



Cisco Intersight Configuration Guide for RDMA over Converged Ethernet (RoCE) Version 2

First Published: 2022-07-18

Last Modified: 2023-08-18

Americas Headquarters

Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
<http://www.cisco.com>
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Communications, Services, Bias-free Language, and Additional Information

- To receive timely, relevant information from Cisco, sign up at [Cisco Profile Manager](#).
- To get the business impact you're looking for with the technologies that matter, visit [Cisco Services](#).
- To submit a service request, visit [Cisco Support](#).
- To discover and browse secure, validated enterprise-class apps, products, solutions and services, visit [Cisco Marketplace](#).
- To obtain general networking, training, and certification titles, visit [Cisco Press](#).
- To find warranty information for a specific product or product family, access [Cisco Warranty Finder](#).

Documentation Feedback

To provide feedback about Cisco technical documentation, use the feedback form available in the right pane of every online document.

Cisco Bug Search Tool

[Cisco Bug Search Tool](#) (BST) is a web-based tool that acts as a gateway to the Cisco bug tracking system that maintains a comprehensive list of defects and vulnerabilities in Cisco products and software. BST provides you with detailed defect information about your products and software.

Bias-Free Language

The documentation set for this product strives to use bias-free language. For purposes of this documentation set, bias-free is defined as language that does not imply discrimination based on age, disability, gender, racial identity, ethnic identity, sexual orientation, socioeconomic status, and intersectionality. Exceptions may be present in the documentation due to language that is hardcoded in the user interfaces of the product software, language used based on standards documentation, or language that is used by a referenced third-party product.



CHAPTER 1

RDMA Over Converged Ethernet (RoCE) Version 2

- [RDMA Over Converged Ethernet \(RoCE\) v2, on page 1](#)

RDMA Over Converged Ethernet (RoCE) v2

RDMA over Converged Ethernet version 2 (RoCE v2) is an internet layer protocol, which means that RoCE v2 packets can be routed. RoCE v2 allows direct memory access over the network by encapsulating an Infiniband (IB) transport packet over Ethernet.

The RoCE v2 protocol exists on top of either the UDP/IPv4 or the UDP/IPv6 protocol. The UDP destination port number 4791 has been reserved for RoCE v2. Since RoCE v2 packets are routable, the RoCE v2 protocol is sometimes called Routable RoCE.

RoCE v2 is supported on Windows, Linux, and ESXi platforms.



CHAPTER 2

Configuring SMB Direct with RoCE v2 in Windows

- [Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet \(RoCE\) v2, on page 3](#)
- [Overview of Configuring RoCE v2 Mode 1 and Mode 2 in Windows, on page 5](#)
- [Windows Requirements, on page 5](#)
- [Configuring Mode 1 on Cisco Intersight, on page 6](#)
- [Configuring SMB Direct Mode 1 on the Host System, on page 11](#)
- [Configuring Mode 2 on Cisco Intersight, on page 14](#)
- [Configuring Mode 2 on the Host System, on page 17](#)
- [Deleting the RoCE v2 Interface Using Cisco Intersight, on page 20](#)

Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet (RoCE) v2

General Guidelines and Limitations:

- Cisco Intersight support Microsoft SMB Direct with RoCE v2 on Microsoft Windows Server 2019 and later. Cisco recommends that you have all KB updates from Microsoft for your Windows Server release.



Note

- RoCE v2 is not supported on Microsoft Windows Server 2016.
 - Refer to [Windows Requirements](#) for specific supported Operating System(OS).
-
- Cisco recommends you check [UCS Hardware and Software Compatibility](#) specific to your UCS Manager release to determine support for Microsoft SMB Direct with RoCE v2 on Microsoft Windows.
 - Microsoft SMB Direct with RoCE v2 is supported only with Cisco UCS VIC 1400 Series, VIC 14000, and VIC 15000 Series adapters. It is not supported with UCS VIC 1200 Series and VIC 1300 Series adapters. SMB Direct with RoCE v2 is supported on all UCS Fabric Interconnects.



Note RoCE v1 is not supported on Cisco UCS VIC 1400 Series, VIC 14000 Series, and VIC 15000 series adapters.

- RoCE v2 configuration is supported only between Cisco adapters. Interoperability between Cisco adapters and third party adapters is not supported.
- RoCE v2 supports two RoCE v2 enabled vNIC per adapter and four virtual ports per adapter interface, independent of SET switch configuration.
- RoCE v2 enabled vNIC interfaces must have the no-drop QoS system class enabled in Cisco Intersight.
- The RoCE Properties queue pairs setting must for be a minimum of four queue pairs and maximum number of queue pairs per adapter is 2048.
- The QoS No Drop class configuration must be properly configured on upstream switches such as Cisco Nexus 9000 series switches. QoS configurations will vary between different upstream switches.
- The maximum number of memory regions per rNIC interface is 131072.
- SMB Direct with RoCE v2 is supported on both IPv4 and IPv6.
- RoCE v2 cannot be used on the same vNIC interface as NVGRE, NetFlow, and VMQ features.
- RoCE v2 cannot be used with usNIC.
- RoCE v2 cannot be used with GENEVE offload.

MTU Properties:

- In older versions of the VIC driver, the MTU was derived from either a Cisco Intersight server profile or from the Cisco IMC vNIC MTU setting in standalone mode. This behavior varies for Cisco UCS VIC 1400 Series, VIC 14000 Series, and VIC 15000 Series adapters, where MTU is controlled from the Windows OS Jumbo Packet advanced property.
- The RoCE v2 MTU value is always power-of-two and its maximum limit is 4096.
- RoCE v2 MTU is derived from the Ethernet MTU.
- RoCE v2 MTU is the highest power-of-two that is less than the Ethernet MTU. For example:
 - If the Ethernet value is 1500, then the RoCE v2 MTU value is 1024
 - If the Ethernet value is 4096, then the RoCE v2 MTU value is 4096
 - If the Ethernet value is 9000, then the RoCE v2 MTU value is 4096

Windows NDPKI Modes of Operation:

- Cisco's implementation of Network Direct Kernel Provider Interface (NDPKI) supports two modes of operation: Mode 1 and Mode 2. Mode 1 and 2 relate to the implementation of Network Direct Kernel Provider Interface (NDKPI): Mode 1 is native RDMA, and Mode 2 involves configuration for the virtual port with RDMA. Cisco does not support NDPKI Mode 3 operation.
- The recommended default adapter policy for RoCE v2 Mode 1 is Win-HPN-SMBd.
- The recommended default adapter policy for RoCE v2 Mode 2 is MQ-SMBd.

- RoCE v2 enabled vNICs for Mode 2 operation require the QoS host control policy set to full.
- Mode 2 is inclusive of Mode 1: Mode 1 must be enabled to operate Mode 2.
- On Windows, the RoCE v2 interface supports MSI & MSIx interrupt modes. By default, it is in MSIx interrupt mode. Cisco recommends you avoid changing interrupt mode when the interface is configured with RoCE v2 properties.

Downgrade Limitations:

- Cisco recommends you remove the RoCE v2 configuration before downgrading to any non-supported RoCE v2 release. If the configuration is not removed or disabled, downgrade will fail.

Overview of Configuring RoCE v2 Mode 1 and Mode 2 in Windows

Configuration of RoCE v2 on the Windows platform requires first configuring RoCE v2 Mode 1, then configuring RoCE v2 Mode 2. Mode 1 and Mode 2 relate to the implementation of Network Direct Kernel Provider Interface (NDKPI): Mode 1 is native RDMA, and Mode 2 involves configuration for the virtual port with RDMA.

- To configure RoCE v2 Mode 1, you will:
 - Configure a no-drop class in System QoS policy. Platinum with CoS 5 is a default setting in Cisco Intersight.
 - Configure Mode 1 in Cisco Intersight by creating an Ethernet Adapter policy or using *Win-HPN-SMBd*, the default (pre-defined) configuration in Ethernet Adapter policy.
 - Configure Mode 1 on the host system.
- To configure RoCE v2 Mode 2, RoCE v2 Mode 1 must be configured first and you will:
 - Configure an Ethernet Adapter policy with VMMQ connection or use the *MQ-SMBd* default (pre-defined) configuration in Ethernet Adapter policy for Mode 2 in Cisco Intersight.
 - Configure Mode 2 on the host system.

Windows Requirements

Configuration and use of RDMA over Converged Ethernet for RoCE v2 in Windows Server requires the following:

- VIC Driver version 5.4.0.x or later
- Cisco UCS M5 B-Series and C-Series with Cisco UCS 1400 Series adapters.
- Cisco UCS M6 B-Series, C-Series, or X-Series servers with Cisco UCS VIC 1400, VIC 14000, or VIC 15000 series adapters.

- Cisco UCS M7 C-Series, or X-Series servers with Cisco UCS VIC 1400, VIC 14000, or VIC 15000 series adapters.



Note All Powershell commands or advanced property configurations are common across Windows 2019 and 2022 unless explicitly mentioned.

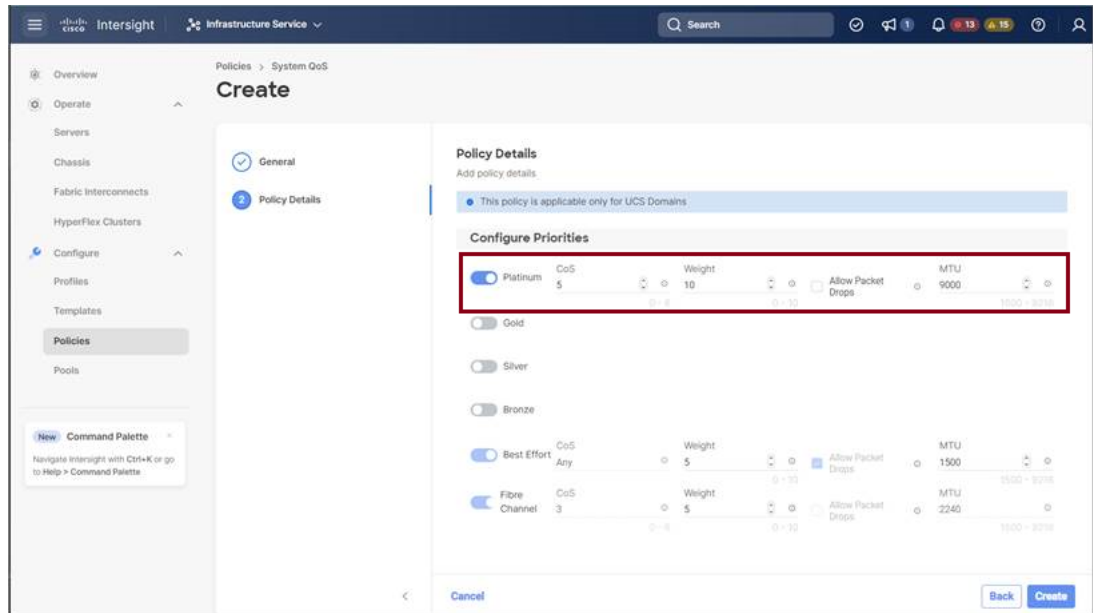
Configuring Mode 1 on Cisco Intersight

Use these steps to configure the RoCE v2 Mode 1 interface on Cisco Intersight.

To avoid possible RDMA packet drops, ensure same no-drop COS is configured across the network. The following steps allows you to configure a no-drop class in System QoS policies and use it for RDMA supported interfaces.

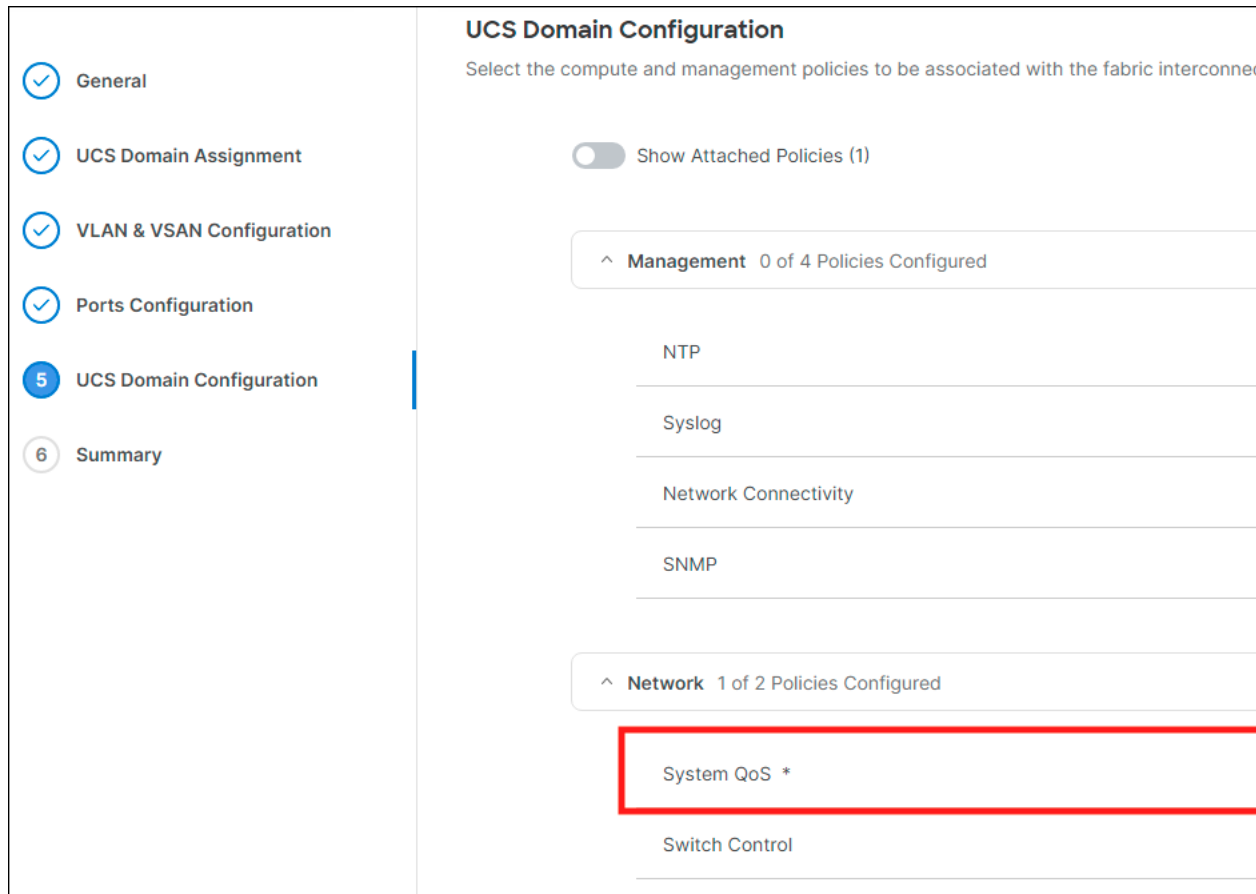
Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Domain** platform type, search or choose **System QoS**, and click **Start**.
- Step 2** In the **General** page, enter the policy name and click **Next**, and then in the **Policy Details** page, configure the property setting for System QoS policy as follows:
- For **Priority**, choose **Platinum**
 - For **Allow Packet Drops**, uncheck the check box.
- Note** For more information on MTU field, see *MTU Properties* in [Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet \(RoCE\) v2](#), on page 3



Step 3 Click **Create**

Step 4 Associate the System QoS policy to the Domain Profile.



Note For more information, see *Creating System QoS Policy* in [Configuring Domain Policies](#) and [Configuring Domain Profiles](#).

The System QoS Policy is successfully created and deployed to the Domain Profile.

What to do next

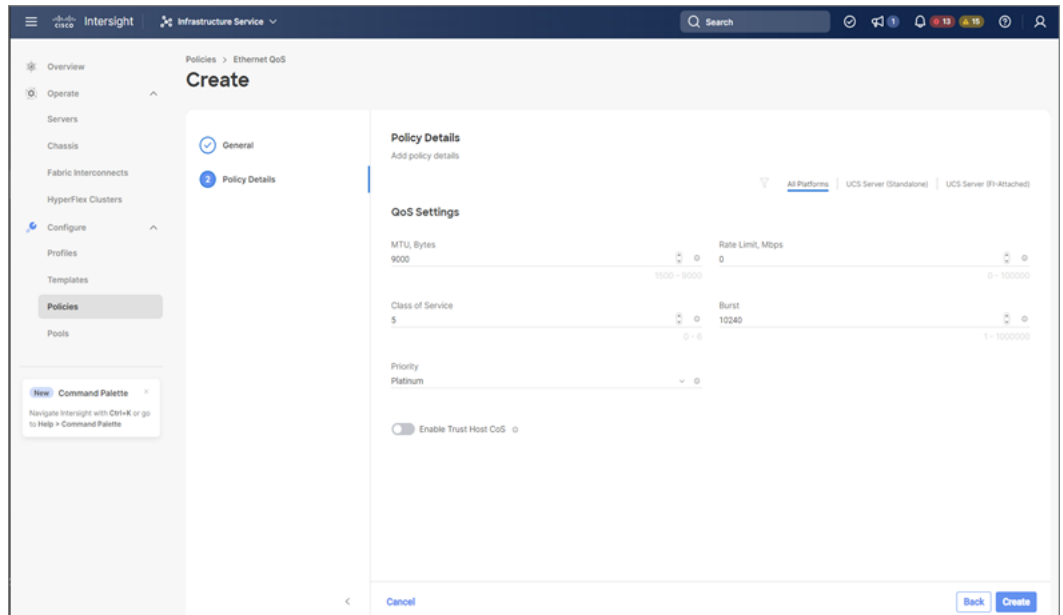
Configure the server profile with RoCE v2 vNIC settings in LAN Connectivity policy.

Enabling RoCE Settings in LAN Connectivity Policy

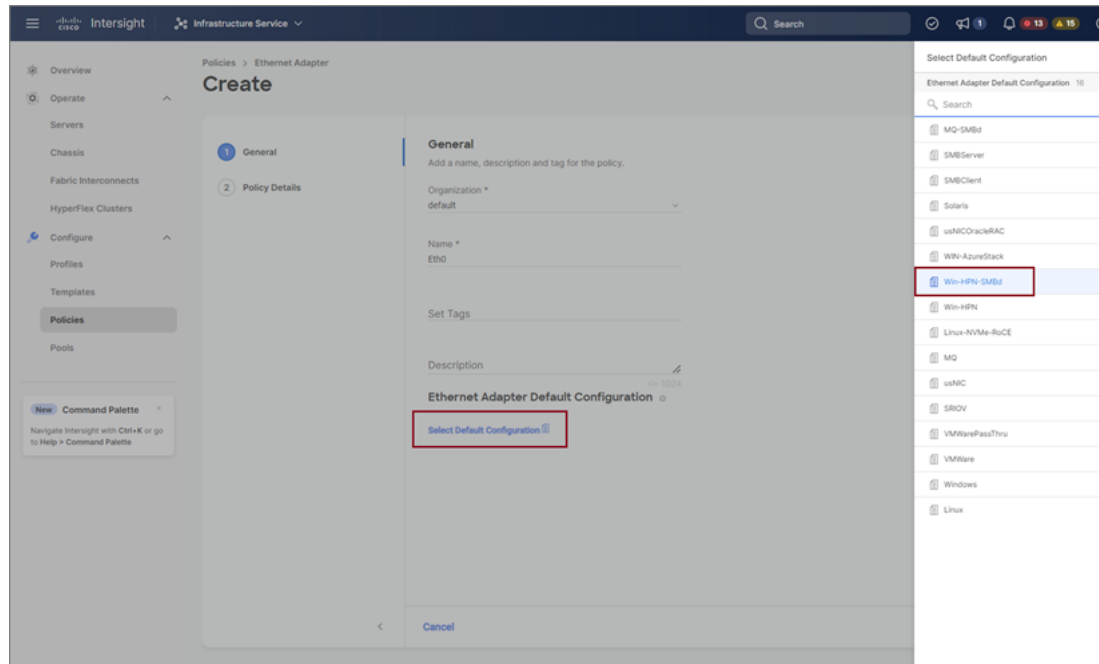
Use these steps to configure the RoCE v2 vNIC settings in Mode 1. In Cisco Intersight LAN Connectivity policy, you can enable the RoCE settings on **Ethernet QoS** policy and **Ethernet Adapter** policy for Mode 1 configuration as follows:

Procedure

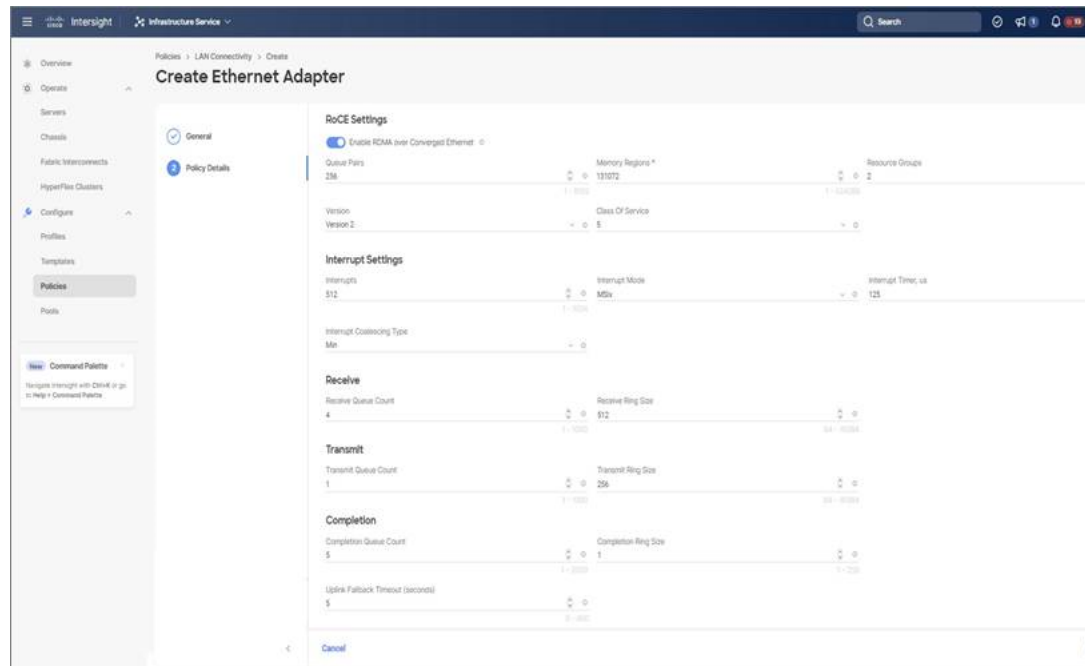
- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Server** platform type, search or choose **LAN Connectivity** policy, and click **Start**.
- Step 2** In the policy **General** page, enter the policy name, select the Target Platform as **UCS Server (Standalone)** or **UCS Server (FI-Attached)**, and click **Next**.
- Step 3** In the **Policy Details** page, click **Add vNIC** to create a new vNIC.
- Step 4** In the **Add vNIC** page, follow the configuration parameters to enable the RoCE vNIC settings:
- In the **General** section, provide a name for virtual ethernet interface.
 - In the **Consistent Device Naming (CDN)** section of the Standalone server or the **Failover** section of FI-attached server, do the following:
 - Click **Select Policy** link below the **Ethernet QoS**. Use the **Create New** button to create a new Ethernet QoS policy with the following property settings:
 - For **MTU**, choose or enter **1500, 4096, or 9000**
 - For **Priority**, choose **Platinum** or **any no-drop**
 - For **Class of Service**, choose or enter **5**



- Click **Select Policy** link below the **Ethernet Adapter**. Follow one of the options to select a default policy or create an Ethernet Adapter policy:
 - **Use the Default Configuration:** Click **Create New** to create a new policy. In the **General** page, enter the name of the policy and click **Select Default Configuration** to search and select **Win-HPN-SMBd**, the pre-defined Ethernet Adapter Default Configuration. Click **Next** and then **Create**.



- **Configure RoCE Settings in the policy:** Click **Create New** to create a new policy. In the **General** page, enter the name of the policy. In the **Policy Details** page, use the following property settings, click **Next**, and then **Create**.
 - For **Enable RDMA over Converged Ethernet**, slide to enable.
 - For **Queue Pairs**, choose or enter **256**
 - For **Memory Regions**, choose or enter **131072**
 - For **Resource Groups**, choose or enter **2**
 - For **Version**, select **Version 2**



- Click **Add** to add and save the new vNIC settings.

Note All the fields with * are mandatory and ensure it is filled out or selected with appropriate policies.

Step 5 Click **Create** to complete the LAN Connectivity policy with RoCE v2 property settings.

Step 6 Associate the LAN Connectivity policy to the server profile.

Note For more information, see *Creating a LAN Connectivity Policy*, *Creating an Ethernet QoS Policy*, and *Creating an Ethernet Adapter Policy* in [Configuring UCS Server Policies](#) and [Configuring UCS Server Profiles](#).

The LAN Connectivity policy with the Ethernet QoS policy and Ethernet Adapter policy vNIC setting is successfully created and the server profile is deployed to enable RoCE v2 configuration.

What to do next

Once the policy configuration for RoCE v2 is complete, reboot the server, and proceed with the RoCE v2 Mode 1 configuration of the host.

Configuring SMB Direct Mode 1 on the Host System

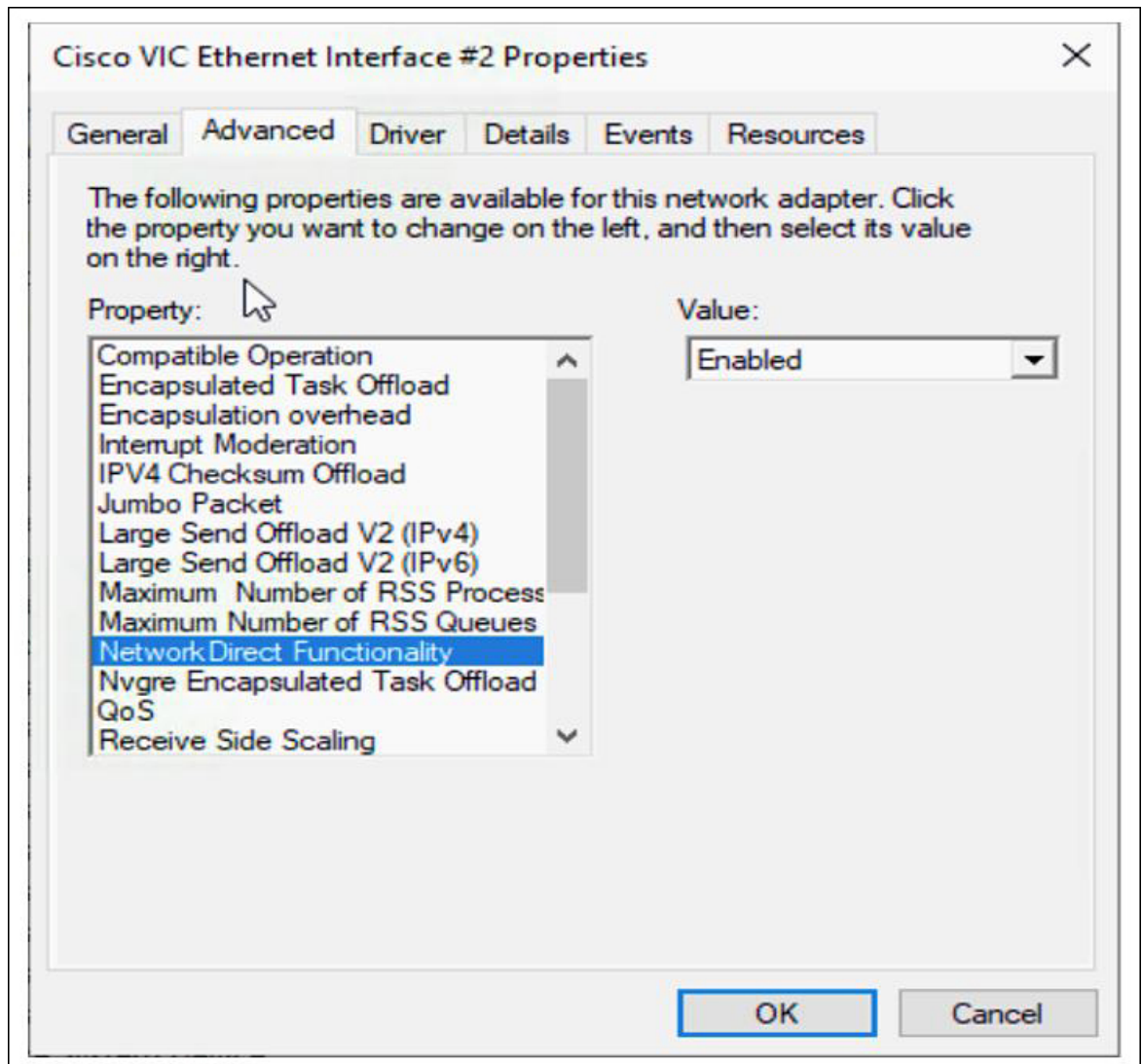
You will configure connection between smb-client and smb-server on two host interfaces. For each of these servers, smb-client and smb-server, configure the RoCE v2-enabled vNIC as described below.

Before you begin

Configure RoCE v2 for Mode 1 in Cisco Intersight.

Procedure

-
- Step 1** In the Windows host, go to the Device Manager and select the appropriate Cisco VIC Internet Interface.
- Step 2** Go to **Tools > Computer Management > Device Manager > Network Adapter** > click on **VIC Network Adapter > Properties > Advanced > Network Direct Functionality**. Perform this operation for both the smb-server and smb-client vNICs.



Step 3 Verify that RoCE is enabled on the host operating system using PowerShell.

The `Get-NetOffloadGlobalSetting` command shows NetworkDirect is enabled.

```
PS C:\Users\Administrator> Get-NetOffloadGlobalSetting
```

```
ReceiveSideScaling      : Enabled
ReceiveSegmentCoalescing : Enabled
Chimney                 : Disabled
TaskOffload             : Enabled
NetworkDirect           : Enabled
NetworkDirectAcrossIPSubnets : Blocked
PacketCoalescingFilter  : Disabled
```

Note If the NetworkDirect setting is showing as disabled, enable it using the command:

```
Set-NetOffloadGlobalSetting -NetworkDirect enabled
```

Step 4 Bring up Powershell and enter the command:

```
get-SmbClientNetworkInterface
```



```
PS C:\Users\Administrator>
PS C:\Users\Administrator> Get-SmbClientNetworkInterface
```

Interface	Index	RSS Capable	RDMA Capable	Speed	IpAddresses	Friendly Name
14		True	False	40 Gbps	{10.37.60.162}	vEthernet (vswitch)
26		True	True	40 Gbps	{10.37.60.158}	vEthernet (vpl)
9		True	True	40 Gbps	{50.37.61.23}	Ethernet 2
5		False	False	40 Gbps	{169.254.10.5}	Ethernet (Kernel Debugger)
8		True	False	40 Gbps	{169.254.4.26}	Ethernet 3

```
PS C:\Users\Administrator>
```

Step 5 Enter `enable - netadapterrdma [-name] ["Ethernetname"]`

Step 6 Verify the overall RoCE v2 Mode 1 configuration at the Host as follows:

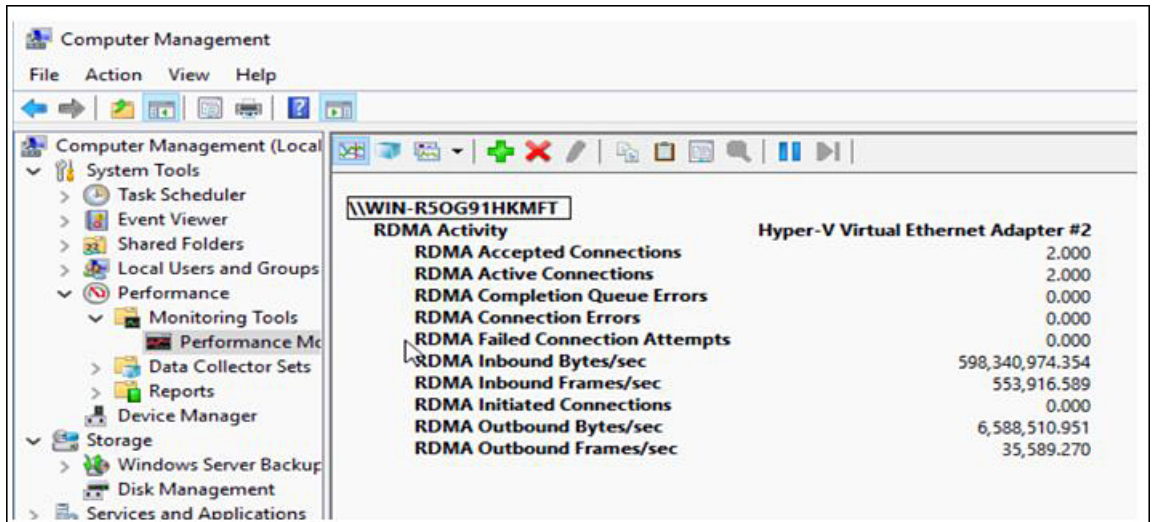
- a) Use the Powershell command `netstat -xan` to verify the listeners in both the smb-client and smb-server Windows host; listeners will be shown in the command output.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan
```

Mode	IfIndex	Type	Local Address	Foreign Address	PID
Kernel	9	Listener	50.37.61.23:445	NA	0
Kernel	26	Listener	10.37.60.158:445	NA	0

```
PS C:\Users\Administrator>
```

- b) Go to the smb-client server fileshare and start an I/O operation.
- c) Go to the performance monitor and check that it displays the RDMA activity.



Step 7 In the Powershell command window, check the connection entries with the `netstat -xan` output command to make sure they are displayed. You can also run `netstat -xan` from the command prompt. If the connection entry shows up in netstat-xan output, the RoCE v2 model1 connections are correctly established between client and server.

```
PS C:\Users\Administrator> netstat -xan
Active NetworkDirect Connections, Listeners, SharedEndpoints
Mode    IfIndex Type           Local Address      Foreign Address    PID
-----  -
Kernel  4    Connection    50.37.61.22:445   50.37.61.71:2240   0
Kernel  4    Connection    50.37.61.22:445   50.37.61.71:2496   0
Kernel  11   Connection    50.37.61.122:445  50.37.61.71:2752   0
Kernel  11   Connection    50.37.61.122:445  50.37.61.71:3008   0
Kernel  32   Connection    10.37.60.155:445  50.37.60.61:49092  0
Kernel  32   Connection    10.37.60.155:445  50.37.60.61:49348  0
Kernel  26   Connection    50.37.60.32:445   50.37.60.61:48580  0
Kernel  26   Connection    50.37.60.32:445   50.37.60.61:48836  0
Kernel  4    Listener      50.37.61.22:445   NA                  0
Kernel  11   Listener      50.37.61.122:445  NA                  0
Kernel  32   Listener      10.37.60.155:445  NA                  0
Kernel  26   Listener      50.37.60.32:445   NA                  0
```

Note IP values are representative only.

Step 8 By default, Microsoft's SMB Direct establishes two RDMA connections per RDMA interface. You can change the number of RDMA connections per RDMA interface to one or any number of connections.

For example, to increase the number of RDMA connections to 4, type the following command in PowerShell:

```
PS C:\Users\Administrator> Set-ItemProperty -Path `
"HKLM:\SYSTEM\CurrentControlSet\Services\LanmanWorkstation\Parameters"
ConnectionCountPerRdmaNetworkInterface -Type DWORD -Value 4 -Force
```

Configuring Mode 2 on Cisco Intersight

Use these steps to configure the RoCE v2 policies in Mode 2. In Cisco Intersight LAN Connectivity Policy, you can enable the RoCE settings on **Ethernet QoS** policy and **Ethernet Adapter** policy, and **VMMQ Adapter** policy for Mode 2 configuration as follows:

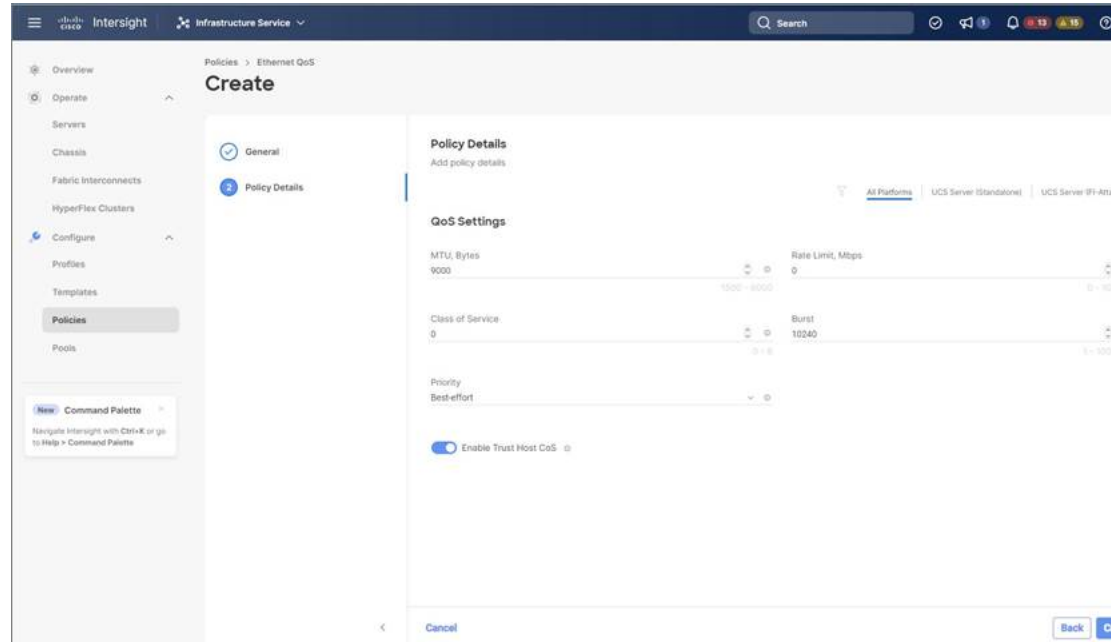
Before you begin

Configure RoCE v2 Policies in Mode 1.

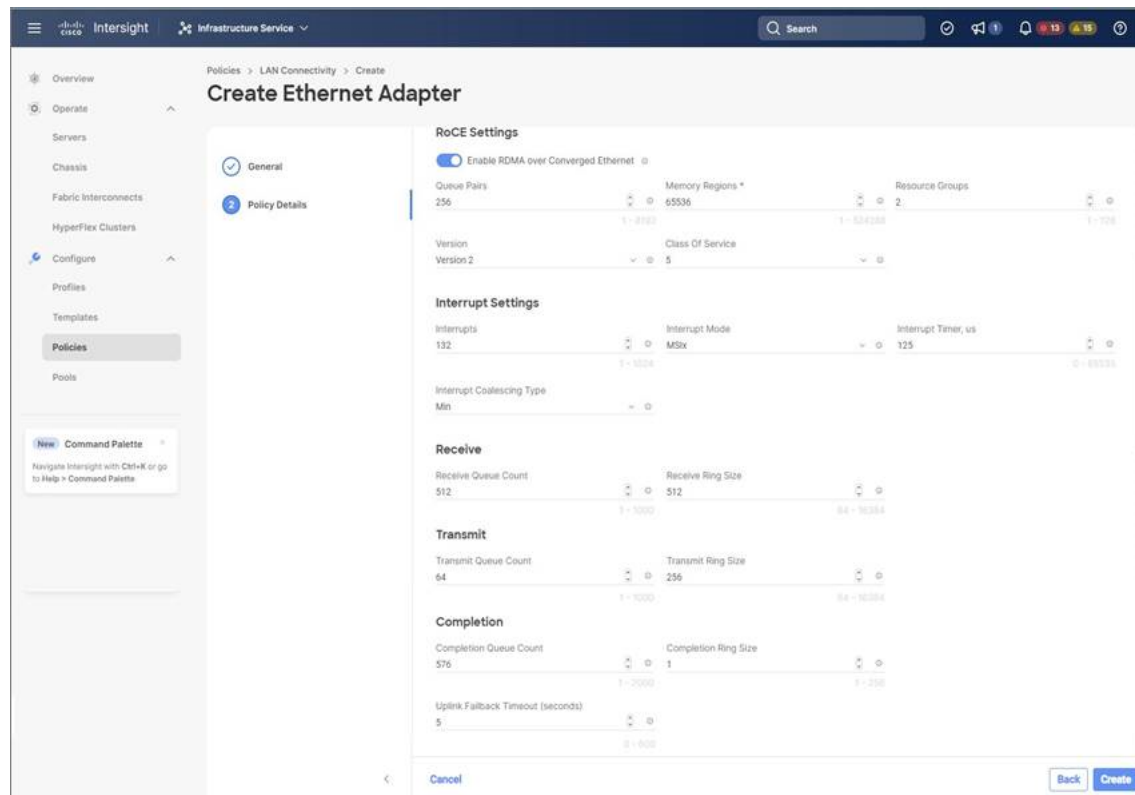
Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Server** platform type, search or choose **LAN Connectivity** policy, and click **Start**.
- Step 2** In the policy **General** page, enter the policy name, select the Target Platform as **UCS Server (Standalone)** or **UCS Server (FI-Attached)**, and click **Next**.
- Step 3** In the **Policy Details** page, click **Add vNIC** to create a new vNIC.
- Step 4** In the **Add vNIC** page, follow the configuration parameters to enable the RoCE vNIC settings:
 - a) In the **General** section, provide a name for virtual ethernet interface.
 - b) In the **Consistent Device Naming (CDN)** section of the Standalone server or the **Failover** section of FI-attached server, do the following:
 - Click **Select Policy** link below the **Ethernet QoS**. Use the **Create New** button to create a new Ethernet QoS policy with the following property settings:

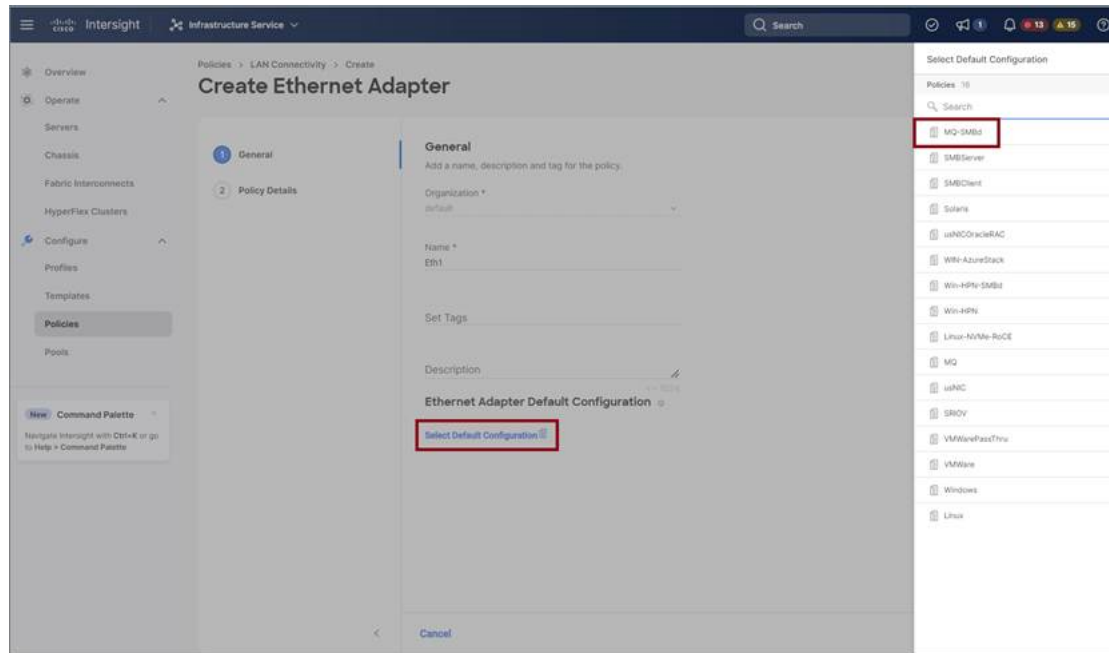
- For **MTU**, choose or enter **1500, 4096, or 9000**
- For **Priority**, choose or enter **Best-effort**
- **Enable Trust Host CoS**, slide to enable



- Click **Select Policy** link below the **Ethernet Adapter**. Use **Create New** button to create a new Ethernet Adapter policy with the following property settings:
 - For **Enable RDMA over Converged Ethernet**, slide to enable.
 - For **Queue Pairs**, select or enter **256**
 - For **Memory Regions**, select or enter **65536**
 - For **Resource Groups**, select or enter **2**
 - For **Version**, choose **Version 2**
 - For **Class of Service**, choose or enter **5**



- In the **Connection** section, use the following property setting for VMQ Connection and to create VMMQ Adapter policy:
 - For connection, select **VMQ**.
 - For **Enable Virtual Machine Multi-Queue**, slide to enable.
 - For **Number of Sub vNICs**, select or enter **4**
 - For **VMMQ Adapter Policy**, click **Select Policy** link below the VMMQ Adapter Policy and do the following:
 - Click **Create New** to create a new policy. In the **General** page, enter the name of the policy and click **Select Default Configuration** to search and select **MQ-SMBd**, the pre-defined VMMQ Adapter Default Configuration.
 - Click **Next** and then **Create**.



- Click **Add** to add and save the new vNIC settings.

Note All the fields with * are mandatory and ensure it is filled out or selected with appropriate policies.

Step 5 Click **Create** to complete the LAN Connectivity policy with RoCE v2 property settings.

Step 6 Associate the LAN Connectivity policy to the server profile.

Note For more information on *Creating an Ethernet QoS, Ethernet Adapter Policy, and VMMQ Adapter Policy*, see [Configuring UCS Server Policies](#) and [Configuring UCS Server Profiles](#).

The LAN Connectivity Policy with Ethernet QoS Policy, Ethernet Adapter Policy, and VMMQ Adapter Policy are successfully created and deployed to enable RoCE v2 configuration.

What to do next

Once the policy configuration for RoCE v2 is complete, reboot the server and proceed with the RoCE v2 Mode 2 configuration of the host.

Configuring Mode 2 on the Host System

This task uses Hyper-V virtualization software that is compatible with Windows Server 2019 and Windows Server 2022.

Before you begin

- Configure and confirm the connection for Mode 1 for both Cisco Intersight and Host.
- Configure Mode 2 in Cisco Intersight.

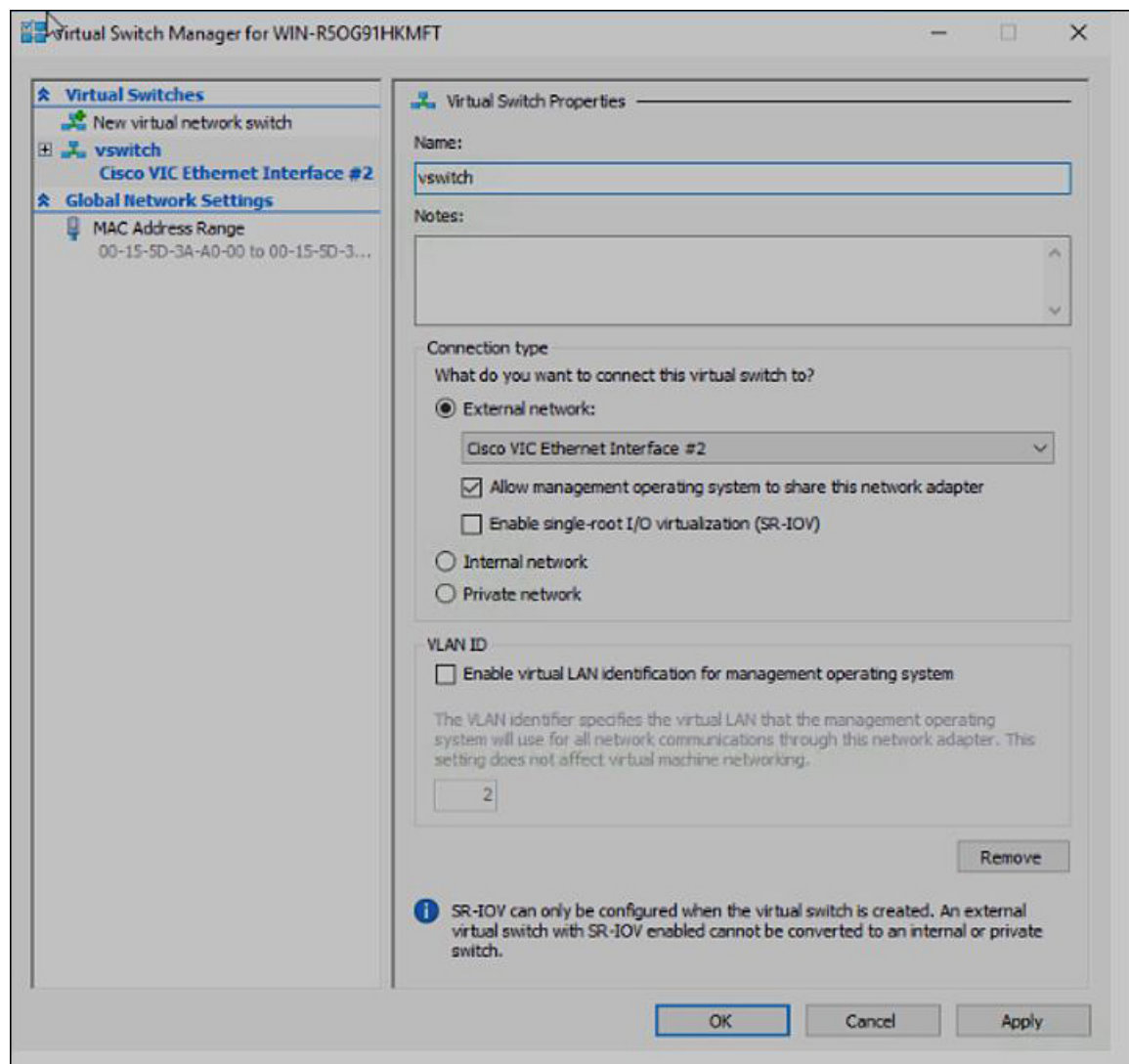
Procedure**Step 1**

Go the Hyper-V switch manager.

Step 2

Create a new Virtual Network Switch (vswitch) for the RoCE v2-enabled Ethernet interface.

- Choose **External Network** and select **VIC Ethernet Interface 2** and **Allow management operating system to share this network adapter**.
- Click **OK** to create the virtual switch.



Bring up the Powershell interface.

Step 3 Configure the non-default vport and enable RDMA with the following Powershell commands:

```
add-vmNetworkAdapter -switchname vswitch -name vp1 -managementOS
enable-netAdapterRdma -name "vEthernet (vp1)"
```

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> add-vmNetworkAdapter -switchName vswitch -name vp1 -managementOS
PS C:\Users\Administrator> enable-netAdapterRdma -name "vEthernet (vp1)"
PS C:\Users\Administrator>
```

a) Configure set-switch using the following Powershell command.

```
new-vmSwitch -name setswitch -netAdapterName "Ethernet x" -enableEmbeddedTeam $true
```

This creates the switch. Use the following to display the interfaces:

```
get-netadapterrdma
add-vmNetworkAdapter -switchname setswitch -name svp1
```

You will see the new vport when you again enter

```
get-netadapterrdma
```

b) Add a vport.

```
add-vmNetworkAdapter -switchname setswitch -name svp1
```

You will see the new vport when you again enter

```
get-netadapterrdma
```

c) Enable the RDMA on the vport:

```
enable-netAdapterRdma -name "vEthernet (svp1)"
```

Step 4 Configure the IPV4 addresses on the RDMA enabled vport in both servers.

Step 5 Create a share in smb-server and map the share in the smb-client.

- For smb-client and smb-server in the host system, configure the RoCE v2-enabled vNIC as described above.
- Configure the IPV4 addresses of the primary fabric and sub-vNICs in both servers, using the same IP subnet and same unique vlan for both.
- Create a share in smb-server and map the share in the smb-client.

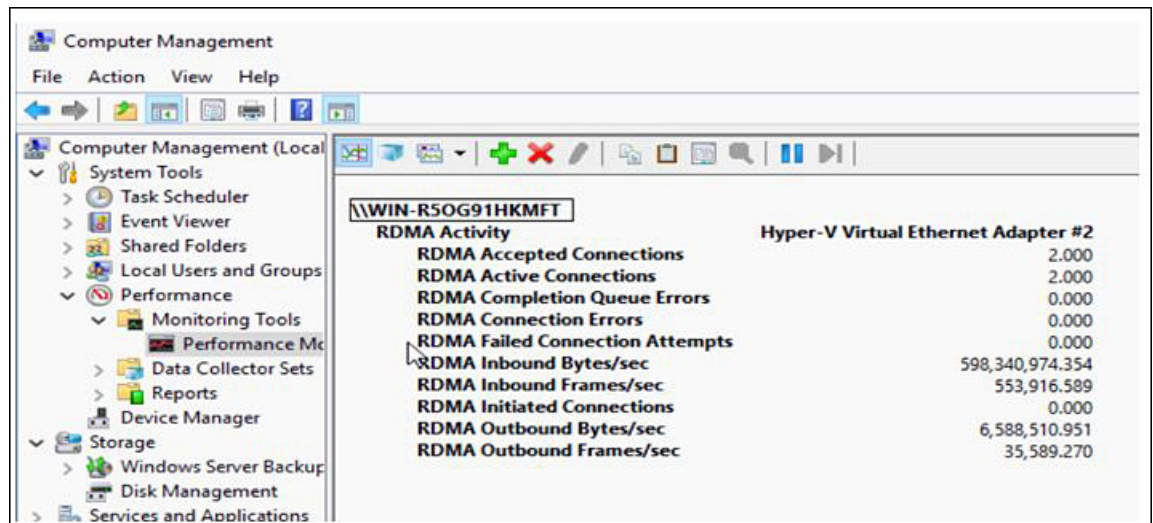
Step 6 Verify the Mode 2 configuration.

a) Use the Powershell command *netstat -xan* to display listeners and their associated IP addresses.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints
Mode    IfIndex Type           Local Address           Foreign Address         PID
-----
Kernel   9 Listener      50.37.61.23:445        NA                       0
Kernel  26 Listener      10.37.60.158:445       NA                       0
PS C:\Users\Administrator>
```

b) Start any RDMA I/O in the file share in smb-client.



- c) Issue the `netstat -xan` command again and check for the connection entries to verify they are displayed.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> netstat -xan

Active NetworkDirect Connections, Listeners, SharedEndpoints

Mode     IfIndex Type                Local Address        Foreign Address      PID
-----
Kernel   9      Connection         50.37.61.23:192     50.37.61.184:445    0
Kernel   9      Connection         50.37.61.23:448     50.37.61.184:445    0
Kernel   9      Connection         50.37.61.23:704     50.37.61.214:445    0
Kernel   9      Connection         50.37.61.23:960     50.37.61.214:445    0
Kernel   9      Connection         50.37.61.23:1216    50.37.61.224:445    0
Kernel   9      Connection         50.37.61.23:1472    50.37.61.224:445    0
Kernel   9      Connection         50.37.61.23:1728    50.37.61.234:445    0
Kernel   9      Connection         50.37.61.23:1984    50.37.61.234:445    0
Kernel   9      Listener           50.37.61.23:445     NA                   0
Kernel   26     Listener           10.37.60.158:445    NA                   0
PS C:\Users\Administrator>
```

What to do next

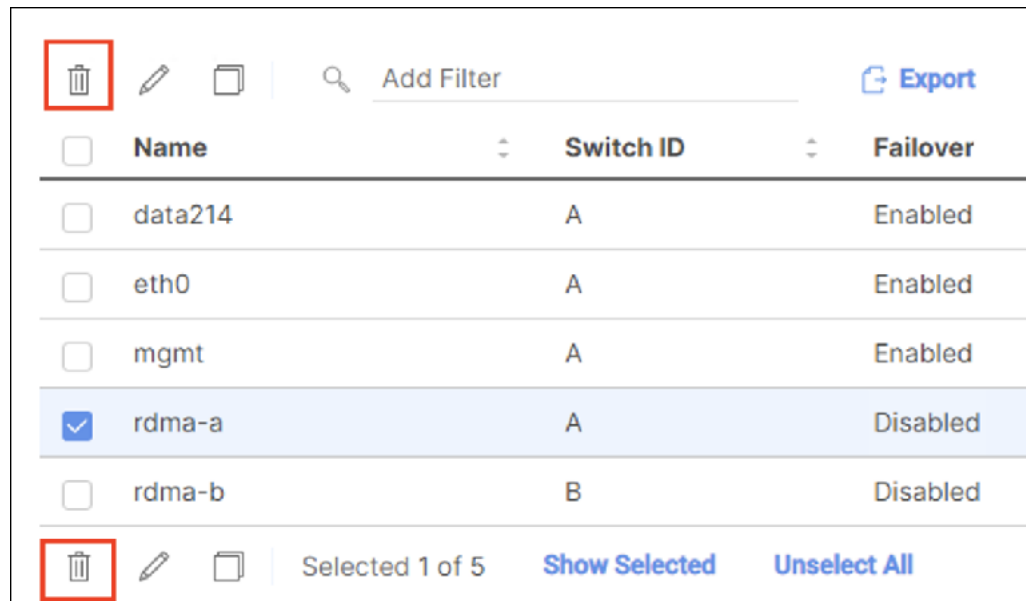
Troubleshoot any items if necessary.

Deleting the RoCE v2 Interface Using Cisco Intersight

Use these steps to remove the RoCE v2 interface.

Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. In the **Add Filter** field, select **Type: LAN Connectivity**.
- Step 2** Select the appropriate LAN Connectivity policy created for RoCE V2 configuration and use the delete icon on the top or bottom of the policy list.
- Step 3** Click **Delete** to delete the policy.



<input type="checkbox"/>	Name	Switch ID	Failover
<input type="checkbox"/>	data214	A	Enabled
<input type="checkbox"/>	eth0	A	Enabled
<input type="checkbox"/>	mgmt	A	Enabled
<input checked="" type="checkbox"/>	rdma-a	A	Disabled
<input type="checkbox"/>	rdma-b	B	Disabled

Selected 1 of 5 Show Selected Unselect All

Step 4 Upon deleting the RoCE v2 configuration, re-deploy the server profile and reboot the server.



CHAPTER 3

Configuring NVMeoF with RoCEv2 in Linux

- [Guidelines for using NVMe over Fabrics \(NVMeoF\) with RoCE v2 on Linux, on page 23](#)
- [Linux Requirements, on page 24](#)
- [Configuring RoCE v2 for NVMeoF on Cisco Intersight, on page 24](#)
- [Configuring RoCE v2 for NVMeoF on the Host System, on page 29](#)
- [Setting Up Device Mapper Multipath, on page 32](#)
- [Deleting the RoCE v2 Interface Using Cisco Intersight, on page 33](#)

Guidelines for using NVMe over Fabrics (NVMeoF) with RoCE v2 on Linux

General Guidelines and Limitations:

- Cisco recommends you check [UCS Hardware and Software Compatibility](#) to determine support for NVMeoF. NVMeoF is supported on Cisco UCS B-Series, C-Series, and X-Series servers.
- NVMe over RDMA with RoCE v2 is supported with the Cisco UCS VIC 1400, VIC 14000, and VIC 15000 Series adapters.
- When creating RoCE v2 interfaces, use Cisco Intersight provided Linux-NVMe-RoCE adapter policy.
- In the Ethernet Adapter policy, do not change values of Queue Pairs, Memory Regions, Resource Groups, and Priority settings other than to Cisco provided default values. NVMeoF functionality may not be guaranteed with different settings for Queue Pairs, Memory Regions, Resource Groups, and Priority.
- When configuring RoCE v2 interfaces, use both the `enic` and `enic_rdma` binary drivers downloaded from Cisco.com and install the matched set of `enic` and `enic_rdma` drivers. Attempting to use the binary `enic_rdma` driver downloaded from Cisco.com with an `inbox` `enic` driver will not work.
- RoCE v2 supports maximum two RoCE v2 enabled interfaces per adapter.
- Booting from an NVMeoF namespace is not supported.
- Layer 3 routing is not supported.
- RoCE v2 does not support bonding.
- Saving a crashdump to an NVMeoF namespace during a system crash is not supported.

- NVMeoF cannot be used with usNIC, VxLAN, VMQ, VMMQ, NVGRE, GENEVE Offload, and DPDK features.
- Cisco Intersight does not support fabric failover for vNICs with RoCE v2 enabled.
- The Quality of Service (QoS) no drop class configuration must be properly configured on upstream switches such as Cisco Nexus 9000 series switches. QoS configurations will vary between different upstream switches.
- Spanning Tree Protocol (STP) may cause temporary loss of network connectivity when a failover or failback event occurs. To prevent this issue from occurring, disable STP on uplink switches.

Linux Requirements

Configuration and use of RoCE v2 in Linux requires the following:

- InfiniBand kernel API module `ib_core`
- A storage array that supports NVMeoF connection

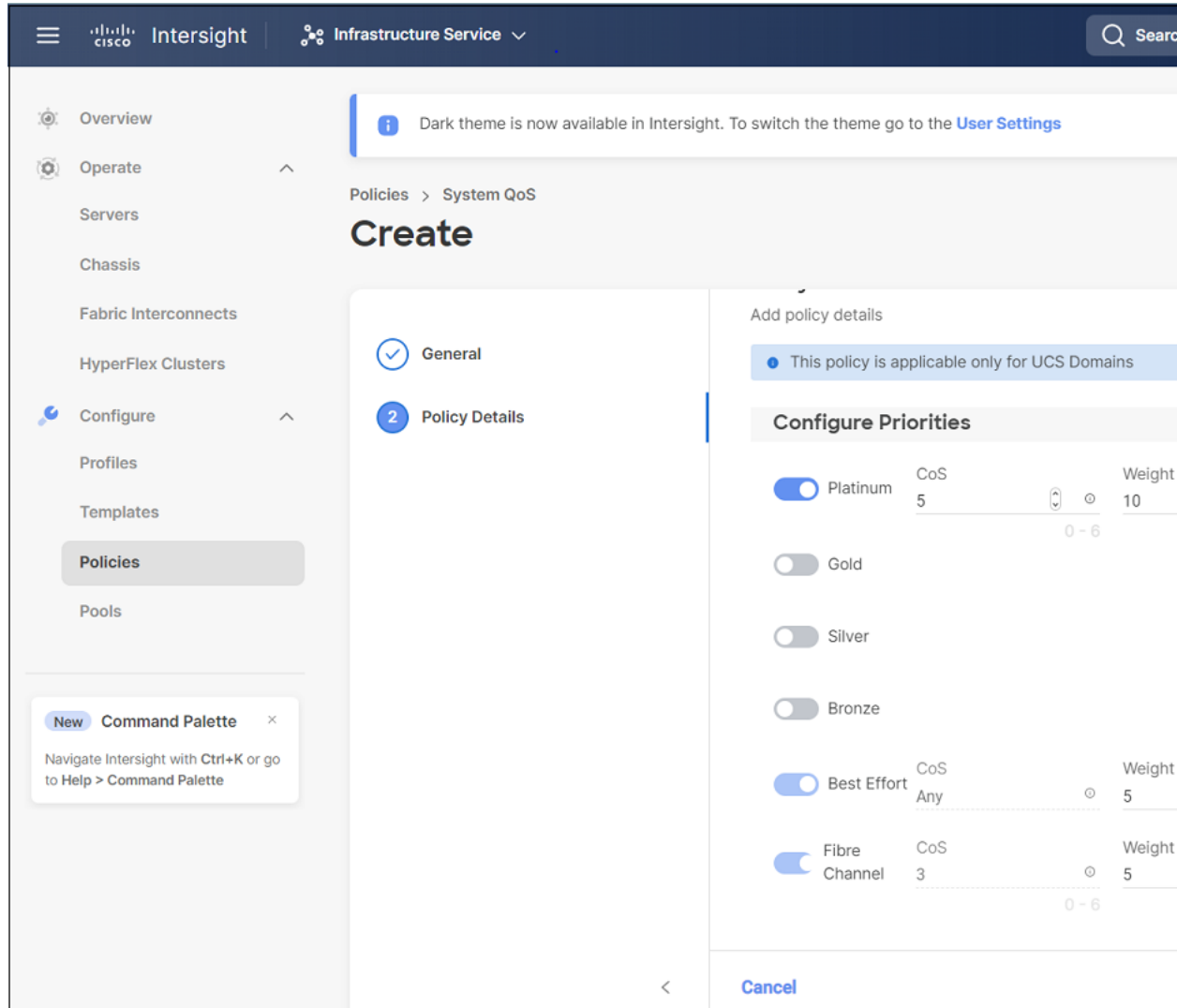
Configuring RoCE v2 for NVMeoF on Cisco Intersight

Use these steps to configure the RoCE v2 interface on Cisco Intersight.

To avoid possible RDMA packet drops, ensure same no-drop COS is configured across the network. The following steps allow you to configure a no-drop class in System QoS policies and use it for RDMA supported interfaces.

Procedure

-
- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Domain** platform type, search or choose **System QoS**, and click **Start**.
- Step 2** In the **General** page, enter the policy name and click **Next**, and then in the **Policy Details** page, configure the property setting for System QoS policy as follows:
- For **Priority**, choose **Platinum**
 - For **Allow Packet Drops**, uncheck the check box.
 - For **MTU**, set the value as **9216**.



Step 3 Click **Create**.

Step 4 Associate the System QoS policy to the Domain Profile.

UCS Domain Configuration

Select the compute and management policies to be associated with the fabric interconnect.

Show Attached Policies (1)

^ **Management** 0 of 4 Policies Configured

NTP

Syslog

Network Connectivity

SNMP

^ **Network** 1 of 2 Policies Configured

System QoS *

Switch Control

Note For more information, see *Creating System QoS Policy* in [Configuring Domain Policies](#) and [Configuring Domain Profiles](#).

The System QoS Policy is successfully created and deployed to the Domain Profile.

What to do next

Configure the server profile with RoCE v2 vNIC settings in LAN Connectivity policy.

Enabling RoCE Settings in LAN Connectivity Policy

Use the following steps to configure the RoCE v2 vNIC. In Cisco Intersight LAN Connectivity policy, you can enable the RoCE settings on **Ethernet Adapter policy** for Linux configuration as follows:

Procedure

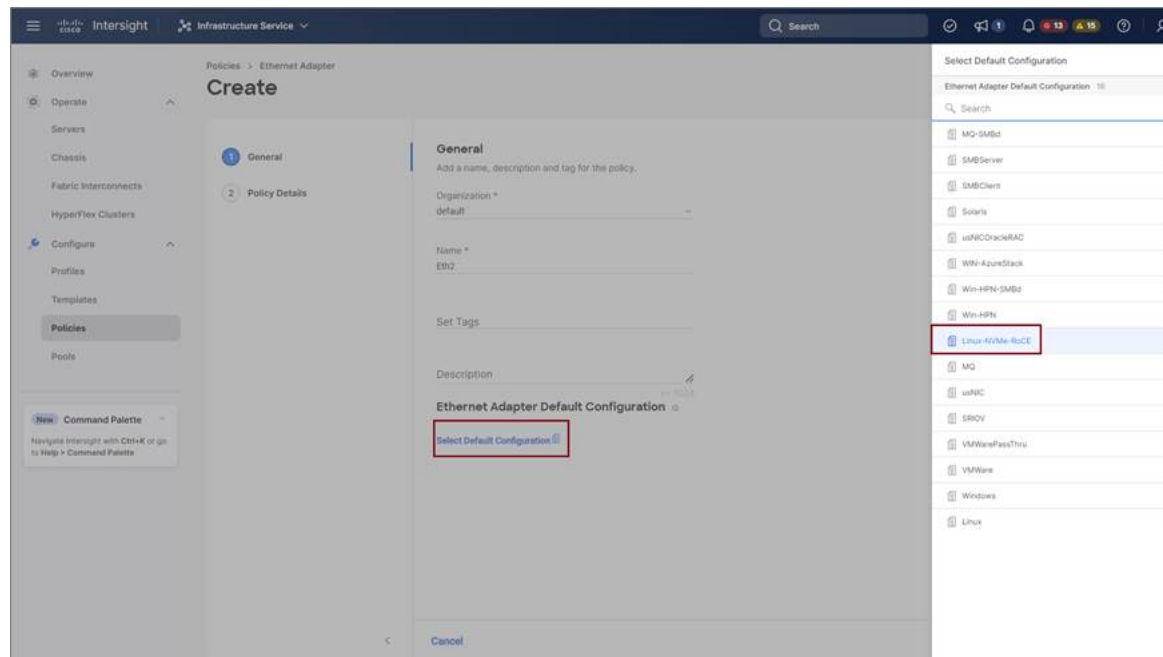
- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Server** platform type, search or choose **LAN Connectivity policy**, and click **Start**.

Step 2 In the policy **General** page, enter the policy name, select the Target Platform as **UCS Server (Standalone)** or **UCS Server (FI-Attached)**, and click **Next**.

Step 3 In the **Policy Details** page, click **Add vNIC** to create a new vNIC.

Step 4 In the **Add vNIC** page, follow the configuration parameters to enable the RoCE v2 vNIC:

- a) In the **General** section, provide a name for virtual ethernet interface.
- b) In the **Consistent Device Naming (CDN)** section of the Standalone server or the **Failover** section of FI-attached server, do the following:
 - Click **Select Policy** under **Ethernet Adapter**.
 - In the **Select Policy** window, click **Create New** to create an Ethernet Adapter policy.
 - On the **General** page, enter the policy name and click **Select Default Configuration**. Search and select **Linux-NVMe-RoCE** in the Default Configuration window and click **Next**.
 - On the **Policy Details**, verify the default configuration parameters for RoCE and click **Create**.



- Click **Add** to save the setting and add the new vNIC.

Note All the fields with * are mandatory and ensure it is filled out or selected with appropriate policies.

Step 5 Click **Create** to complete the LAN Connectivity policy with RoCE v2 settings.

Step 6 Associate the LAN Connectivity policy to the Server Profile.

Note For more information, see *Creating a LAN Connectivity Policy* and *Creating an Ethernet Adapter Policy* in [Configuring UCS Server Policies](#) and [Configuring UCS Server Profiles](#).

The LAN Connectivity Policy with the Ethernet Adapter policy vNIC setting is successfully created and deployed to enable RoCE v2 configuration.

What to do next

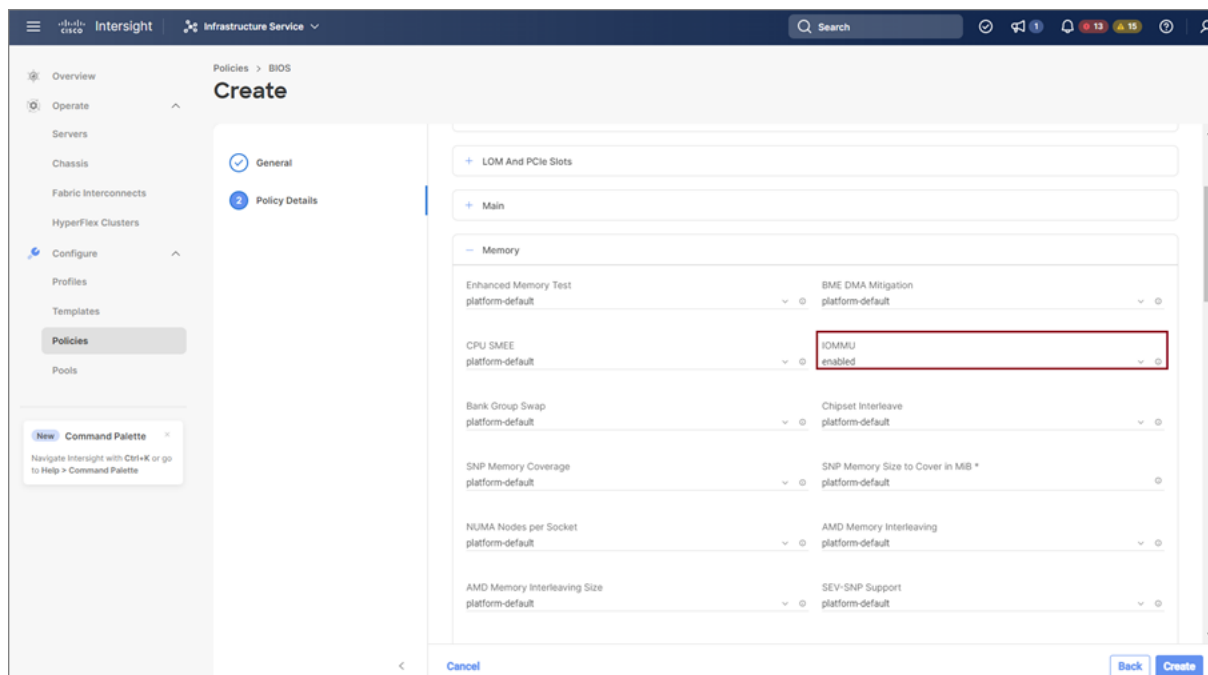
Once the policy configuration for RoCE v2 is complete, proceed to enable IOMMU in the BIOS policy.

Enabling an IOMMU BIOS Settings

Use the following steps to configure the server profile with the RoCE v2 vNIC and enable the IOMMU BIOS policy before enabling the IOMMU in the Linux kernel.

Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Server** platform type, search or choose **BIOS**, and click **Start**.
- Step 2** On the **General** page, enter the policy name and click **Next**.
- Step 3** On the **Policy Details** page, configure the following BIOS:
 - a) Select **All Platforms**.
 - b) Expand the **Memory** group.
 - c) In the **IOMMU** drop-down list, select the BIOS value **enabled** for setting IOMMU configuration.



- Step 4** Click **Create**.
- Step 5** Associate the BIOS policy to the server profile and reboot the server.

Note For more information, see *Creating a BIOS Policy* in [Configuring Server Policies](#) and [Configuring Server Profile](#).

The BIOS Policy is successfully created and deployed on the server profile.

What to do next

Configure RoCE v2 for NVMeoF on the Host System.

Configuring RoCE v2 for NVMeoF on the Host System

Before you begin

Configure the Server Profile with RoCE v2 vNIC and the IOMMU enabled BIOS policy.

Procedure

- Step 1** Open the `/etc/default/grub` file for editing.
- Step 2** Add `intel_iommu=on` to the end of `GRUB_CMDLINE_LINUX`.
- ```
sample /etc/default/grub configuration file after adding intel_iommu=on:
cat /etc/default/grub
GRUB_TIMEOUT=5
GRUB_DISTRIBUTOR="$(sed 's, release .*$,,g' /etc/system-release)"
GRUB_DEFAULT=saved
GRUB_DISABLE_SUBMENU=true
GRUB_TERMINAL_OUTPUT="console"
GRUB_CMDLINE_LINUX="crashkernel=auto rd.lvm.lv=rhel/root rd.lvm.lv=rhel/swap biosdevname=1
rhgb quiet intel_iommu=on
GRUB_DISABLE_RECOVERY="true"
```
- Step 3** After saving the file, generate a new `grub.cfg` file.
- For Legacy boot:
- ```
# grub2-mkconfig -o /boot/grub2/grub.cfg
```
- For UEFI boot:
- ```
grub2-mkconfig -o /boot/grub2/efi/EFI/redhat/grub.cfg
```
- Step 4** Reboot the server. You must reboot your server for the changes to take after enabling IOMMU.
- Step 5** Verify the server is booted with `intel_iommu=on` option.
- ```
cat /proc/cmdline | grep iommu
```
- Note its inclusion at the end of the output.

```
[root@localhost basic-setup]# cat /proc/cmdline | grep iommu
BOOT_IMAGE=vmlinux-3.10.0-957.27.2.el7.x86_64 root=/dev/mapper/rhel-root ro crashkernel=auto
rd.lvm.lv=rhel/root rd.lvm.lv=rhel/swap rhgb quiet intel_iommu=on LANG=en US.UTF-8
```

What to do next

Download the enic and enic_rdma drivers.

Installing Cisco enic and enic_rdma Drivers

The enic_rdma driver requires enic driver. When installing enic and enic_rdma drivers, download and use the matched set of enic and enic_rdma drivers on Cisco.com. Attempting to use the binary enic_rdma driver downloaded from Cisco.com with an inbox enic driver, will not work.

Procedure

Step 1 Install the enic and enic_rdma rpm packages:

```
# rpm -ivh kmod-enic-<version>.x86_64.rpm kmod-enic_rdma-<version>.x86_64.rpm
```

Note During enic_rdma installation, the enic_rdmalibnvdimm module may fail to install on RHEL 7.7 because the nvdimm-security.conf dracut module needs spaces in the add_drivers value. For workaround, please follow the instruction from the following links:

<https://access.redhat.com/solutions/4386041>

https://bugzilla.redhat.com/show_bug.cgi?id=1740383

Step 2 The enic_rdma driver is now installed but not loaded in the running kernel. Reboot the server to load enic_rdma driver into the running kernel.

Step 3 Verify the installation of enic_rdma driver and RoCE v2 interface:

```
[root@localhost ~]# dmesg | grep enic_rdma
[  3.137083] enic_rdma: Cisco VIC Ethernet NIC RDMA Driver, ver 1.2.0.28-877.2
2 init
[  3.242663] enic 0000:1b:00.1 eno6: enic_rdma: FW v3 RoCEv2 enabled
[  3.284856] enic 0000:1b:00.4 eno9: enic_rdma: FW v3 RoCEv2 enabled
[ 16.441662] enic 0000:1b:00.1 eno6: enic_rdma: Link UP on enic_rdma_0
[ 16.458754] enic 0000:1b:00.4 eno9: enic_rdma: Link UP on enic_rdma_1
```

Step 4 Load the nvme-rdma kernel module:

```
# modprobe nvme-rdma
```

After server reboot, nvme-rdma kernel module is unloaded. To load nvme-rdma kernel module every server reboot, create nvme_rdma.conf file using:

```
# echo nvme_rdma > /etc/modules-load.d/nvme_rdma.conf
```

Note For more information about enic_rdma after installation, use the `rpm -q -l kmod-enic_rdma` command to extract the README file.

What to do next

Discover targets and connect to NVMe namespaces. If your system needs multipath access to the storage, go to the section for [Setting Up Device Mapper Multipath, on page 32](#).

Discovering the NVMe Target

Use this procedure to discover the NVMe target and connect NVMe namespaces.

Before you begin

Install `nvme-cli` version 1.6 or later if it is not installed already.



Note Skip to Step 2 below if `nvme-cli` version 1.7 or later is installed.

Configure the IP address on the RoCE v2 interface and make sure the interface can ping the target IP.

Procedure

Step 1 Create an `nvme` folder in `/etc`, then manually generate host `nqn`.

```
# mkdir /etc/nvme
# nvme gen-hostnqn > /etc/nvme/hostnqn
```

Step 2 Create a `settos.sh` file and run the script to set priority flow control (PFC) in IB frames.

Note To avoid failure of sending NVMeoF traffic, you *must* create and run this script after *every* server reboot.

```
# cat settos.sh
#!/bin/bash
for f in `ls /sys/class/infiniband`;
do
    echo "setting TOS for IB interface:" $f
    mkdir -p /sys/kernel/config/rdma_cm/$f/ports/1
    echo 186 > /sys/kernel/config/rdma_cm/$f/ports/1/default_roce_tos
done
```

Step 3 Discover the NVMe target by entering the following command.

```
nvme discover --transport=rdma --traddr=<IP address of transport target port>
```

For example, to discover the target at 50.2.85.200:

```
# nvme discover --transport=rdma --traddr=50.2.85.200

Discovery Log Number of Records 1, Generation counter 2
=====Discovery Log Entry 0=====
trtype: rdma
adrfam: ipv4
subtype: nvme subsystem
treq: not required
portid: 3
trsvcid: 4420
subnqn: nqn.2010-06.com.purestorage:flasharray.9a703295ee2954e
traddr: 50.2.85.200
rdma_prtype: roce-v2
rdma_qptype: connected
rdma_cms: rdma-cm
rdma_pkey: 0x0000
```

Note To discover the NVMe target using IPv6, put the IPv6 target address next to the `traddr` option.

Step 4 Connect to the discovered NVMe target by entering the following command.

```
nvme connect --transport=rdma --traddr=<IP address of transport target port>> -n <subnqn value from nvme discover>
```

For example, to discover the target at 50.2.85.200 and the subnqn value found above:

```
# nvme connect --transport=rdma --traddr=50.2.85.200 -n
nqn.2010-06.com.purestorage:flasharray.9a703295ee2954e
```

Note To connect to the discovered NVMe target using IPv6, put the IPv6 target address next to the `traddr` option.

Step 5 Use the `nvme list` command to check mapped namespaces:

```
# nvme list
Node          SN              Model          Namespace
Usage                Format          FW Rev
-----
/dev/nvme0n1    09A703295EE2954E  Pure Storage FlashArray  72656
4.29 GB / 4.29 GB  512 B + 0 B  99.9.9
/dev/nvme0n2    09A703295EE2954E  Pure Storage FlashArray  72657
5.37 GB / 5.37 GB  512 B + 0 B  99.9.9
```

Setting Up Device Mapper Multipath

If your system is configured with Device Mapper multipathing (DM Multipath), use the following steps to set up Device Mapper multipath.

Procedure

Step 1 Install the `device-mapper-multipath` package if it is not installed already

Step 2 Enable and start multipathd:

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

Step 3 Edit the `etc/multipath.conf` file to use the following values :

```
defaults {
    polling_interval      10
    path_selector         "queue-length 0"
    path_grouping_policy  multibus
    fast_io_fail_tmo     10
    no_path_retry         0
    features              0
    dev_loss_tmo          60
    user_friendly_names  yes
}
```

Step 4 Flush with the updated multipath device maps.

```
# multipath -F
```

Step 5 Restart multipath service:

```
# systemctl restart multipathd.service
```

Step 6 Rescan multipath devices:

```
# multipath -v2
```

Step 7 Check the multipath status:

```
# multipath -ll
```

Deleting the RoCE v2 Interface Using Cisco Intersight

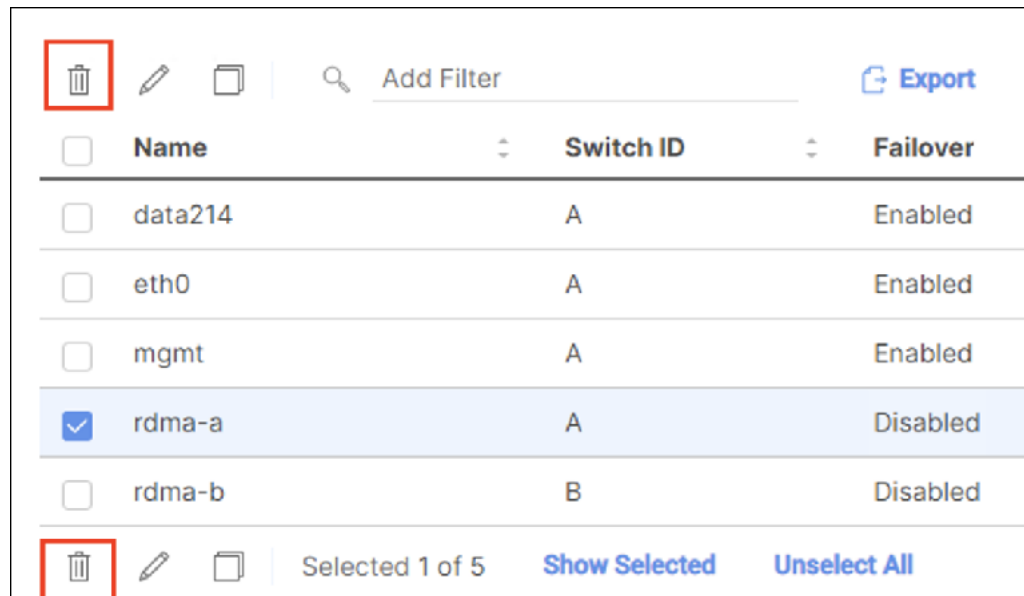
Use these steps to remove the RoCE v2 interface.

Procedure

Step 1 Navigate to **CONFIGURE > Policies**. In the **Add Filter** field, select **Type: LAN Connectivity**.

Step 2 Select the appropriate LAN Connectivity policy created for RoCE V2 configuration and use the delete icon on the top or bottom of the policy list.

Step 3 Click **Delete** to delete the policy.



<input type="checkbox"/>	Name	Switch ID	Failover
<input type="checkbox"/>	data214	A	Enabled
<input type="checkbox"/>	eth0	A	Enabled
<input type="checkbox"/>	mgmt	A	Enabled
<input checked="" type="checkbox"/>	rdma-a	A	Disabled
<input type="checkbox"/>	rdma-b	B	Disabled

Selected 1 of 5 [Show Selected](#) [Unselect All](#)

Step 4 Upon deleting the RoCE v2 configuration, re-deploy the server profile and reboot the server.



CHAPTER 4

Configuring NVMeoF with RoCEv2 in ESXi

- [Guidelines for using NVMe over Fabrics \(NVMeoF\) with RoCE v2 on ESXi, on page 35](#)
- [ESXi Requirements, on page 36](#)
- [Configuring RoCE v2 for NVMeoF on Cisco Intersight, on page 36](#)
- [NENIC Driver Installation, on page 41](#)
- [ESXi NVMe RDMA Host Side Configuration, on page 42](#)
- [Deleting the RoCE v2 Interface Using Cisco Intersight, on page 49](#)

Guidelines for using NVMe over Fabrics (NVMeoF) with RoCE v2 on ESXi

General Guidelines and Limitations:

- Cisco recommends you to check the [UCS Hardware and Software Compatibility](#) to determine support for NVMeoF. NVMeoF is supported on Cisco UCS B-Series, C-Series, and X-Series servers.
- Nonvolatile Memory Express (NVMe) over RDMA with RoCE v2 is currently supported only with Cisco VIC 15000 Series adapters.
- When creating RoCE v2 interfaces, use Cisco recommended Queue Pairs, Memory Regions, Resource Groups, and Class of Service settings. NVMeoF functionality may not be guaranteed with different settings for Queue Pairs, Memory Regions, Resource Groups, and Class of Service.
- RoCE v2 supports maximum two RoCE v2 enabled interfaces per adapter.
- Booting from an NVMeoF namespace is not supported.
- Layer 3 routing is not supported.
- Saving a crashdump to an NVMeoF namespace during a system crash is not supported.
- NVMeoF cannot be used with usNIC, VxLAN, VMQ, VMMQ, NVGRE, GENEVE Offload, ENS, and DPDK features.
- Cisco Intersight does not support fabric failover for vNICs with RoCE v2 enabled.
- The Quality of Service (QoS) no drop class configuration must be properly configured on upstream switches such as Cisco Nexus 9000 series switches. QoS configurations will vary between different upstream switches.

- During the failover or failback event, the Spanning Tree Protocol (STP) can result temporary loss of network connectivity. To prevent this connectivity issue, disable STP on uplink switches.

Downgrade Guidelines: Remove the RoCEv2 configuration first and then downgrade to the release version lower than Cisco UCS Manager release 4.2(3b) version.

ESXi Requirements

Configuration and use of RoCE v2 in ESXi requires the following:

- VMWare ESXi version 7.0 Update 3.
- Cisco UCS Manager Release 4.2(3b) or later versions.
- VIC firmware 5.2(3x) or later versions.
- The driver version, *nenic-2.0.4.0-IOEM.700.1.0.15843807.x86_64.vib* that provides both standard eNIC and RDMA support with the Cisco UCS Manager 4.2(3b) release package.
- A storage array that supports NVMeoF connection.

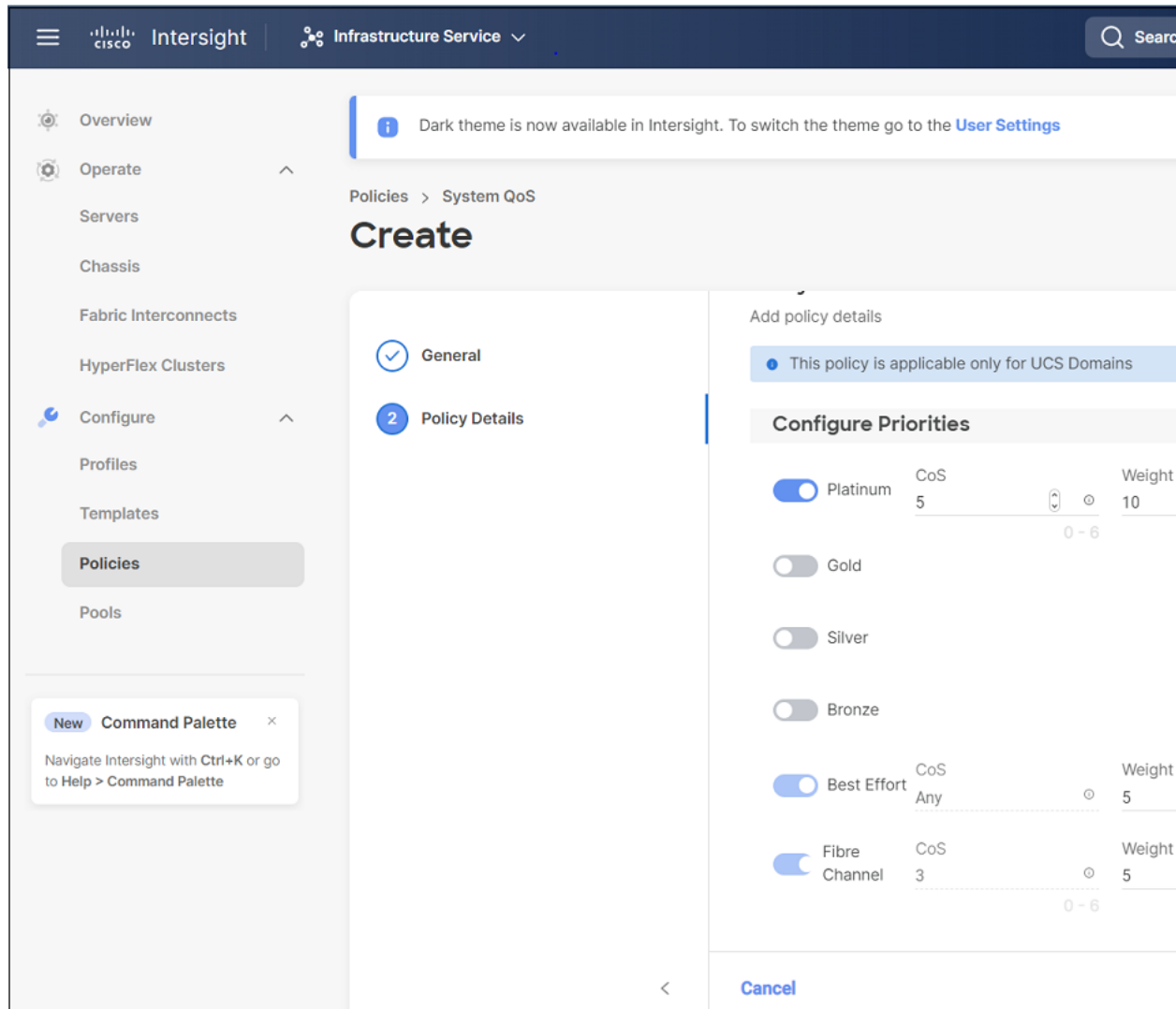
Configuring RoCE v2 for NVMeoF on Cisco Intersight

Use these steps to configure the RoCE v2 interface on Cisco Intersight.

To avoid possible RDMA packet drops, ensure same no-drop COS is configured across the network. The following steps allows you to configure a no-drop class in System QoS policies and use it for RDMA supported interfaces.

Procedure

-
- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Domain** platform type, search or choose **System QoS**, and click **Start**.
- Step 2** In the **General** page, enter the policy name and click **Next**, and then in the **Policy Details** page, configure the property setting for System QoS policy as follows:
- For **Priority**, choose **Platinum**
 - For **Allow Packet Drops**, uncheck the check box.
 - For **MTU**, set the value as **9216**.



Step 3 Click **Create**.

Step 4 Associate the System QoS policy to the Domain Profile.

UCS Domain Configuration

Select the compute and management policies to be associated with the fabric interconnect.

Show Attached Policies (1)

^ **Management** 0 of 4 Policies Configured

NTP

Syslog

Network Connectivity

SNMP

^ **Network** 1 of 2 Policies Configured

System QoS *

Switch Control

Note For more information, see *Creating System QoS Policy* in [Configuring Domain Policies](#) and [Configuring Domain Profiles](#).

The System QoS Policy is successfully created and deployed to the Domain Profile.

What to do next

Configure the server profile with RoCE v2 vNIC settings in LAN Connectivity policy.

Enabling RoCE Settings in LAN Connectivity Policy

Use the following steps to configure the RoCE v2 vNIC. In Cisco Intersight LAN Connectivity policy, you can enable the RoCE settings on **Ethernet Adapter policy** for Linux configuration as follows:

Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. Click **Create Policy**, select **UCS Server** platform type, search or choose **LAN Connectivity policy**, and click **Start**.

- Step 2** In the policy **General** page, enter the policy name, select the Target Platform as **UCS Server (Standalone)** or **UCS Server (FI-Attached)**, and click **Next**.
- Step 3** In the **Policy Details** page, click **Add vNIC** to create a new vNIC.
- Step 4** In the **Add vNIC** page, follow the configuration parameters to enable the RoCE v2 vNIC:
- a) In the **General** section, provide a name for virtual ethernet interface.
 - b) In case of a Standalone server, click the **Consistent Device Naming (CDN)** or click the **Failover** of a FI-attached server, and do the following:
 - Click **Select Policy** under **Ethernet Adapter**.
 - In the **Select Policy** window, click **Create New** to create an Ethernet Adapter policy.
 - In the **General** page of the Ethernet Adapter Policy, enter the policy name and click **Next**.
 - In the **Policy Details** page of the Ethernet Adapter Policy, modify the following property setting:
 - **RoCE Settings**
 - For **Enable RDMA over Converged Ethernet**, slide to enable and set the RoCE on this virtual interface.
 - For **Queue Pairs**, select or enter **1024**
 - For **Memory Regions**, select or enter **131072**
 - For **Resource Groups**, select or enter **8**
 - For **Version**, select **Version 2**
 - For **Class of Service**, select **5**
 - **Interrupt Settings**
 - For **Interrupts**, select or enter **256**.
 - For **Interrupt mode**, select **MSIX**.
 - For **Interrupt Timer, us**, select **125**.
 - For **Interrupt Coalescing Type**, select **Min**.
 - **Receive Settings**
 - For **Receive Queue Count**, select or enter **1**.
 - For **Receiving Ring Size**, select or enter **512**.
 - **Transmit Settings**
 - For **Transmit Queue Count**, select or enter **1**.
 - For **Transmit Ring Size**, select or enter **256**.
 - **Completion Settings**
 - For **Completion Queue Count**, select or enter **2**.
 - For **Completion Ring Size**, select or enter **1**.

- For **Uplink Failback Timeout(seconds)**, select or enter **5**
- Click **Create** to create an Ethernet Adapter Policy with the above defined settings.

The screenshot displays the Cisco Intersight interface for creating an Ethernet Adapter Policy. The left sidebar shows the navigation menu with 'Policies' highlighted. The main area is titled 'Create Ethernet Adapter' and features two tabs: 'General' (active) and 'Policy Details'. On the right side, there are several toggle switches for various settings, including 'Enable Virtual E...', 'Enable Network...', 'Enable Accelerat...', 'Enable Precision...', 'Enable Advance...', 'Enable Interrupt...', and 'Enable GENEVE...'. Below these is the 'RoCE Settings' section, which includes a toggle for 'Enable RDMA o...', a 'Queue Pairs' field set to '1024', and a 'Version' dropdown set to 'Version 2'. A 'Cancel' button is visible at the bottom right. A 'Command Palette' notification is also present in the bottom left corner.

- Click **Add** to save the setting and add the new vNIC.

Note All the fields with * are mandatory and ensure it is filled out or selected with appropriate policies.

Step 5 Click **Create** to complete the LAN Connectivity policy with RoCE v2 settings.

Step 6 Associate the LAN Connectivity policy to the Server Profile.

Note For more information, see *Creating a LAN Connectivity Policy* and *Creating an Ethernet Adapter Policy* in [Configuring UCS Server Policies](#) and [Configuring UCS Server Profiles](#).

The LAN Connectivity Policy with the Ethernet Adapter policy vNIC setting is successfully created and deployed to enable RoCE v2 configuration.

What to do next

Once the policy configuration for RoCE v2 is complete, configure RoCE v2 for NVMeoF on the Host System.

NENIC Driver Installation

Before you begin

The Ethernet Network Interface Card (eNIC) Remote Direct Memory Access (RDMA) driver requires nenic driver.

Procedure

Step 1 Copy the eNIC vSphere Installation Bundle (VIB) or offline bundle to the ESXi server.

Step 2 Use the command to install nenic driver:

```
esxcli software vib install -v {VIBFILE}
or
esxcli software vib install -d {OFFLINE_BUNDLE}
```

Example:

```
esxcli software vib install -v /tmp/nenic-2.0.4.0-10EM.700.1.0.15843807.x86_64.vib
```

Note Depending on the certificate used to sign the VIB, you may need to change the host acceptance level. To do this, use the command:

```
esxcli software acceptance set --level=<level>
```

Depending on the type of VIB installed, you may need to put ESX into maintenance mode. This can be done through the client, or by adding the `--maintenance-mode` option to the above `esxcli`.

What to do next

Configure the Host side for ESXi NVMe RDMA.

ESXi NVMe RDMA Host Side Configuration

NENIC RDMA Functionality

One of the major difference between RDMA on Linux and ESXi is listed below:

- In ESXi, the physical interface (vmnic) MAC is not used for RoCEv2 traffic. Instead, the VMkernel port (vmk) MAC is used.

Outgoing RoCE packets use the vmk MAC in the Ethernet source MAC field, and incoming RoCE packets use the vmk MAC in the Ethernet destination mac field. The vmk MAC address is a VMware MAC address assigned to the vmk interface when it is created.

- In Linux, the physical interface MAC is used in source MAC address field in the ROCE packets. This Linux MAC is usually a Cisco MAC address configured to the VNIC using UCS Manager.

If you ssh into the host and use the `esxcli network ip interface list` command, you can see the MAC address.

```
vmk0
Name: vmk0
MAC Address: 2c:f8:9b:a1:4c:e7
Enabled: true
Portset: vSwitch0
Portgroup: Management Network
Netstack Instance: defaultTcpipStack
VDS Name: N/A
VDS UUID: N/A
VDS Port: N/A
VDS Connection: -1
Opaque Network ID: N/A
Opaque Network Type: N/A
External ID: N/A
MTU: 1500
TSO MSS: 65535
RXDispQueue Size: 2
Port ID: 67108881
```

You must create a vSphere Standard Switch to provide network connectivity for hosts, virtual machines, and to handle VMkernel traffic. Depending on the connection type that you want to create, you can create a new vSphere Standard Switch with a VMkernel adapter, only connect physical network adapters to the new switch, or create the switch with a virtual machine port group.

Create Network Connectivity Switches

Use these steps to create a vSphere Standard Switch to provide network connectivity for hosts, virtual machines, and to handle VMkernel traffic.

Before you begin

Ensure you have nenic drivers. Download and install nenic drivers before proceeding with below steps:

Procedure

- Step 1** In the vSphere Client, navigate to the host.
- Step 2** On the **Configure** tab, expand **Networking** and select **Virtual Switches**.
- Step 3** Click on **Add Networking**.
- The available network adapter connection types are:
- **Vmkernel Network Adapter**
Creates a new VMkernel adapter to handle host management traffic
 - **Physical Network Adapter**
Adds physical network adapters to a new or existing standard switch.
 - **Virtual Machine Port Group for a Standard Switch**
Creates a new port group for virtual machine networking.
- Step 4** Select connection type **Vmkernel Network Adapter**.
- Step 5** Select **New Standard Switch** and click **Next**.
- Step 6** Add physical adapters to the new standard switch.
- a) Under **Assigned Adapters**, select **New Adapters**.
 - b) Select one or more adapters from the list and click **OK**. To promote higher throughput and create redundancy, add two or more physical network adapters to the Active list.
 - c) (Optional) Use the up and down arrow keys to change the position of the adapter in the Assigned Adapters list.
 - d) Click **Next**.
- Step 7** For the new standard switch you just created for the VMadapter or a port group, enter the connection settings for the adapter or port group.
- a) Enter a label that represents the traffic type for the VMkernel adapter.
 - b) Set a VLAN ID to identify the VLAN the VMkernel uses for routing network traffic.
 - c) Select IPV4 or IPV6 or both.
 - d) Select an MTU size from the drop-down menu. Select Custom if you wish to enter a specific MTU size. The maximum MTU size is 9000 bytes.
- Note** You can enable Jumbo Frames by setting an MTU greater than 1500.
- e) After setting the TCP/IP stack for the VMkernel adapter, select a TCP/IP stack.
To use the default TCP/IP stack, select it from the available services.
- Note** Be aware that the TCP/IP stack for the VMkernel adapter cannot be changed later.
- f) Configure IPV4 and/or IPV6 settings.
- Step 8** On the **Ready to Complete** page, click **Finish**.

Step 9 Check the VMkernel ports for the VM Adapters or port groups with NVMe RDMA in the vSphere client, as shown in the Results below.

The VMkernel ports for the VM Adapters or port groups with NVMe RDMA are shown below.

The screenshot shows the vSphere Client interface with the 'Configure' tab selected. The left-hand navigation pane is expanded to 'Networking' > 'VMkernel adapters'. The main area displays a table of VMkernel adapters.

Device	Network Label	Switch	IP Address	TCP/IP Stack	Enabled
vmk0	Management Network	vSwitch0	10.193.176.52	Default	Management
vmk1	vmk284	vSwitch1	50.284::210	Default	--
vmk2	vmk283	vSwitch2	50.2.83.210	Default	--

The VRDMA Port groups created with NVMeRDMA supported vmnic appear as below.

The screenshot shows the vSphere Client interface with the 'Configure' tab selected. The left-hand navigation pane is expanded to 'Networking' > 'RDMA adapters'. The main area displays a table of RDMA adapters.

Name	Driver	State	Paired Uplink	RoCE v1	RoCE v2	iWARP
vmrdma0	nenic	Active	vmnic2	Disabled	Enabled	Disabled
vmrdma1	nenic	Active	vmnic3	Disabled	Enabled	Disabled

Below the table, the 'RDMA Device: vmrdma1' section is visible, showing properties for 'Bound VMkernel Adapters':

VMkernel Adapter	TCP/IP Stack	IP Address
vmk2	Default	50.2.83.210

What to do next

Create vmhba ports on top of vmrdma ports.

Create VMVHBA Ports in ESXi

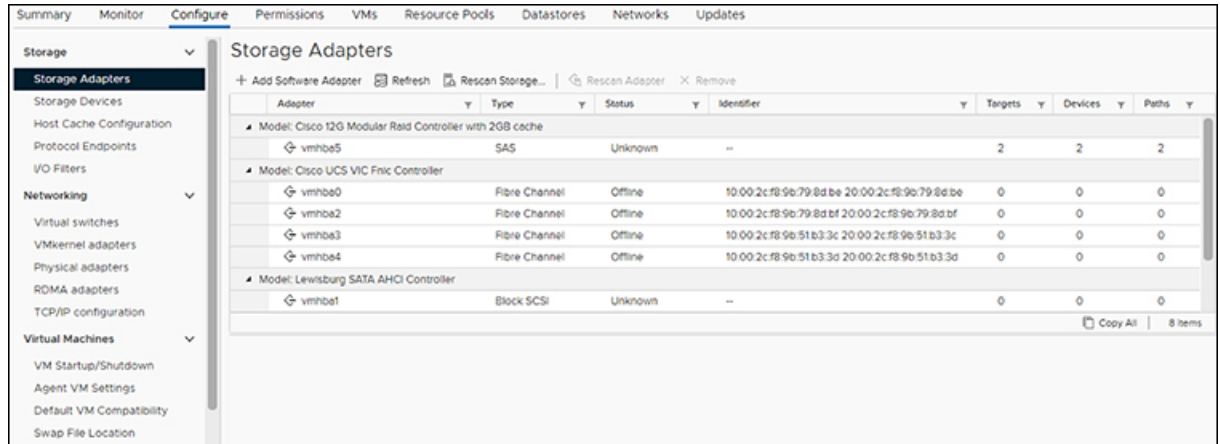
Use the following steps for creating vmhba ports on top of the vmrdma adapter ports.

Before you begin

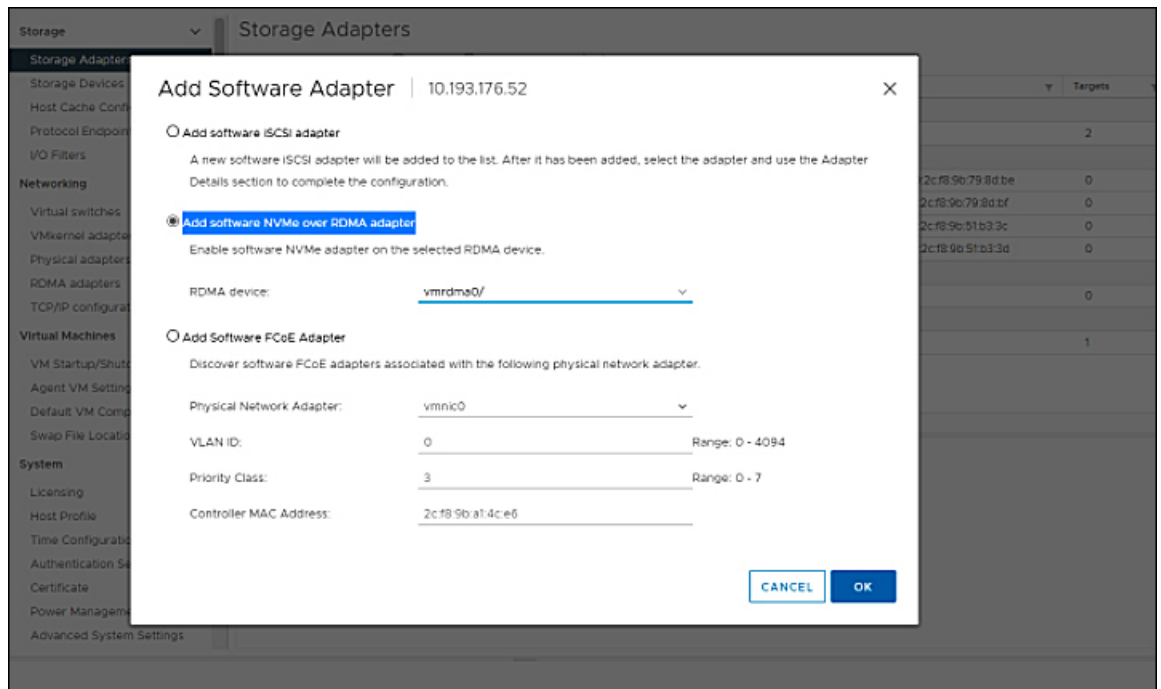
Create the adapter ports for storage connectivity.

Procedure

- Step 1** Go to vCenter where your ESXi host is connected.
- Step 2** Click on **Host>Configure>Storage adapters**.

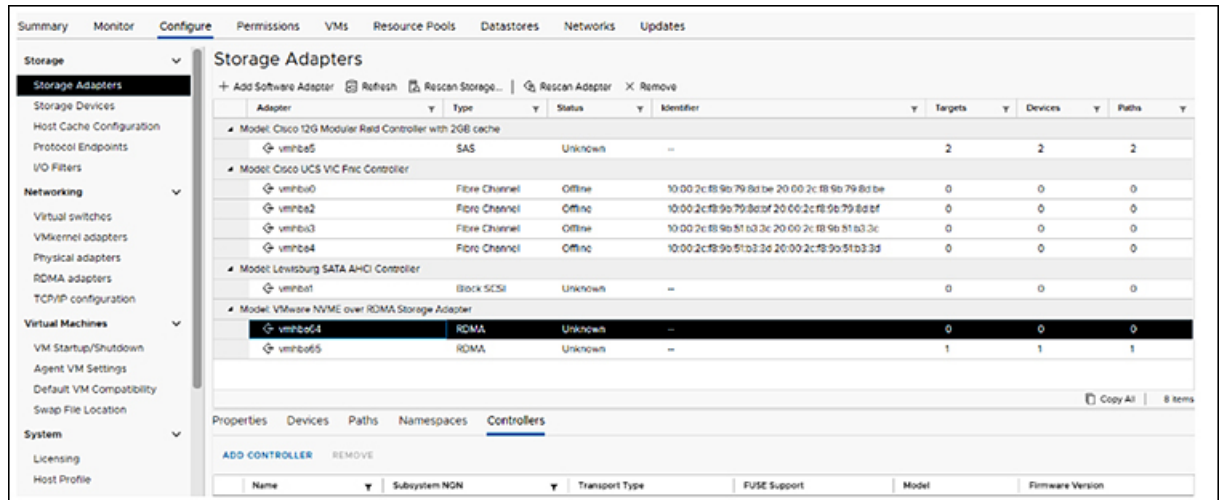


- Step 3** Click **+Add Software Adapter**. The following dialog box will appear.



- Step 4** Select **Add software NVMe over RDMA adapter** and the vmrdma port you want to use.
- Step 5** Click **OK**

The vmhba ports for the VMware NVMe over RDMA storage adapter will be shown as in the example below



Displaying vmnic and vmrDMA Interfaces

ESXi creates a vmnic interface for each nenic VNIC configured to the host.

Before you begin

Create Network Adapters and VHBA ports.

Procedure

- Step 1** Use `ssh` to access the host system.
- Step 2** Enter `esxcfg-nics -l` to list the vmnics on ESXi.

Name	PCI	Driver	Link	Speed	Duplex	MAC Address	MTU	Description
vmnic0	0000:3b:00:0	ixgben	Down	0Mbps	Half	2c:f8:9b:a1:4c:e6	1500	Intel(R) Ethernet Controller X550
vmnic1	0000:3b:00:1	ixgben	Up	1000Mbps	Full	2c:f8:9b:a1:4c:e7	1500	Intel(R) Ethernet Controller X550
vmnic2	0000:1d:00:0	nenic	Up	5000Mbps	Full	2c:f8:9b:79:8d:bc	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic3	0000:1d:00:1	nenic	Up	5000Mbps	Full	2c:f8:9b:79:8d:bd	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic4	0000:63:00:0	nenic	Down	0Mbps	Half	2c:f8:9b:51:b3:3a	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic5	0000:63:00:1	nenic	Down	0Mbps	Half	2c:f8:9b:51:b3:3b	1500	Cisco Systems Inc Cisco VIC Ethernet NIC

esxcli network nic list

Name	PCI Device	Driver	Admin Status	Link Status	Speed	Duplex	MAC Address	MTU	Description
vmnic0	0000:3b:00:0	ixgben	Up	Down	0	Half	2c:f8:9b:a1:4c:e6	1500	Intel(R) Ethernet Controller X550
vmnic1	0000:3b:00:1	ixgben	Up	Up	1000	Full	2c:f8:9b:a1:4c:e7	1500	Intel(R) Ethernet Controller X550
vmnic2	0000:1d:00:0	nenic	Up	Up	50000	Full	2c:f8:9b:79:8d:bc	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic3	0000:1d:00:1	nenic	Up	Up	50000	Full	2c:f8:9b:79:8d:bd	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic4	0000:63:00:0	nenic	Up	Down	0	Half	2c:f8:9b:51:b3:3a	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic5	0000:63:00:1	nenic	Up	Down	0	Half	2c:f8:9b:51:b3:3b	1500	Cisco Systems Inc Cisco VIC Ethernet NIC

- Step 3** Use `esxcli rdma device list` to list the vmrDMA devices. When the enic driver registers with ESXi the RDMA device for a RDMA capable VNIC, ESXi creates a vmrDMA device and links it to the corresponding vmnic.

```
[root@ESXi7U3:~] esxcli rdma device list
Name      Driver  State  MTU  Speed  Paired Uplink  Description
-----
vmdma0    nenic   Active 4096  50 Gbps  vmnic1        Cisco UCS VIC 15XXX (A0)
vmdma1    nenic   Active 4096  50 Gbps  vmnic2        Cisco UCS VIC 15XXX (A0)
[root@ESXi7U3:~] esxcli rdma device vmknic list
Device    Vmknic  NetStack
-----
vmdma0    vmk1    defaultTcpipStack
vmdma1    vmk2    defaultTcpipStack
```

Step 4 Use `esxcli rdma device protocol list` to check the protocols supported by the vmdma interface.

For enic, RoCE v2 is the only protocol supported from this list. The output of this command should match the RoCEv2 configuration on the VNIC.

```
[root@ESXi7U3:~] esxcli rdma device protocol list
Device    RoCE v1  RoCE v2  iWARP
-----
vmdma0    false    true     false
vmdma1    false    true     false
[root@ESXi7U3:~]
```

Step 5 Use `esxcli nvme adapter list` to list the NVMe adapters and the vmdma and vmnic interfaces it is configured on.

```
[root@ESXi7U3:~] esxcli nvme adapter list
Adapter  Adapter Qualified Name  Transport Type  Driver
-----
vmhba64  aqn:nvmerdma:2c-f8-9b-79-8d-bc  RDMA            nvmerdma
vmhba65  aqn:nvmerdma:2c-f8-9b-79-8d-bd  RDMA            nvmerdma
[root@ESXi7U3:~]
```

Step 6 All vmhbases in the system can be listed using `esxcli storage core adapter list`. The vmhba configured over RDMA.

```
[root@ESXi7U3:~] esxcli storage core adapter list
HBA Name  Driver  Link State  UID  Capabilities
-----
vmhba0    nfnic   link-down   fc.10002cf89b798dbe:20002cf89b798dbe  Second Level Lun
vmhba1    vmw_ahci  link-n/a   sata.vmhba1
vmhba2    nfnic   link-down   fc.10002cf89b798dbf:20002cf89b798dbf  Second Level Lun
vmhba3    nfnic   link-down   fc.10002cf89b51b33c:20002cf89b51b33c  Second Level Lun
vmhba4    nfnic   link-down   fc.10002cf89b51b33d:20002cf89b51b33d  Second Level Lun
vmhba5    lsi_mr3  link-n/a   sas.5cc167e9732f9b00
vmhba64  nvmerdma  link-n/a   rdma.vmknic2:2c:f8:9b:79:8d:bc
vmhba65  nvmerdma  link-n/a   rdma.vmknic3:2c:f8:9b:79:8d:bd
[root@ESXi7U3:~]
```

Note For vmhba64 and vmhba65, you may observe that the driver's Link State displays *link-n/a* instead of *Online*. This is a known issue in ESXi 7.0 Update 3. For more information, see [ESXi](#).

NVMe Fabrics and Namespace Discovery

This procedure is performed through the ESXi command line interface.

Before you begin

Create and configure NVMe on the adapter's VMHBAs. The maximum number of adapters is two, and it is a best practice to configure both for fault tolerance.

Procedure

Step 1 Check and enable NVMe on the vmrdma device.

```
esxcli nvme fabrics enable -p RDMA -d vmrdma0
```

The system should return a message showing if NVMe is enabled.

Step 2 Discover the NVMe fabric on the array by entering the following command:

```
esxcli nvme fabrics discover -a vmhba64 -l transport_address
```

figure with `esxcli nvme fabrics discover -a vmhba64 -l 50.2.84.100`

The output will list the following information: Transport Type, Address Family, Subsystem Type, Controller ID, Admin Queue, Max Size, Transport Address, Transport Service ID, and Subsystem NQN

You will see output on the NVMe controller.

Step 3 Perform NVMe fabric interconnect.

```
esxcli nvme fabrics discover -a vmhba64 -l transport_address p Transport Service ID -s Subsystem NQN
```

Step 4 Repeat steps 1 through 4 to configure the second adapter.

Step 5 Verify the configuration.

a) Display the controller list to verify the NVMe controller is present and operating.

```
esxcli nvme controller list RDMA -d vmrdma0
```

```
[root@ESXi7U3:~] esxcli nvme controller list
Name
-----
nqn.2010-06.com.purestorage:flasharray.5ab274df5b161455#vmhba64#50.2.84.100:4420
nqn.2010-06.com.purestorage:flasharray.5ab274df5b161455#vmhba65#50.2.83.100:4420
[root@ESXi7U3:~] esxcli nvme namespace list
Name                               Controller Number  Namespace ID      Block Size  Capacity
-----
eui.00e6d65b65a8f34024a9374e00011745  258                71493              512
eui.00e6d65b65a8f34024a9374e00011745  259                71493              512
[root@ESXi7U3:~]
```

- b) Verify that the fabric is enabled on the controller through the adapter, and verify the controller is accessible through the port on the adapter.

```
[root@ESXiUCSA:~] esxcli nvme fabrics enable -p RDMA -d vmrdma0
NVMe already enabled on vmrdma0
[root@ESXiUCSA:~] esxcli nvme fabrics discover -a vmhba64 -l 50.2.84.100
Transport Type Address Family Subsystem Type Controller ID Admin Queue Max Size Transport
Address Transport Service ID Subsystem NQN
-----
RDMA          IPV4          NVM          65535          31
50.2.84.100   4420
nq.210-06.com.purestorage:flasharray:2dp1239anjkl484
[root@ESXiUCSA:~] esxcli nvme fabrics discover -a vmhba64 -l 50.2.84.100 p 4420 -s
nq.210-06.com.purestorage:flasharray:2dp1239anjkl484
Controller already connected
```

Deleting the RoCE v2 Interface Using Cisco Intersight

Use these steps to remove the RoCE v2 interface.

Procedure

- Step 1** Navigate to **CONFIGURE > Policies**. In the **Add Filter** field, select **Type: LAN Connectivity**.
- Step 2** Select the appropriate LAN Connectivity policy created for RoCE V2 configuration and use the delete icon on the top or bottom of the policy list.
- Step 3** Click **Delete** to delete the policy.

<input type="checkbox"/>	Name	Switch ID	Failover
<input type="checkbox"/>	data214	A	Enabled
<input type="checkbox"/>	eth0	A	Enabled
<input type="checkbox"/>	mgmt	A	Enabled
<input checked="" type="checkbox"/>	rdma-a	A	Disabled
<input type="checkbox"/>	rdma-b	B	Disabled

Selected 1 of 5 Show Selected Unselect All

Step 4 Upon deleting the RoCE v2 configuration, re-deploy the server profile and reboot the server.



CHAPTER 5

Known Issues

- [Windows](#), on page 51
- [Linux](#), on page 52
- [ESXi](#), on page 52

Windows

Symptom	Conditions	Workaround
<p>On VIC 1400 Series adapters, the neNIC driver for Windows 2019 can be installed on Windows 2016 and the Windows 2016 driver can be installed on Windows 2019. However, this is an unsupported configuration.</p>	<p>Case 1 : Installing Windows 2019 nenic driver on Windows 2016 succeeds-but on Windows 2016 RDMA is not supported.</p> <p>Case 2 : Installing Windows 2016 nenic driver on Windows 2019 succeeds-but on Windows 2019 RDMA comes with default disabled state, instead of enabled state.</p>	<p>The driver binaries for Windows 2016 and Windows 2019 are in folders that are named accordingly. Install the correct binary on the platform that is being built/updated.</p>

Linux

Symptom	Conditions	Workaround
<p>When sending high bandwidth NVMe traffic on some Cisco Nexus 9000 switches, the switch port that connected to the storage sometimes reaches the max PFC peak and does not automatically clear the buffers. In Nexus 9000 switches, the nxos command "show hardware internal buffer info pkt-stats input peak" shows that the <code>Peak_cell</code> or <code>PeakQos</code> value for the port reaches more than 1000.</p>	<p>The NVMe traffic will drop.</p>	<p>To recover the switch from this error mode.</p> <ol style="list-style-type: none"> 1. Log into the switch. 2. Locate the port that connected to the storage and shut down the port using "shutdown" command 3. Execute the following commands one by one: <pre># clear counters # clear counter buffers module 1 # clear qos statistics</pre> 4. Run no shutdown on the port that was shut down.

ESXi

Symptom	Conditions	Workaround
<p>When using the command esxcli storage core adapter list to list the vmhba, the Driver's Link State for vmhba64 and vmhba65 rdma ports displays <i>Link-n/a</i> instead of <i>Online</i>.</p> <p>Note VMware Developer Center Partner Network (DCPN) Case ID - 00113157</p>	<p>This is a known issue in ESXi 7.0 Update 3.</p>	<p>None</p>