SANtricity® System Manager 11.42

Installing and Configuring for Linux® NVMe™ over InfiniBand®

Express Guide

April 2018 | 215-12291_C0
doccomments@netapp.com

NetApp®

Contents

Deciding whether to use this Express Guide

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

The express method includes the following steps:

1. Setting up an NVMe over InfiniBand communication environment.
2. Creating namespaces on the storage array and assigning a namespace ID to each one.
3. Making the namespaces available to the data host, and verifying that data can be written to them.

This guide is based on the following assumptions:

Component	Assumptions
Hardware	• You are using an E5700 or EF570 storage array with NVMe over InfiniBand host ports. • Each E5700 or EF570 controller contains at least 64 Gb of RAM. **Note:** If you are using controllers with less than 64 GB of RAM and an NVMe over Infiniband submodel configuration, a message is displayed directing you to either upgrade your controller memory to meet the required amount, or call Technical Support to help you revert to a different submodel configuration. • 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 array controllers. • You have applied power to the storage array. • You have installed all other hardware (for example, management station, switches) and made the necessary connections.
Host	• You have installed the Linux operating system. More specifically, you have installed SUSE Linux Enterprise Server 12 SP3. In the future, other operating systems may be supported. See the *NetApp Interoperability Matrix Tool* for more information.

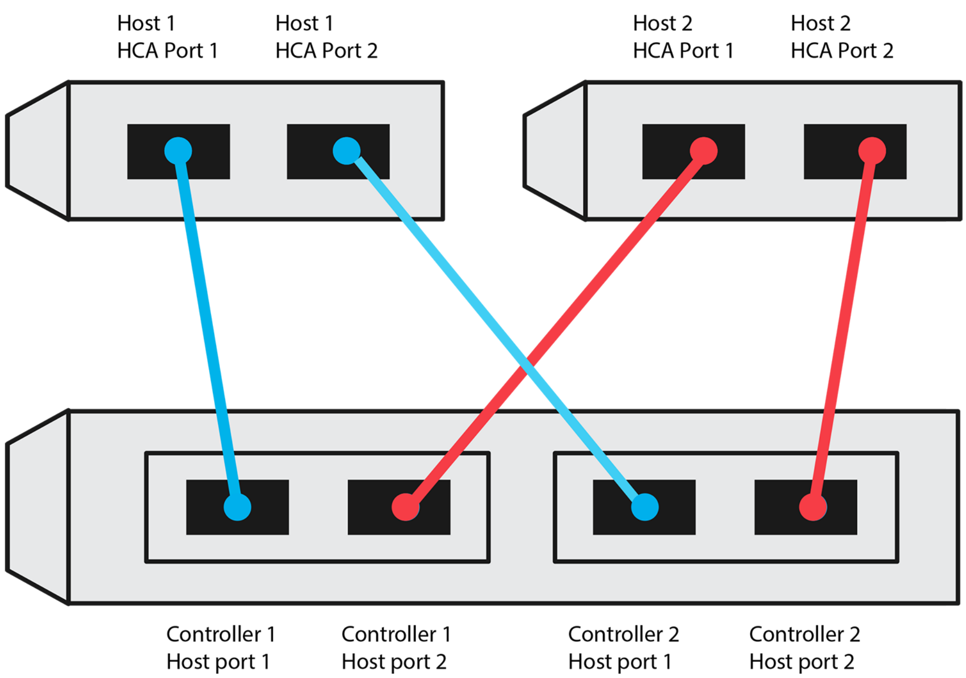
Component	Assumptions
Storage management station	• You are using a 1 Gbps or faster management network. • You are using out-of-band management, in which a storage management station sends commands to the storage array through the Ethernet connections to the controller. • For initial configuration of the storage management ports, you have attached the management station to the same subnet as the ports.
IP addressing	• You have not yet made an Ethernet connection between the management station and the storage array.
Protocol and fabric	• You have received the 100G host interface cards in an E5700 or EF570 pre-configured with the NVMe over InfiniBand protocol or the controllers were ordered with standard IB ports and need to be converted to NVMe-oF ports. • You are using an InfiniBand fabric, or you desire to directly connect to the storage array.

NVMe topologies

The EF570 and E5700 controllers include a port for implementing NVMe (Non-Volatile Memory Express) over an InfiniBand fabric. NVMe allows for high-performance communication between hosts and the storage array.

Direct connect topology

In a direct connect topology, one or more hosts are directly connected to the subsystem. In the SANtricity OS 11.42 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 E-Series controller port it is connected to, but on a different subnet from the other HCA port.

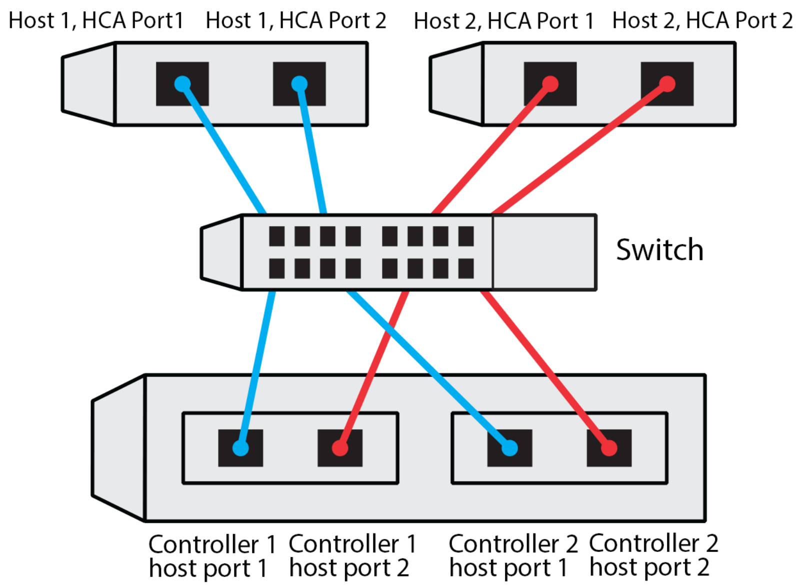


An example configuration that satisfies the requirements consists of four network subnets as follows:

- 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

Switch connect topology

In a fabric topology, one or more switches are used. Refer to *NetApp Interoperability Matrix Tool* for a list of supported switches.



Supported NVMe commands

The controller does not support every NVMe command. The supported NVMe commands are:

- Abort
- Asynchronous Event Request
- NVMe-oF Discover and Connect
- Flush
- Get Features
- Get Log Page
- Identify
- Keep Alive
- Read
- Reservation Register
- Reservation Acquire
- Reservation Release
- Reservation Report
- Set Features
- Write

See *http://www.nvmexpress.org/* for a list of all commands and their descriptions.

Supported NVMe-oF commands

The controller does not support every NVMe-oF command. The supported NVMe-oF commands are:

- Connect

- Property Get
- Property Set

See *http://www.nvmexpress.org/* for a list of all commands and their descriptions.

Verifying that the NVMe over InfiniBand configuration is supported

Verify your configuration

Verify your configuration, using the *NetApp Interoperability Matrix Tool*.

About the hardware

For the SANtricity OS 11.42 release, only the E5700 or EF570 controller can be configured for NVMe over InfiniBand. This controller must have the dual-100GB InfiniBand host port.

Feature pack

This product can be ordered with host ports pre-configured to use InfiniBand NVMe host ports. If you have ordered an EF570 or E5700 with a different IB configuration, you may convert the feature pack using the instructions in *Converting the protocol of E5700 or EF570 host ports to or from NVMe* on page 33.

Restrictions

The following restrictions are in effect for the 11.42 release. See the *NetApp Interoperability Matrix Tool* for a complete list of requirements.

Controller restrictions

- This protocol can be used only for an E5700 or EF570 controller with a minimum of 64 GB of physical memory. If the minimum memory requirements for the E5700 controller are not met during start of day operations, a message is displayed that helps you diagnose the problem.
- No simplex (i.e. single controller) configurations are supported.

Host, host protocol and host operating system restrictions

- The host must be running SUSE Linux Enterprise Server 12 SP3 operating system. See the *NetApp Interoperability Matrix Tool* for a complete list of requirements.
- The Linux OS requires a minimum kernel version of XXXX for required functionality.
- The only supported host channel adapters are from Mellanox. See the *NetApp Interoperability Matrix Tool* for more information.
- The only supported host interface card (HIC) is the dual 100G EDR IB HIC, which also supports iSER and SRP (but iSER and SRP are not supported simultaneously).
- There is no support for mixed NVMe over InfiniBand and SCSI host interfaces.

Storage and disaster recovery restrictions

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

Configuring management port IP addresses using the Quick Connect utility

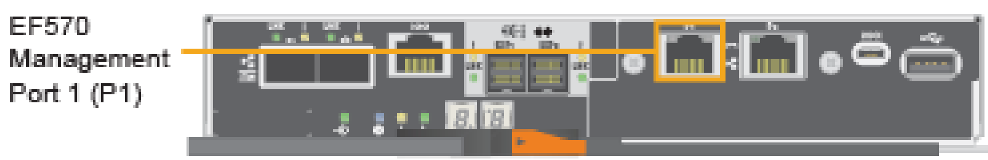
In this best-practices method for configuring communications, you configure the management station and array controllers to communicate using the Quick Connect utility.

Before you begin

- You have obtained the network configuration information from your network administrator for the controllers (IP address, subnet mask, and gateway or IP address and routable IP address).
- You have turned on the legacy management interface (SYMbol). If you have disabled the interface, see the *SANtricity System Manager online help* or the *Command Line Reference* for information on re-enabling it.

About this task

The following figure shows the location of management port 1 on the controller. Note that the EF570 back view is identical to the E5700 back view.



Steps

1. Go to *SANtricity Quick Connect*. Download and install the utility.
2. Follow the directions on the wizard screens to configure your management port and to configure the IP address of each controller.
3. Connect an Ethernet cable to management port 1 (labeled P1) on each controller, and connect the other end to your network.

   Note: Do not use port 2 on either controller. These ports are reserved for use by NetApp technical personnel.

Accessing SANtricity System Manager and using Setup wizard

You use the Setup wizard in SANtricity System Manager to configure your storage array.

Before you begin

- You have ensured that the device from which you will access SANtricity System Manager contains one of the following browsers:

Browser	Minimum version
Google Chrome	47
Microsoft Internet Explorer	11
Microsoft Edge	EdgeHTML 12
Mozilla Firefox	31
Safari	9

- 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 SANtricity 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. See *SANtricity System Manager online help* for more information on the four local user roles.

2. Enter the System Manager password for the `admin` role in the `Set Administrator Password` and `Confirm Password` fields, then select the **Set Password** button. Log out, then log back in to System Manager with the `admin` credentials.

   When you open System Manager and no pools, volumes groups, workloads, or notifications have been configured, the Setup wizard launches.

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 SANtricity System Manager.

Configuring subnet manager

Using an InfiniBand switch to run subnet manager might cause unexpected path loss during high loads. To avoid path loss, configure the subnet manager on one or more of your hosts using opensm.

Steps

1. Install the `opensm` package on any hosts that will be running the subnet manager.

   These are the instructions for SUSE Linux Enterprise Server 12 SP3:

   ```
   # zypper install opensm
   ```

2. Use the `ibstat -p` command to find `GUID0` and `GUID1` of the HCA ports. For example:

   ```
   # ibstat -p
    0x248a070300a80a80
    0x248a070300a80a81
   ```

3. The way that you configure Subnet Manager depends on your configuration:

   - If you are using a single switch, start and enable the `opensm` service, then add the HCA port identifier values you found in *step 2* on page 13 to the `opensm.conf` file on each port. Repeat for the other port.

     - Start the service immediately:

       ```
       systemctl start opensm
       ```

     - Enable the service on boot:

       ```
       systemctl enable opensm
       ```

     - Edit the `/etc/rdma/opensm.conf` file to add the identifier for that port:

       ```
       opensm -c /etc/rdma/opensm.conf

       # The port GUID on which the OpenSM is running
       guid 0x248a070300a80a80
       ```

   - If you are using the direct connect method, or if you have multiple switches, enable Subnet Manager on each port of the connected HCA on the host:

     - Add the following two lines to `/etc/rc.d/after.local` (for SUSE Linux Enterprise Server 12 SP3). Substitute the values you found in *step 2* on page 13 for `GUID0` and `GUID1`. For `P0` and `P1`, use the subnet manager priorities, with 1 being the lowest and 15 the highest:

       ```
        opensm -B -g GUID0 -p P0 -f /var/log/opensm-ib0.log
        opensm -B -g GUID1 -p P1 -f /var/log/opensm-ib1.log
       ```

An example of the command with value substitutions.

```
# cat /etc/rc.d/rc.local
 opensm -B -g 0x248a070300a80a80 -p 15 -f /var/log/opensm-ib0.log
 opensm -B -g 0x248a070300a80a81 -p 1 -f /var/log/opensm-ib1.log
```

Set up NVMe over InfiniBand on the host side

Configuring an NVMe initiator in an InfiniBand environment includes installing and configuring the `opensm`, `InfiniBand`, and `RDMA` packages, configuring initiator IP addresses, and setting up the NVMe-oF layer on the host.

Steps

1. Install the `rdma`, `nvme-cli`, and `infiniband` packages:

   These are the instructions for SUSE Linux Enterprise Server 12 SP3:

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

2. Enable `ipoib`. Edit the `/etc/rdma/rdma.conf` file and modify the entry for loading `ipoib`:

   ```
   IPOIB_LOAD=yes
   ```

3. Start and enable the `rdma` service and verify that the service is enabled and active (running):

   ```
   # systemctl start rdma
   # systemctl enable rdma
   # systemctl status rdma
   ```

   Note: If you receive an error message from the host indicating that the `rdma` service could not be started because of dependencies, reboot the host.

4. Check that both ib port links are up and the State = Active:

   ```
    # ibstat
   ```

   ```
   CA 'mlx4_0'
           CA type: MT4099
           Number of ports: 2
           Firmware version: 2.40.7000
           Hardware version: 1
           Node GUID: 0x0002c90300317850
           System image GUID: 0x0002c90300317853
           Port 1:
                   State: Active
                   Physical state: LinkUp
                   Rate: 40
                   Base lid: 4
                   LMC: 0
                   SM lid: 4
                   Capability mask: 0x0259486a
                   Port GUID: 0x0002c90300317851
                   Link layer: InfiniBand
           Port 2:
                   State: Active
                   Physical state: LinkUp
                   Rate: 56
                   Base lid: 5
                   LMC: 0
                   SM lid: 4
                   Capability mask: 0x0259486a
   ```

```
                 Port GUID: 0x0002c90300317852
                 Link layer: InfiniBand
```

5. Set up IPv4 IP addresses on the ib ports.

   For SUSE Linux Enterprise Server 12, SP3, create the file `/etc/sysconfig/network/ifcfg-ib0`

```
  BOOTPROTO='static'
  BROADCAST=
  ETHTOOL_OPTIONS=
  IPADDR='10.10.10.100/24'
  IPOIB_MODE='connected'
  MTU='65520'
  NAME=
  NETWORK=
  REMOTE_IPADDR=
  STARTMODE='auto'
```

   Then, create the file `/etc/sysconfig/network/ifcfg-ib1`.

```
  BOOTPROTO='static'
  BROADCAST=
  ETHTOOL_OPTIONS=
  IPADDR='11.11.11.100/24'
  IPOIB_MODE='connected'
  MTU='65520'
  NAME=
  NETWORK=
  REMOTE_IPADDR=
  STARTMODE='auto'
```

6. Enable the `ib` interface:

```
# ifup ib0
# ifup ib1
```

7. Verify the IP addresses you will use to connect to the array. Run this command for both `ib0` and `ib1`:

```
# ip addr show ib0
# ip addr show ib1
```

   As shown in the example below, the IP address for `ib0` is `10.10.10.255`.

```
10: ib0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 65520 qdisc pfifo_fast
state UP group default qlen 256
    link/infiniband 80:00:02:08:fe:
80:00:00:00:00:00:00:00:02:c9:03:00:31:78:51 brd 00:ff:ff:ff:ff:
12:40:1b:ff:ff:00:00:00:00:00:00:ff:ff:ff:ff
    inet 10.10.10.255 brd 10.10.10.255 scope global ib0
       valid_lft forever preferred_lft forever
    inet6 fe80::202:c903:31:7851/64 scope link
       valid_lft forever preferred_lft forever
```

As shown in the example below, the IP address for `ib1` is `11.11.11.255`.

```
10: ib1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 65520 qdisc pfifo_fast
state UP group default qlen 256
    link/infiniband 80:00:02:08:fe:
80:00:00:00:00:00:00:00:02:c9:03:00:31:78:51 brd 00:ff:ff:ff:ff:
12:40:1b:ff:ff:00:00:00:00:00:00:ff:ff:ff:ff
    inet 11.11.11.255 brd 11.11.11.255 scope global ib0
       valid_lft forever preferred_lft forever
    inet6 fe80::202:c903:31:7851/64 scope link
       valid_lft forever preferred_lft forever
```

8. Set up the NVMe-oF layer on the host.

a. Create the following files 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
```

b. Verify the `nvme-rdma` module is being loaded:

```
 # lsmod | grep nvme
```

```
nvme_rdma              32768  128
nvme_fabrics           20480  1 nvme_rdma
nvme_core              65536  66 nvme_fabrics,nvme_rdma
rdma_cm                65536  4 rpcrdma,nvme_rdma,ib_iser,rdma_ucm
ib_core               233472  13
rdma_cm,ib_cm,iw_cm,rpcrdma,mlx5_ib,ib_srp,ib_ucm,nvme_rdma,ib_iser
,ib_umad,ib_uverbs,rdma_ucm,ib_ipoib
```

c. Enable and start the `system-modules-load` service, then verify that it is running:

```
 # systemctl enable systemd-modules-load
 # systemctl start systemd-modules-load
 # systemctl status systemd-modules-load
```

Configure storage array NVMe over InfiniBand connections

If your controller includes an NVMe over InfiniBand port, you can configure the IP address of each port using SANtricity System Manager

Steps

1. Select **Hardware**.
2. If the graphic shows the drives, click **Show back of shelf**.

   The graphic changes to show the controllers instead of the drives.
3. Click the controller with the NVMe over InfiniBand ports you want to configure.

   The controller's context menu appears.
4. Select **Configure NVMe over InfiniBand ports**.

   Note: The Configure NVMe over InfiniBand ports option appears only if System Manager detects NVMe over InfiniBand ports on the controller.

   The **Configure NVMe over InfiniBand Ports** dialog box opens.
5. In the drop-down list, select the HIC port you want to configure, and then enter the IP address of the port.
6. Click **Configure**.
7. Repeat steps 5 and 6 for the other HIC ports that will be used.

Discovering and connecting to the storage from the host

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

Steps

1. Discover available subsystems on the NVMe-oF target for the first path. Use the command `nvme discover -t rdma -a` *`target_ip_address`*, where *`target_ip_address`* is the IP address of the target port.

```
# nvme discover  -t rdma -a 10.10.10.100
 Discovery Log Number of Records 1, 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:5700.600a098000af41580000000058ed54be
 traddr:
10.10.10.100

 rdma_prtype: infiniband
 rdma_qptype: connected
 rdma_cms:    rdma-cm
 rdma_pkey: 0x0000
```

2. Discover available subsystems on the NVMe-oF target for the second path. Note that there may be more than two connections, if so, repeat for any others.

```
# nvme discover  -t rdma -a 11.11.11.100
 Discovery Log Number of Records 1, 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:5700.600a098000af41580000000058ed54be
 traddr:
11.11.11.100

 rdma_prtype: infiniband
 rdma_qptype: connected
 rdma_cms:    rdma-cm
 rdma_pkey: 0x0000
```

3. Connect to the discovered subsystems 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`*

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

   Important: 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.

Important: 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 10.10.10.100 -n  nqn.1992-08.com.netapp:
5700.600a098000af41580000000058ed54be -Q 1024 -l 3600
```

4. Use the `nvme list` command to see a list of the NVMe devices currently connected. In the example below, it is `nvme0n1`.

```
 # nvme list

 Node          SN            Model
Namespace
 -------------------------------------------------------
 /dev/nvme0n1  021648023161  NetApp E-Series      1
```

```
Usage                 Format           FW Rev
--------------------------------------------------------------
5.37 GB /5.37 GB      512 B + 0 B      0842XXXX
```

5. Connect to the discovered subsystems on the second path:

```
  # nvme connect -t rdma -a 11.11.11.100 -n  nqn.1992-08.com.netapp:
5700.600a098000af41580000000058ed54be -Q 1024 -l 3600
```

6. Use the Linux `lsblk` and `grep` commands to show additional information about each block device:

```
 # lsblk | grep nvme

 nvme0n1    259:0    0     5G  0 disk
 nvme1n1    259:0    0     5G  0 disk
```

7. Use the `nvme list` command to see a new list of the NVMe devices currently connected. In the example below, it is `nvme0n1` and `nvme0n1` .

```
 # nvme list
 Node          SN            Model
Namespace
 --------------------------------------------------------------
 /dev/nvme0n1  021648023161  NetApp E-Series          1
 /dev/nvme1n1  021648023161  NetApp E-Series          1
```

```
Usage                 Format           FW Rev
--------------------------------------------------------------
5.37 GB /5.37 GB          512 B + 0 B      0842XXXX
5.37 GB /5.37 GB          512 B + 0 B      0842XXXX
```

Defining a host

Using SANtricity System Manager, you define the hosts that send data to the storage array.

About this task

Keep these guidelines in mind when you create 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.

   Field details

Setting	Description
Name	Type a name for the new host.
Host operating system type	Select *`Linux DM-MP (Kernel 3.10 or later)`* from the drop-down list.
Host interface type	Select the host interface type that you want to use.
Host ports	Do one of the following: • **Select I/O Interface** If the host ports have logged in, you can select host port identifies 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, SANtricity 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`.

Assigning 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.

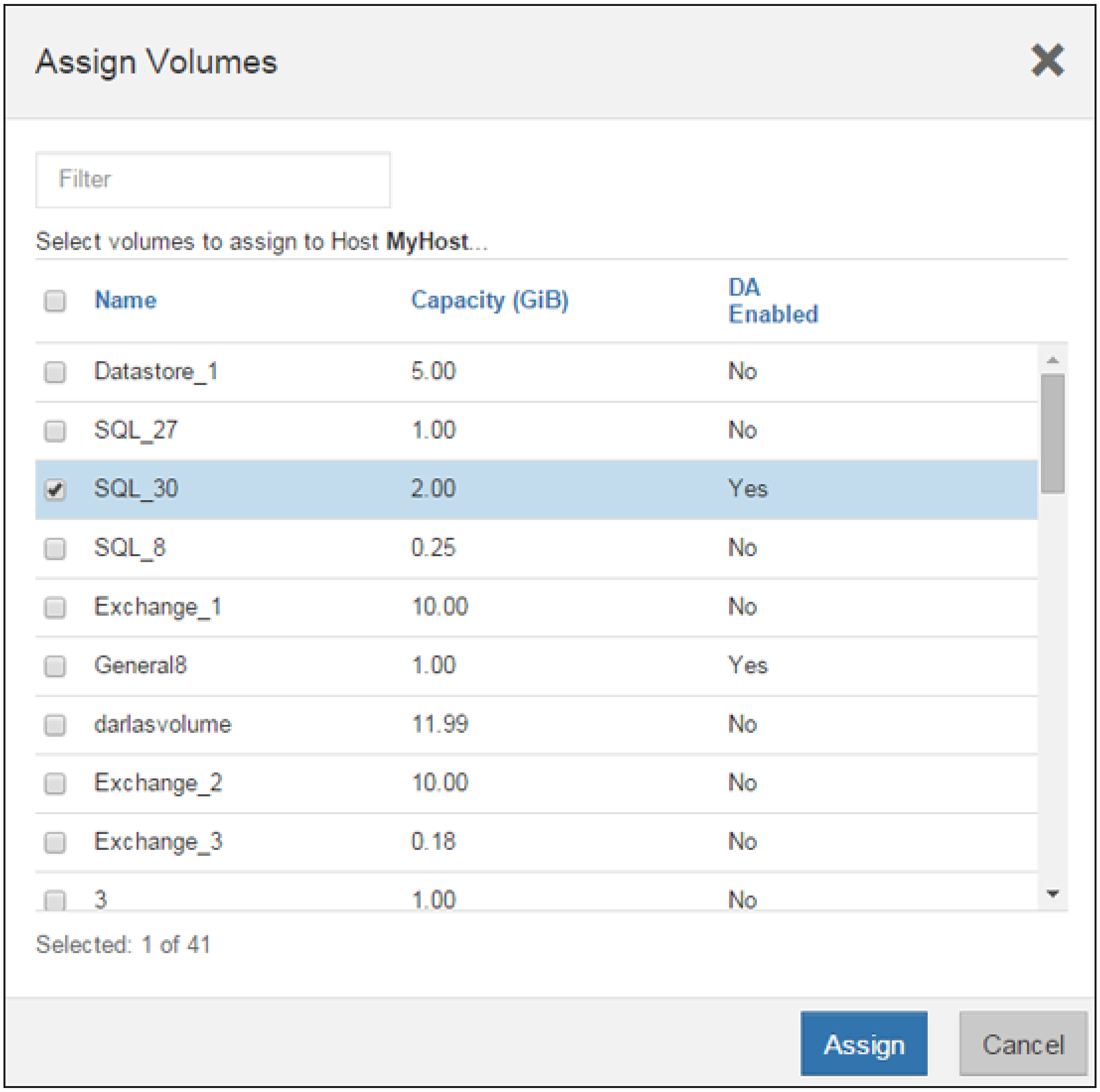
The ability to assign a volume is unavailable under these conditions:

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

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 check box next to each volume that you want to assign or select the check box 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.

Displaying the volumes visible to the host using SMdevices

Use the `SMdevices` tool, part of the `nvme-cli` package, to view the volumes currently visible on the host. This is an alternative to the `nvme list` command.

Step

1. To view information about each NVMe path to an E-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
```

Setting up failover on SLES hosts

The NVMe prioritizer package, which runs on the SUSE Linux Enterprise Server host, is an add on for Device Mapper Multipath. The package provides path failover and failback of I/O to namespaces. Any volume ownership transfers are a result of the prioritizer directing I/O to a new controller as a result of path failures or path restores.

About this task

The NVMe failover package is a special prioritizer in the host to set the active path. It runs periodically to check path states and to switch paths if the namespace is no longer accessible via the active path. As part of the Express workflow, you will install the prioritizer in *Configuring the host to use failover* on page 26.

The controller provides a list of optimal namespaces for each path. The host reads this data to understand that I/O should be directed to this path. If a path fails, the host redirects I/O to the surviving path. The surviving path goes to a different controller in the storage array. This provides the failover.

If the path that was failed becomes restored, the next time the failover prioritizer runs, it calculates that the optimal path is not the one that is being used, and redirects I/O to the optimal path. Approximately 5 minutes later, the controller firmware transfers ownership of the volume to the controller receiving the I/O on the optimal path. This provides the failback.

Important: Failover is supported only on SUSE Linux Enterprise Server 12 SP3.

Configuring the host to run failover

The SUSE Linux Enterprise Server host requires configuration changes to run failover. The failover solution uses DM-MP.

Steps

1. Install the NetApp NVMe `prio` package. It can be found at *https://mysupport.netapp.com/NOW/download/software/eseries_cfw/0.3.0/* .

```
# zypper install netapp-nvme-prio-0.3.0-release.x86_64.rpm
```

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

```
devices {
        device {
                vendor "NVME"
                product "NetApp E-Series*"
                path_grouping_policy group_by_prio
                prio netappnvme
                failback immediate
                no_path_retry 30
        }
}
```

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

```
# systemctl enable multipathd
```

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

```
# systemctl start multipathd
```

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

```
# systemctl status multipathd
```

Running I/O when using NVMe failover

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

Virtual devices are I/O targets

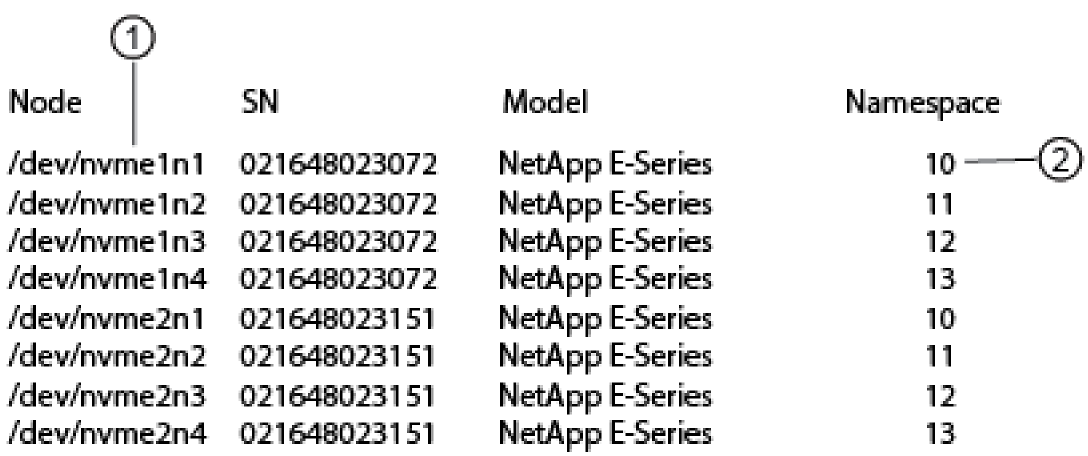
Make sure you are running I/O only to the virtual devices created by DM-MP and not to the physical device paths. If you are running I/O to the physical paths, device mapper cannot manage a failover event and the I/O fails.

To run raw I/O, make sure your targets are virtual devices created by DM-MP to manage the physical paths. 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
```

Example

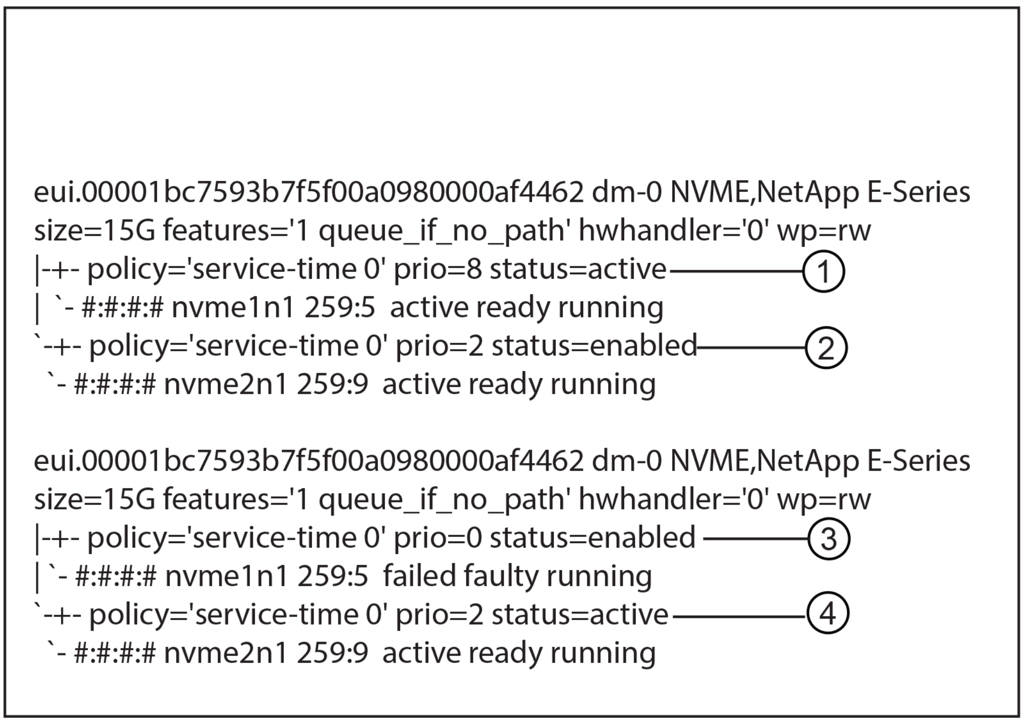
The following example output from the `nvme list` command shows the host node name and its correlation with the namespace ID.



Callout	Description
1	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.
2	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 8, while the non-optimized paths are shown with a `prio` value of 2.

Linux 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.



Callout	Description
1	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 8 and a `status` value of `active`. This namespace is owned by controller A.
2	This line shows the failover path for namespace 10, with a `prio` value of 2 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.
3	This example shows `multipath -ll` 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`.

Callout	Description
4	Note that the `active` path refers to `nvme2`, so the I/O is being directed on this path to controller B.

Creating partitions and filesystems

You can create a partition on the multipath device, optionally create a file system on the namespace, and mount the partition.

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,NetApp E-Series
size=10G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=8 status=active
| |- #:#:#:# nvme0n19 259:19  active ready running
| `- #:#:#:# nvme1n19 259:115 active ready running
`-+- policy='service-time 0' prio=2 status=enabled
  |- #:#:#:# nvme2n19 259:51  active ready running
  `- #:#:#:# nvme3n19 259:83  active ready running
eui.00001fd25a94fef000a0980000af4444 dm-16 NVME,NetApp E-Series
size=16G features='1 queue_if_no_path' hwhandler='0' wp=rw
|-+- policy='service-time 0' prio=8 status=active
| |- #:#:#:# nvme0n16 259:16  active ready running
| `- #:#:#:# nvme1n16 259:112 active ready running
`-+- policy='service-time 0' prio=2 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/dm` device.

   The method for creating a file system varies depending on the file system chosen. In this example, we are creating an `ext4` file system.

Example

```
# 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 partition.

Example

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

4. Mount the partition.

Example

```
# mount /dev/mapper/dm-19 /mnt/ext4
```

Verifying storage access on the host

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

Before you begin

You must have initialized the namespace and formatted it 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.

Converting the protocol of E5700 or EF570 host ports to or from NVMe

E5700 or EF570 host ports can be converted from one host protocol to another. However, you only need to do this if you have not ordered the storage array already configured with the correct protocol.

Note: Any existing host connections will not be available after the host protocol conversion.

Obtaining the feature pack (NVMe related)

To obtain the feature pack, you need the serial number from the controller shelf, a Feature Activation Code, and the Feature Enable Identifier for the storage array.

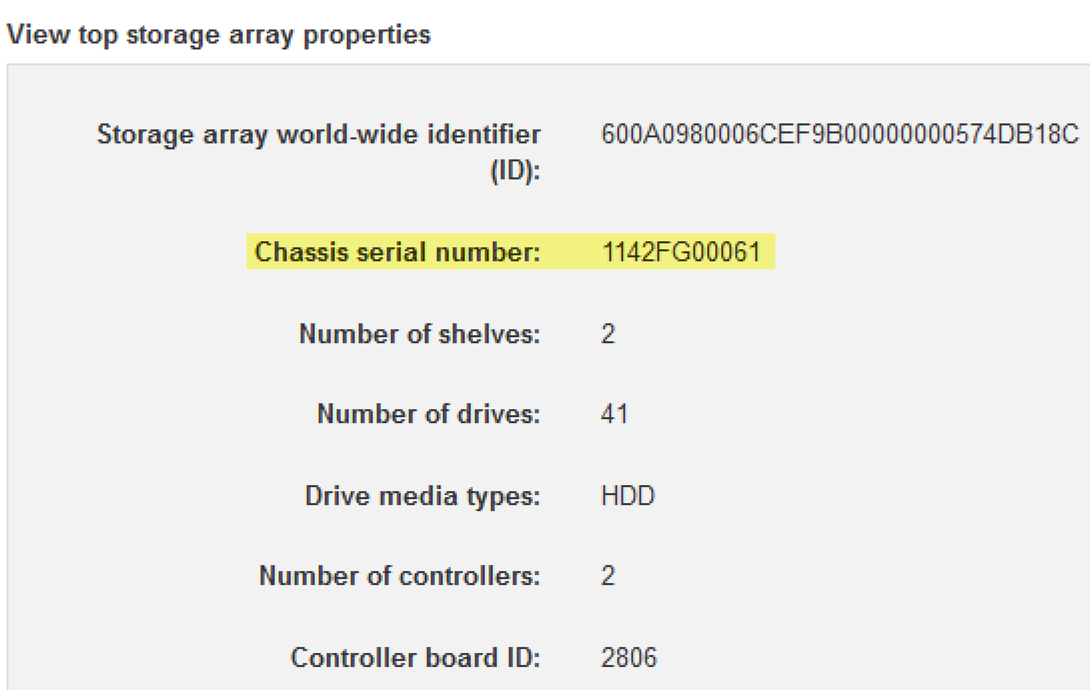
Steps

1. Locate the serial number.

   a. From SANtricity System Manager, select **Support > Support Center**.

   b. With the **Support Resources** tab selected, scroll to the **View top storage array properties** section.

   c. Locate the **Chassis Serial Number**, and copy this value to a text file.



2. Locate the **feature pack submodel ID**.

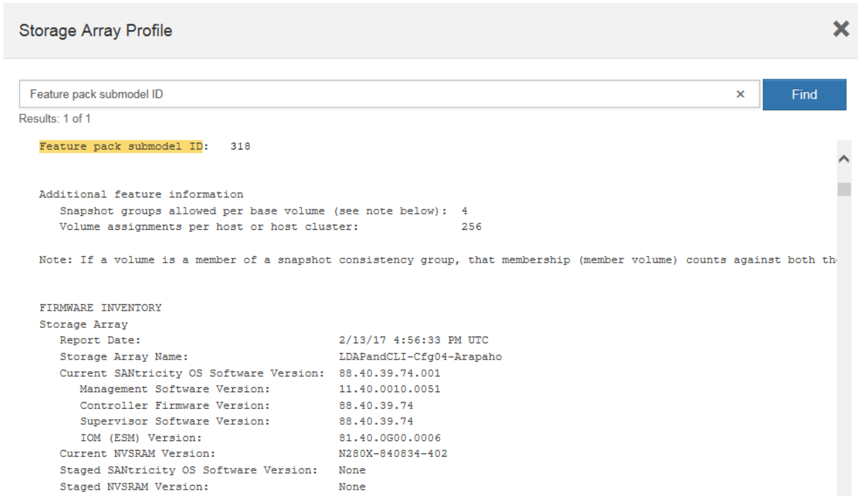
   a. From the SANtricity System Manager, select **Support**.

   b. Select the **Support Center** tile.

c. On the Support Resources tab, locate and select the **Storage Array Profile** link.

d. Type **feature pack submodel ID** in the text box, and click **Find**.

e. Locate the feature pack submodel ID for the starting configuration.



3. Using the feature pack submodel ID, locate the corresponding Controller submodel ID for the starting configuration and find the Feature Activation Code for the desired ending configuration within the table below. Then, copy that Feature Activation Code to a text file.

Note: If you are not using the IB HIC, you can ignore the *HIC Ports* column in the following tables:

Table 1: Encryption capable Feature Activation Codes for E5700/EF570

Starting configuration			Ending configuration			Feature Activation Code
Controller submodel ID	Baseboard ports	HIC ports	Controller submodel ID	Baseboard ports	HIC ports	
360	FC	iSER	382	FC	NVMe	KGI-ISB-ZDHQF
			383	iSCSI	NVMe	ZGQ-5TB-ZHVU4
361	FC	SRP	382	FC	NVMe	TGV-8TB-ZKTH6
			383	iSCSI	NVMe	1G3-UTB-Z09D4
362	iSCSI	iSER	382	FC	NVMe	BGN-7UB-ZY1YY
			383	iSCSI	NVMe	BGS-AUB-Z2YNG
363	iSCSI	SRP	382	FC	NVMe	YGZ-WUB-Z6ERJ
			383	iSCSI	NVMe	NG5-ZUB-Z8C8J
382	FC	NVMe	360	FC	iSER	3GN-TVB-ZINKM
			361	FC	SRP	4GK-AUB-ZG69H
			362	iSCSI	iSER	BGP-DVB-ZJ4YC
			363	iSCSI	SRP	BGU-GVB-ZM3KW
			383	iSCSI	NVMe	4GX-ZVB-ZNJVD

Starting configuration			Ending configuration			Feature Activation Code
Controller submodel ID	Baseboard ports	HIC ports	Controller submodel ID	Baseboard ports	HIC ports	
383	iSCSI	NVMe	360	FC	iSER	TGS-WVB-ZKL9T
			361	FC	SRP	RGZ-JVB-Z017I
			362	iSCSI	iSER	WG2-3VB-ZQHLF
			363	iSCSI	SRP	QG7-6VB-ZSF8M
			382	FC	NVMe	PGA-PVB-ZUWMX

Table 2: Non-encryption Feature Activation Codes for E5700/EF570

Starting configuration			Ending configuration			Feature Activation Code
Controller submodel ID	Baseboard ports	HIC ports	Controller submodel ID	Baseboard ports	HIC ports	
365	FC	iSER	384	FC	NVMe	DGH-YWB-ZHX5H
			385	iSCSI	NVMe	UGM-2XB-ZKV0B
366	FC	SRP	384	FC	NVMe	DGW-LWB-Z76KE
			385	iSCSI	NVMe	KG2-0WB-Z9477
367	iSCSI	iSER	384	FC	NVMe	TGD-1TB-ZT5TL
			385	iSCSI	NVMe	9GG-KTB-ZUM15
368	iSCSI	SRP	384	FC	NVMe	3GI-QUB-ZFP1Y
			385	iSCSI	NVMe	8GD-NUB-ZDRCD
384	FC	NVMe	365	FC	iSER	SGF-SVB-ZWU9M
			366	FC	SRP	7GH-CVB-ZYBGV
			367	iSCSI	iSER	6GK-VVB-ZZSRN
			368	iSCSI	SRP	RGM-FWB-Z195H
			385	iSCSI	NVMe	FGP-YWB-Z3QK0
385	iSCSI	NVMe	365	FC	iSER	GG5-8WB-ZBKEM
			366	FC	SRP	KG7-RWB-ZC2RZ
			367	iSCSI	iSER	NGC-VWB-ZFZEN
			368	iSCSI	SRP	4GE-FWB-ZGGQJ
			384	FC	NVMe	1G9-CWB-ZDIZR

Note: If your controller submodel ID is not listed, contact *NetApp Support*.

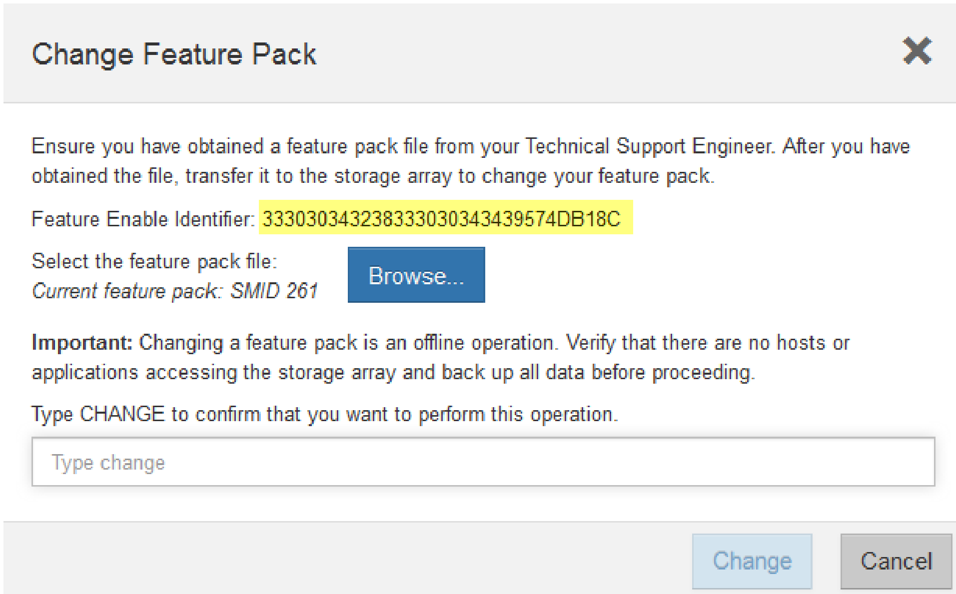
4. In System Manager, locate the Feature Enable Identifier.

   a. Go to **Settings > System**.

b. Scroll down to **Add-ons**.

c. Under **Change Feature Pack**, locate the **Feature Enable Identifer**.

d. Copy and paste this 32-digit number to a text file.



5. Go to *NetApp License Activation: Storage Array Premium Feature Activation*, and enter the information required to obtain the feature pack.

- Chassis serial number
- Feature Activation Code
- Feature Enable Identifier

Important: The Premium Feature Activation web site includes a link to "Premium Feature Activation Instructions." Do not attempt to use those instructions for this procedure.

6. Choose whether to receive the key file for the feature pack in an email or download it directly from the site.

Stopping host I/O

You must stop all I/O operations from the host before converting the protocol of the host ports. You cannot access data on the storage array until you successfully complete the conversion.

Steps

1. Ensure that no I/O operations are occurring between the storage array and all connected hosts. For example, you can perform these steps:

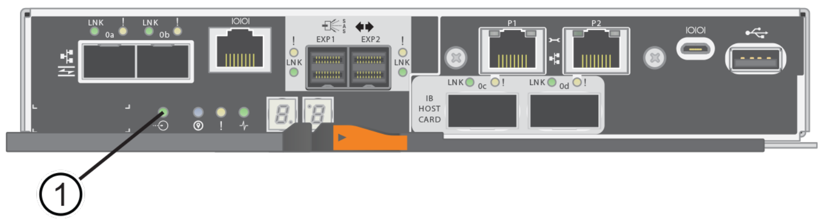
- Stop all processes that involve the LUNs mapped from the storage to the hosts.
- Ensure that no applications are writing data to any LUNs mapped from the storage to the hosts.
- Unmount all file systems associated with volumes on the array.

Note: The exact steps to stop host I/O operations depend on the host operating system and the configuration, which are beyond the scope of these instructions. If you are not sure how to stop host I/O operations in your environment, consider shutting down the host.

Attention: Possible data loss - If you continue this procedure while I/O operations are occurring, the host application might lose data because the storage array will not be accessible.

2. Wait for any data in cache memory to be written to the drives.

   The green Cache Active LED on the back of each controller is on when cached data needs to be written to the drives. You must wait for this LED to turn off.



1	Cache Active LED

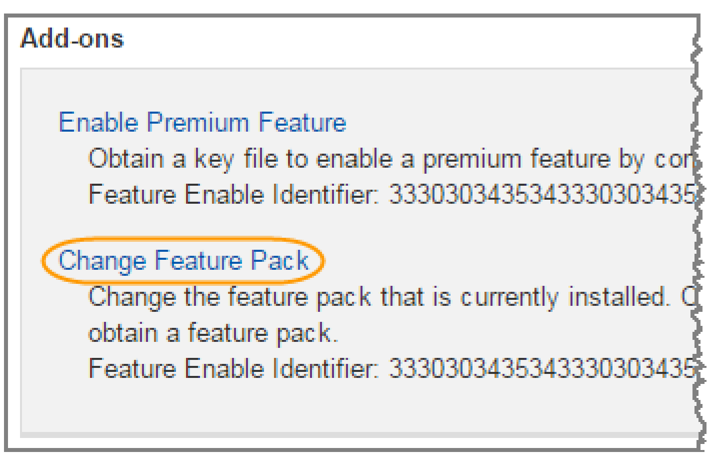
3. From the Home page of SANtricity System Manager, select **View Operations in Progress**.
4. Wait for all operations to complete before continuing with the next step.

Changing the feature pack

You can change the feature pack to convert the host protocol of the baseboard host ports, the IB HIC ports, or both types of ports.

Steps

1. From SANtricity System Manager, select **Settings** > **System**.
2. Under **Add-ons**, select **Change Feature Pack**.



3. Click **Browse**, and then select the feature pack you want to apply.

4. Type **CHANGE** in the field.
5. Click **Change**.

   The feature pack migration begins. Both controllers automatically reboot twice to allow the new feature pack to take effect. The storage array returns to a responsive state after the reboot is complete.

6. Confirm the host ports have the protocol you expect.

   a. From SANtricity System Manager, select **Hardware**.

   b. Click **Show back of shelf**.

   c. Select the graphic for either Controller A or Controller B.

   d. Select **View settings** from the context menu.

   e. Select the **Host Interfaces** tab.

   f. Click **Show more settings**.

   g. Review the details shown for the baseboard ports and the HIC ports (labeled “slot 1”), and confirm that each type of port has the protocol you expect (i.e. "NVMe over Infiniband").

Copyright information

Trademark information

How to send comments about documentation and receive update notifications

You can help us to improve the quality of our documentation by sending us your feedback. You can receive automatic notification when production-level (GA/FCS) documentation is initially released or important changes are made to existing production-level documents.

If you have suggestions for improving this document, send us your comments by email.

doccomments@netapp.com

To help us direct your comments to the correct division, include in the subject line the product name, version, and operating system.

If you want to be notified automatically when production-level documentation is released or important changes are made to existing production-level documents, follow Twitter account @NetAppDoc.

You can also contact us in the following ways:

- NetApp, Inc., 495 East Java Drive, Sunnyvale, CA 94089 U.S.
- Telephone: +1 (408) 822-6000
- Fax: +1 (408) 822-4501
- Support telephone: +1 (888) 463-8277