE storage configuration information. You need this information to perform provisioning tasks.

6.17.1. Direct connect topology

In a direct connect topology, one or more hosts are directly connected to the subsystem. In the ThinkSystem SAN OS 11.60.2 release, we support a single connection from each host to a subsystem controller, as shown below. In this configuration, one HCA (host channel adapter) port from each host should be on the same subnet as the DE Series controller port it is connected to, but on a different subnet from the other HCA port.

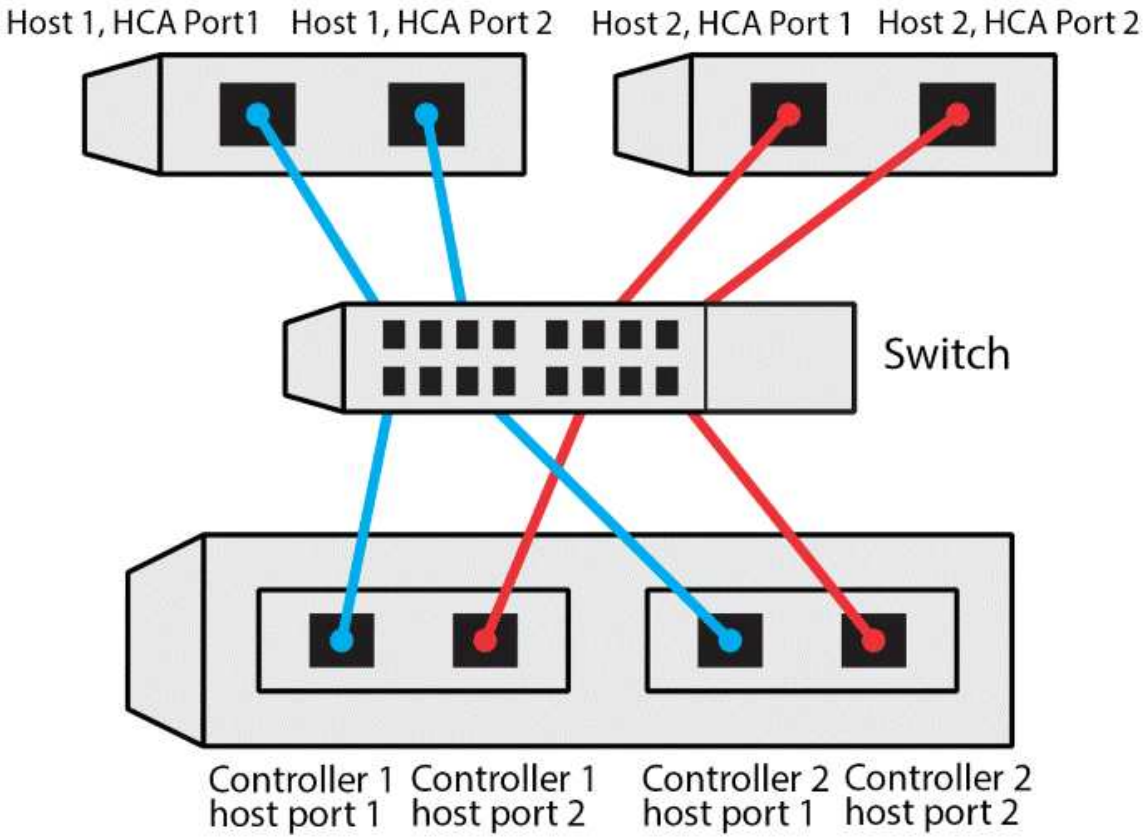


An example configuration that satisfies the requirements consists of four network subnets as follows:

- Subnet 1: Host 1 HCA Port 1 and Controller 1 Host port 1
- Subnet 2: Host 1 HCA Port 2 and Controller 2 Host port 1
- Subnet 3: Host 2 HCA Port 1 and Controller 1 Host port 2
- Subnet 4: Host 2 HCA Port 2 and Controller 2 Host port 2

6.17.2. Switch connect topology

In a fabric topology, one or more switches are used. Refer to Lenovo Storage Interoperation Center (LSIC) for a list of supported switches.



6.17.3. Host identifiers

Locate and document the initiator NQN from each host.

Host port connections	Software initiator NQN
Host (initiator) 1	
Host (initiator) 2	

6.17.4. Target NQN

Document the target NQN for the storage array.

Array name	Target NQN
Array controller (target)	

6.17.5. Target NQNs

Document the NQNs to be used by the array ports.

Array controller (target) port connections	NQN
Controller A, port 1	
Controller B, port 1	
Controller A, port 2	
Controller B, port 2	

6.17.6. Mapping host name

The mapping host name is created during the workflow.

Mapping host name	
Host OS type	

Chapter 7. NVMe over Fibre Channel Setup

7.1. Verify Linux support and review restrictions

As a first step, you should verify that your Linux configuration is supported and also review the controller, host, and recovery restrictions.

7.1.1. Verify the Linux configuration is supported

To ensure reliable operation, you create an implementation plan and then use the Lenovo Storage Interoperation Center (LSIC) to verify that the entire configuration is supported.

Steps

1. Go to Lenovo Storage Interoperation Center (LSIC) for interop support configuration.
2. Choose your storage model, firmware, protocol, HBA, and operating system and click here for guidance on how to use LSIC to search the products support configuration.

3. Select the criteria you know you want for your configuration, and then see what compatible configuration elements apply.
4. As necessary, make the updates for your operating system and protocol that are prescribed in the tool.

   Detailed information for your chosen configuration is accessible on the View Supported Configurations page by clicking the right page arrow.

7.1.2. Review restrictions for NVMe over FC

Before using NVMe over Fibre Channel, see Lenovo Storage Interoperation Center (LSIC) to review the latest controller, host, and recovery restrictions.

Storage and disaster recovery restrictions

- Asynchronous and synchronous mirroring are not supported.

- Thin provisioning (the creation of thin volumes) is not supported.

7.2. Configure IP addresses using DHCP

To configure communications between the management station and the storage array, use Dynamic Host Configuration Protocol (DHCP) to provide IP addresses.

What you'll need

A DHCP server installed and configured on the same subnet as the storage management ports.

About this task

Each storage array has two storage management ports. Each management port will be assigned an IP address.

The following instructions refer to a storage array with two controllers.

Steps

1. If you have not already done so, connect an Ethernet cable to the management station and to management port 1 on each controller (A and B).

   The DHCP server assigns an IP address to port 1 of each controller.

   Do not use management port 2 on either controller. Port 2 is reserved for use by Lenovo technical personnel.

   If you disconnect and reconnect the Ethernet cable, or if the storage array is power-cycled, DHCP assigns IP addresses again. This process occurs until static IP addresses are configured. It is recommended that you avoid disconnecting the cable or power-cycling the array.

   If the storage array cannot get DHCP-assigned IP addresses within 30 seconds, the following default IP addresses are set:

   - Controller A, port 1: 169.254.128.101
   - Controller B, port 1: 169.254.128.102
   - Subnet mask: 255.255.0.0

2. Locate the MAC address label on the back of each controller, and then provide your network administrator with the MAC address for port 1 of each controller.

   Your network administrator needs the MAC addresses to determine the IP address for each controller. You will need the IP addresses to connect to your storage system through your browser.

7.3. Access ThinkSystem System Manager and use Setup wizard

To configure your storage array, you can use the Setup wizard in ThinkSystem System Manager.

ThinkSystem System Manager is a web-based interface embedded on each controller. To access the user interface, you point a browser to the controller's IP address. A setup wizard helps you get started with system configuration.

What you'll need

- Out-of-band management.
- A management station for accessing ThinkSystem System Manager that includes one of the following browsers:

Browser	Minimum version
Google Chrome	79
Microsoft Internet Explorer	11
Microsoft Edge	79
Mozilla Firefox	70
Safari	12

About this task

The wizard automatically relaunches when you open System Manager or refresh your browser and *at least one* of the following conditions is met:

- No pools and volume groups are detected.
- No workloads are detected.
- No notifications are configured.

Steps

1. From your browser, enter the following URL: `https://<DomainNameOrIPAddress>`

   `IPAddress` is the address for one of the storage array controllers.

   The first time ThinkSystem System Manager is opened on an array that has not been configured, the Set Administrator Password prompt appears. Role-based access management

configures four local roles: admin, support, security, and monitor. The latter three roles have random passwords that cannot be guessed. After you set a password for the admin role, you can change all of the passwords using the admin credentials. For more information about the four local user roles, see the online help available in the ThinkSystem System Manager user interface.

2. Enter the System Manager password for the admin role in the Set Administrator Password and Confirm Password fields, and then click **Set Password**.

   The Setup wizard launches if there are no pools, volumes groups, workloads, or notifications configured.

3. Use the Setup wizard to perform the following tasks:
   - **Verify hardware (controllers and drives)** — Verify the number of controllers and drives in the storage array. Assign a name to the array.
   - **Verify hosts and operating systems** — Verify the host and operating system types that the storage array can access.
   - **Accept pools** — Accept the recommended pool configuration for the express installation method. A pool is a logical group of drives.
   - **Configure alerts** — Allow System Manager to receive automatic notifications when a problem occurs with the storage array.
   - **Enable AutoSupport** — Automatically monitor the health of your storage array and have dispatches sent to technical support.
4. If you have not already created a volume, create one by going to **Storage › Volumes › Create › Volume**.

   For more information, see the online help for ThinkSystem System Manager.

7.4. Configure the FC switches

Configuring (zoning) the Fibre Channel (FC) switches enables the hosts to connect to the storage array and limits the number of paths. You zone the switches using the management interface for the switches.

What you'll need

- Administrator credentials for the switches.
- The WWPN of each host initiator port and of each controller target port connected to the switch. (Use your HBA utility for discovery.)

About this task

For details about zoning your switches, see the switch vendor's documentation.

Each initiator port must be in a separate zone with all of its corresponding target ports.

Steps

1. Log in to the FC switch administration program, and then select the zoning configuration option.
2. Create a new zone that includes the first host initiator port and that also includes all of the target ports that connect to the same FC switch as the initiator.
3. Create additional zones for each FC host initiator port in the switch.
4. Save the zones, and then activate the new zoning configuration.

7.5. Set up NVMe over Fibre Channel on the host side

NVMe initiator configuration in a Fibre Channel environment includes installing and configuring the nvme-cli package and for enabling the NVMe/FC initiator on the host.

About this task

The following procedure is for RHEL 8, SLES 12, and SLES 15 using Broadcom Emulex or QLogic NVMe/FC capable FC HBAs. For more information on which versions of these OS's or HBA's are supported, consult the Lenovo Storage Interoperation Center (LSIC).

NVMe initiator configuration in a Fibre Channel environment includes installing and configuring the nvme-cli package, and enabling the NVMe/FC initiator on the host.

These are the instructions for SUSE Linux Enterprise Server 12 SP4 and 32GB FC HBAs.

Steps

1. Install the `nvme-cli` package:

   SLES 12 or SLES 15

   ```
   # zypper install nvme-cli
   ```

   RHEL 8

   ```
   # yum install nvme-cli
   ```

   a. For Qlogic, modify `/lib/systemd/system/nvmefc-boot-connections.service` after installing the Broadcom NVMe/FC autoconnect script to contain the following:

```
[Unit]
Description=Auto-connect to subsystems on FC-NVME devices found during boot

[Service]
Type=oneshot
ExecStart=/bin/sh -c "echo add > /sys/class/fc/fc_udev_device/nvme_discovery"

[Install]
WantedBy=default.target
```

2. Enable and start the `nvmefc-boot-connections` service.

```
systemctl enable nvmefc-boot-connections.service
```

```
systemctl start nvmefc-boot-connections.service
```

Host-side setup for Emulex HBAs:

The following steps are for Emulex HBAs only.

1. Set `lpfc_enable_fc4_type` to `3` to enable SLES12 SP4 as an NVMe/FC initiator.

```
# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_enable_fc4_type=3
```

2. Re-build the `initrd` to get the Emulex change and the boot parameter change.

```
# dracut --force
```

3. Reboot the host to load the changes to the `lpfc` driver.

```
# reboot
```

The host is rebooted and the NVMe/FC initiator is enabled on the host.

After completing the host-side setup, connection of the NVMe over Fibre Channel ports occur automatically.

7.6. Define a host

Using ThinkSystem System Manager, you define the hosts that send data to the

storage array. Defining a host is one of the steps required to let the storage array know which hosts are attached to it and to allow I/O access to the volumes.

About this task

Keep these guidelines in mind when you define a host:

- You must define the host identifier ports that are associated with the host.
- Make sure that you provide the same name as the host's assigned system name.
- This operation does not succeed if the name you choose is already in use.
- The length of the name cannot exceed 30 characters.

Steps

1. Select **Storage › Hosts**.
2. Click **Create › Host**.

   The Create Host dialog box appears.

3. Select the settings for the host as appropriate.

Setting	Description
Name	Type a name for the new host.
Host operating system type	Select the following: `Linux DM-MP (Kernel 3.10 or later)`
Host interface type	Select the host interface type that you want to use. If the array you configure only has one available host interface type, this setting might not be available to select.

Setting	Description
Host ports	Do one of the following: • **Select I/O Interface** If the host ports have logged in, you can select host port identifiers from the list. This is the recommended method. • **Manual add** If the host ports have not logged in, look at /etc/nvme/hostnqn on the host to find the hostnqn identifiers and associate them with the host definition. You can manually enter the host port identifiers or copy/paste them from the /etc/nvme/hostnqn file (one at a time) into the **Host ports** field. You must add one host port identifier at a time to associate it with the host, but you can continue to select as many identifiers that are associated with the host. Each identifier is displayed in the **Host ports** field. If necessary, you also can remove an identifier by selecting the **X** next to it.

4. Click **Create**.

Result

After the host is successfully created, ThinkSystem System Manager creates a default name for each host port configured for the host.

The default alias is `<Hostname_Port Number>`. For example, the default alias for the first port created for `host IPT is IPT_1`.

7.7. Assign a volume

You must assign a volume (namespace) to a host or host cluster so it can be used for I/O operations. This assignment grants a host or host cluster access to one or more namespaces in a storage array.

About this task

Keep these guidelines in mind when you assign volumes:

- You can assign a volume to only one host or host cluster at a time.
- Assigned volumes are shared between controllers in the storage array.

- The same namespace ID (NSID) cannot be used twice by a host or a host cluster to access a volume. You must use a unique NSID.

Assigning a volume fails under these conditions:

- All volumes are assigned.
- The volume is already assigned to another host or host cluster.

The ability to assign a volume is unavailable under these conditions:

- No valid hosts or host clusters exist.
- All volume assignments have been defined.

All unassigned volumes are displayed, but functions for hosts with or without Data Assurance (DA) apply as follows:

- For a DA-capable host, you can select volumes that are either DA-enabled or not DA-enabled.
- For a host that is not DA-capable, if you select a volume that is DA-enabled, a warning states that the system must automatically turn off DA on the volume before assigning the volume to the host.

Steps

1. Select **Storage › Hosts**.
2. Select the host or host cluster to which you want to assign volumes, and then click **Assign Volumes**.

   A dialog box appears that lists all the volumes that can be assigned. You can sort any of the columns or type something in the **Filter** box to make it easier to find particular volumes.

3. Select the checkbox next to each volume that you want to assign or select the checkbox in the table header to select all volumes.
4. Click **Assign** to complete the operation.

Result

After successfully assigning a volume or volumes to a host or a host cluster, the system performs the following actions:

- The assigned volume receives the next available NSID. The host uses the NSID to access the volume.
- The user-supplied volume name appears in volume listings associated to the host.

7.8. Display the volumes visible to the host

You can use the SMdevices tool to view volumes currently visible on the host. This tool is part of the nvme-cli package, and can be used as an alternative to the `nvme list` command.

To view information about each NVMe path to a DE-Series volume, use the `nvme netapp smdevices [-o <format>]` command.

The output `<format>` can be normal (the default if -o is not used), column, or json.

```
# nvme netapp smdevices
/dev/nvme1n1, Array Name ICTM0706SYS04, Volume Name NVMe2, NSID 1, Volume ID
000015bd5903df4a00a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n2, Array Name ICTM0706SYS04, Volume Name NVMe3, NSID 2, Volume ID
000015c05903e24000a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n3, Array Name ICTM0706SYS04, Volume Name NVMe4, NSID 4, Volume ID
00001bb0593a46f400a0980000af4462, Controller A, Access State unknown, 2.15GB
/dev/nvme1n4, Array Name ICTM0706SYS04, Volume Name NVMe6, NSID 6, Volume ID
00001696593b424b00a0980000af4112, Controller A, Access State unknown, 2.15GB
/dev/nvme2n1, Array Name ICTM0706SYS04, Volume Name NVMe2, NSID 1, Volume ID
000015bd5903df4a00a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n2, Array Name ICTM0706SYS04, Volume Name NVMe3, NSID 2, Volume ID
000015c05903e24000a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n3, Array Name ICTM0706SYS04, Volume Name NVMe4, NSID 4, Volume ID
00001bb0593a46f400a0980000af4462, Controller B, Access State unknown, 2.15GB
/dev/nvme2n4, Array Name ICTM0706SYS04, Volume Name NVMe6, NSID 6, Volume ID
00001696593b424b00a0980000af4112, Controller B, Access State unknown, 2.15GB
```

7.9. Set up failover on the host

To provide a redundant path to the storage array, you can configure the host to run failover.

What you'll need

You must install the required packages on your system.

- For Red Hat (RHEL) hosts, verify the packages are installed by running `rpm -q device-mapper-multipath`
- For SLES hosts, verify the packages are installed by running `rpm -q multipath-tools`

About this task

SLES 12 uses Device Mapper Multipath (DMMP) for multipathing when using NVMe over Fibre Channel. RHEL 8 and SLES 15 use a built in Native NVMe Failover. Depending on which OS you are running, some additional configuration of multipath is required to get it running properly.

7.9.1. Enable Device Mapper Multipath (DMMP) for SLES 12

By default, DM-MP is disabled in RHEL and SLES. Complete the following steps to enable DM-MP components on the host.

Steps

1. Add the NVMe DE Series device entry to the devices section of the /etc/multipath.conf file, as shown in the following example:

```
devices {
        device {
                vendor "NVME"
                product "Lenovo DE_Series *"
                path_grouping_policy group_by_prio
                failback immediate
                no_path_retry 30
        }
}
```

2. Configure `multipathd` to start at system boot.

```
# systemctl enable multipathd
```

3. Start `multipathd` if it is not currently running.

```
# systemctl start multipathd
```

4. Verify the status of `multipathd` to make sure it is active and running:

```
# systemctl status multipathd
```

7.9.2. Set up Native NMVe Multipathign for RHEL 8

About this task

Native NVMe Multipathing is disabled by default in RHEL 8 and must be enabled using the steps below.

Steps

1. Setup `modprobe` rule to turn on Native NVMe Multipathing.

```
# echo "options nvme_core multipath=y" >> /etc/modprobe.d/50-nvme_core.conf
```

2. Remake `initramfs` with new modprobe parameter.

```
# dracut -f
```

3. Reboot server to bring it up with the Native NVMe Multipathing enabled

```
# reboot
```

4. Verify Native NVMe Multipathing has been enabled after the host boots back up.

```
# cat /sys/module/nvme_core/parameters/multipath
```

   a. If the command output is `N`, then Native NVMe Multipathing is still disabled.

   b. If the command output is `Y`, then Native NVMe Multipathing is enabled and any NVMe devices you discover will use it.

For SLES 15, Native NVMe Multipathing is enabled by default and no additional configuration is required.

7.10. Access NVMe volumes for virtual device targets

You can configure the I/O directed to the device target based on which OS (and by extension multipathing method) you are using.

For SLES 12, I/O is directed to virtual device targets by the Linux host. DM-MP manages the physical paths underlying these virtual targets.

7.10.1. Virtual devices are I/O targets

Make sure you are running I/O only to the virtual devices created by DM-MP and not to the physical device paths. If you are running I/O to the physical paths, DM-MP cannot manage a failover event and the I/O fails.

You can access these block devices through the `dm` device or the `symlink` in `/dev/mapper`; for example:

```
/dev/dm-1
/dev/mapper/eui.00001bc7593b7f5f00a0980000af4462
```

7.10.2. Example

The following example output from the `nvme list` command shows the host node name and its correlation with the namespace ID.

```
NODE         SN           MODEL           NAMESPACE

/dev/nvme1n1 021648023072 Lenovo DE_Series  10
/dev/nvme1n2 021648023072 Lenovo DE_Series  11
/dev/nvme1n3 021648023072 Lenovo DE_Series  12
/dev/nvme1n4 021648023072 Lenovo DE_Series  13
/dev/nvme2n1 021648023151 Lenovo DE_Series  10
/dev/nvme2n2 021648023151 Lenovo DE_Series  11
/dev/nvme2n3 021648023151 Lenovo DE_Series  12
/dev/nvme2n4 021648023151 Lenovo DE_Series  13
```

Column	Description
`Node`	The node name includes two parts: • The notation `nvme1` represents controller A and `nvme2` represents controller B. • The notation `n1`, `n2`, and so on represent the namespace identifier from the host perspective. These identifiers are repeated in the table, once for controller A and once for controller B.
`Namespace`	The Namespace column lists the namespace ID (NSID), which is the identifier from the storage array perspective.

In the following `multipath -ll` output, the optimized paths are shown with a `prio` value of 50, while the non-optimized paths are shown with a `prio` value of 10.

The Linux operating system routes I/O to the path group that is shown as `status=active`, while the path groups listed as `status=enabled` are available for failover.

```
eui.00001bc7593b7f500a0980000af4462 dm-0 NVME, Lenovo DE_Series
size=15G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- #:#:#:# nvme1n1 259:5 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  `- #:#:#:# nvme2n1 259:9  active ready running

eui.00001bc7593b7f5f00a0980000af4462 dm-0 NVME, Lenovo DE_Series
size=15G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=0 status=enabled
| `- #:#:#:# nvme1n1 259:5 failed faulty running
`-+- policy='service-time 0' prio=10 status=active
  `- #:#:#:# nvme2n1 259:9  active ready running
```

Line item	Description
`policy='service-time 0' prio=50 status=active`	This line and the following line show that `nvme1n1`, which is the namespace with an NSID of 10, is optimized on the path with a `prio` value of 50 and a `status` value of `active`. This namespace is owned by controller A.
`policy='service-time 0' prio=10 status=enabled`	This line shows the failover path for namespace 10, with a `prio` value of 10 and a `status` value of `enabled`. I/O is not being directed to the namespace on this path at the moment. This namespace is owned by controller B.
`policy='service-time 0' prio=0 status=enabled`	This example shows `multipath -l`output from a different point in time, while controller A is rebooting. The path to namespace 10 is shown as `failed faulty running` with a `prio` value of 0 and a `status` value of `enabled`.
`policy='service-time 0' prio=10 status=active`	Note that the `active` path refers to `nvme2`, so the I/O is being directed on this path to controller B.

7.11. Access NVMe volumes for physical NVMe device targets

You can configure the I/O directed to the device target based on which OS (and by extension multipathing method) you are using.

For RHEL 8 and SLES 15, I/O is directed to the physical NVMe device targets by the Linux host. A native NVMe multipathing solution manages the physical paths underlying the single apparent physical device displayed by the host.

It is best practice to use the links in /dev/disk/by-id/ rather than /dev/nvme0n1. For example:

```
# ls /dev/disk/by-id/ -l lrwxrwxrwx 1 root root 13 Oct 18 15:14
nvme-eui.0000320f5cad32cf00a0980000af4112 -> ../../nvme0n1
```

7.11.1. Physical NVMe devices are I/O targets

Run I/O to the physical nvme device path. There should only be one of these devices present for each namespace using the following format:

```
/dev/nvme[subsys#]n[id#]
```

All paths are virtualized using the native multipathing solution underneath this device.

You can view your paths by running:

```
# nvme list-subsys
```

Example output:

```
nvme-subsys0 - NQN=nqn.1992-08.com.netapp:5700.600a098000a522500000000589aa8a6
\
+- nvme0 rdma traddr=192.4.21.131 trsvcid=4420 live
+- nvme1 rdma traddr=192.4.22.141 trsvcid=4420 live
```

If you specify a namespace device when using the `nvme list-subsys` command, it provides additional information about the paths to that namespace:

```
# nvme list-subsys /dev/nvme0n1
nvme-subsys0 - NQN=nqn.1992-08.com.netapp:5700.600a098000af44620000000058d5dd96
\
 +- nvme0 rdma traddr=192.168.130.101 trsvcid=4420 live non-optimized
 +- nvme1 rdma traddr=192.168.131.101 trsvcid=4420 live non-optimized
 +- nvme2 rdma traddr=192.168.130.102 trsvcid=4420 live optimized
 +- nvme3 rdma traddr=192.168.131.102 trsvcid=4420 live optimized
```

There are also hooks into the multipath commands to allow you to view your path information for native failover through them as well:

```
#multipath -ll
```

To view the path information, the following must be set in `/etc/multipath.conf`:

```
defaults {
        enable_foreign nvme
}
```

Example output:

```
eui.0000a0335c05d57a00a0980000a5229d [nvme]:nvme0n9 NVMe,Lenovo DE_Series ,08520001
size=4194304 features='n/a' hwhandler='ANA' wp=rw
|-+- policy='n/a' prio=50 status=optimized
| `- 0:0:1 nvme0c0n1 0:0 n/a optimized    live
`-+- policy='n/a' prio-10 status=non-optimized
`- 0:1:1 nvme0c1n1 0:0 n/a non-optimized    live
```

7.12. Create filesystems

You can create a file system on the namespace or native NVMe device and mount the filesystem.

7.12.1. Create filesystems (SLES 12)

For SLES 12, you create a file system on the desired dm device and mount the filesystem.

Steps

1. Run the `multipath -ll` command to get a list of `/dev/mapper/dm` devices.

   ```
   # multipath -ll
   ```

   The result of this command shows two devices, `dm-19` and `dm-16`:

   ```
   eui.00001ffe5a94ff8500a0980000af4444 dm-19 NVME, Lenovo DE_Series
   size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- #:#:#:# nvme0n19 259:19  active ready running
   | `- #:#:#:# nvme1n19 259:115 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- #:#:#:# nvme2n19 259:51  active ready running
     `- #:#:#:# nvme3n19 259:83  active ready running
   eui.00001fd25a94fef000a0980000af4444 dm-16 NVME, Lenovo DE_Series
   size=16G features='1 queue_if_no_path' hwhandler='0' wp=rw
   |-+- policy='service-time 0' prio=50 status=active
   | |- #:#:#:# nvme0n16 259:16  active ready running
   | `- #:#:#:# nvme1n16 259:112 active ready running
   `-+- policy='service-time 0' prio=10 status=enabled
     |- #:#:#:# nvme2n16 259:48  active ready running
     `- #:#:#:# nvme3n16 259:80  active ready running
   ```

2. Create a file system on the partition for each `/dev/mapper/eui-` device.

   The method for creating a file system varies depending on the file system chosen. This example shows creating an `ext4` file system.

```
# mkfs.ext4 /dev/mapper/dm-19
mke2fs 1.42.11 (09-Jul-2014)
Creating filesystem with 2620928 4k blocks and 655360 inodes
Filesystem UUID: 97f987e9-47b8-47f7-b434-bf3ebbe826d0
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

3. Create a folder to mount the new device.

```
# mkdir /mnt/ext4
```

4. Mount the device.

```
# mount /dev/mapper/eui.00001ffe5a94ff8500a0980000af4444 /mnt/ext4
```

7.12.2. Create filesystems (RHEL 8 and SLES 15)

For RHEL 8 and SLES 15, you create a filesystem on the native nvme device and mount the filesystem.

Steps

1. Run the multipath -ll command to get a list of /dev/nvme devices.

```
# multipath -ll
```

   The result of this command shows device nvme0n6.

```
eui.000082dd5c05d39300a0980000a52225 [nvme]:nvme0n6 NVMe,Lenovo DE_Series, 08520000
size=4194304 features='n/a' hwhandler='ANA' wp=rw
|-+- policy='n/a' prio=50 status=optimized
| `- 0:0:1 nvme0c0n1 0:0 n/a optimized     live
|-+- policy='n/a' prio=50 status=optimized
| `- 0:1:1 nvme0c1n1 0:0 n/a optimized     live
|-+- policy='n/a' prio=10 status=non-optimized
| `- 0:2:1 nvme0c2n1 0:0 n/a non-optimized live
`-+- policy='n/a' prio=10 status=non-optimized
  `- 0:3:1 nvme0c3n1 0:0 n/a non-optimized live
```

2. Create a file system on the partition for each /dev/nvme0n# device.

   The method for creating a file system varies depending on the file system chosen. This example

shows creating an ext4 file system.

```
# mkfs.ext4 /dev/disk/by-id/nvme-eui.000082dd5c05d39300a0980000a52225
mke2fs 1.42.11 (22-Oct-2019)
Creating filesystem with 2620928 4k blocks and 655360 inodes
Filesystem UUID: 97f987e9-47b8-47f7-b434-bf3ebbe826d0
Superblock backups stored on blocks:
        32768, 98304, 163840, 229376, 294912, 819200, 884736, 1605632

Allocating group tables: done
Writing inode tables: done
Creating journal (32768 blocks): done
Writing superblocks and filesystem accounting information: done
```

3. Create a folder to mount the new device.

```
# mkdir /mnt/ext4
```

4. Mount the device.

```
# mount /dev/disk/by-id/nvme-eui.000082dd5c05d39300a0980000a52225 /mnt/ext4
```

7.13. Verify storage access on the host

Before using the namespace, you verify that the host can write data to the namespace and read it back.

What you'll need

An initialized namespace that is formatted with a file system.

Steps

1. On the host, copy one or more files to the mount point of the disk.
2. Copy the files back to a different folder on the original disk.
3. Run the diff command to compare the copied files to the originals.

After you finish

Remove the file and folder that you copied.

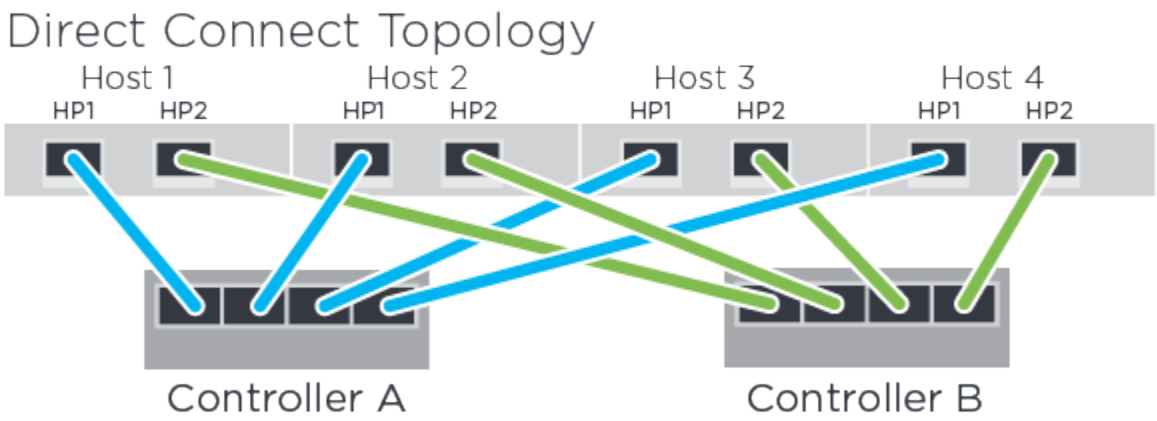
7.14. Record your NVMe over FC configuration

You can generate and print a PDF of this page, and then use the following

worksheet to record NVMe over Fibre Channel storage configuration information. You need this information to perform provisioning tasks.

7.14.1. Direct connect topology

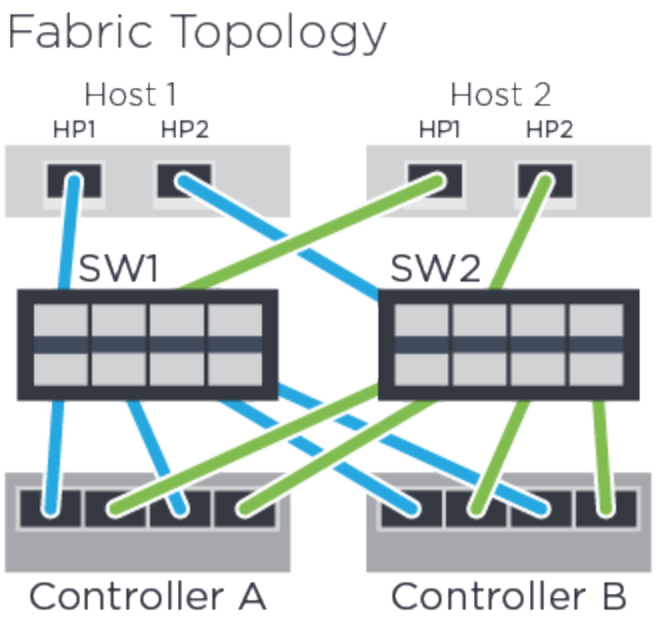
In a direct connect topology, one or more hosts are directly connected to the controller.



- Host 1 HBA Port 1 and Controller A Host port 1
- Host 1 HBA Port 2 and Controller B Host port 1
- Host 2 HBA Port 1 and Controller A Host port 2
- Host 2 HBA Port 2 and Controller B Host port 2
- Host 3 HBA Port 1 and Controller A Host port 3
- Host 3 HBA Port 2 and Controller B Host port 3
- Host 4 HBA Port 1 and Controller A Host port 4
- Host 4 HBA Port 2 and Controller B Host port 4

7.14.2. Switch connect topology

In a fabric topology, one or more switches are used. See Lenovo Storage Interoperation Center (LSIC) for a list of supported switches.



7.14.3. Host identifiers

Locate and document the initiator NQN from each host.

Host port connections	Host NQN
Host (initiator) 1	
Host (initiator) 2	

7.14.4. Target NQN

Document the target NQN for the storage array.

Array name	Target NQN
Array controller (target)	

7.14.5. Target NQNs

Document the NQNs to be used by the array ports.

Array controller (target) port connections	NQN
Controller A, port 1	
Controller B, port 1	

Array controller (target) port connections	NQN
Controller A, port 2	
Controller B, port 2	

7.14.6. Mapping host name

The mapping host name is created during the workflow.

Mapping host name	
Host OS type	

Chapter 8. Appendix

8.1. Contacting support

You can contact Support to obtain help for your issue.

You can receive hardware service through a Lenovo Authorized Service Provider. To locate a service provider authorized by Lenovo to provide warranty service, go to https://datacentersupport.lenovo.com/serviceprovider and use filter searching for different countries. For Lenovo support telephone numbers, see https://datacentersupport.lenovo.com/supportphonelist for your region support details.

8.2. Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area.

Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document is not an offer and does not provide a license under any patents or patent applications. You can send inquiries in writing to the following:

```
Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing
```

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information

contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

8.3. Trademarks