Cisco UCS Manager Configuration Guide for RDMA over Converged Ethernet (RoCE) Version 2, Release 4.3

First Published: 2023-12-08

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

CONTENTS

PREFACE

Preface v

Audience v

Conventions v

Related Cisco UCS Documentation vii

Documentation Feedback vii

CHAPTER 1

RDMA Over Converged Ethernet (RoCE) version 2 1

RDMA Over Converged Ethernet (RoCE) v2 1

CHAPTER 2

Configuring SMB Direct with RoCEv2 in Windows 3

Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet (RoCE) v2 3

Overview of Configuring RoCEv2 Modes 1 and 2 in Windows 5

Windows Requirements 5

Configuring SMB Direct Mode 1 on UCS Manager 6

Configuring SMB Direct Mode 1 on the Host System 8

Configuring Mode 2 on UCS Manager 11

Configuring Mode 2 on the Host System 14

CHAPTER 3

Configuring NVMe Over Fabrics (NVMeoF) with RoCEv2 in Linux 19

Guidelines for using NVMe over Fabrics (NVMeoF) with RoCEv2 on Linux 19

Linux Requirements 20

Configuring RoCEv2 for NVMeoF on UCS Manager 21

Enabling an SRIOV BIOS Policy 21

Configuring RoCEv2 for NVMeoF on the Host System 22

Installing Cisco enic and enic_rdma Drivers 23

Discovering the NVMe Target 23
Setting Up Device Mapper Multipath 25
Deleting the RoCEv2 Interface Using UCS Manager 26

CHAPTER 4 **Configuring NVMe with ROCE V2 in ESXi 27**
Guidelines for using RoCEv2 Protocol in the Native ENIC driver on ESXi 27
ESXi nENIC RDMA Requirements 28
Installing NENIC Driver 28
Configuring and Enabling RoCEv2 on UCS Manager 29
Configuring RoCEv2 for VMware NVMeoF on UCS Manager 29
ESXi NVMe RDMA Host Side Configuration 30
NENIC RDMA Functionality 30
Create Network Connectivity Switches 31
Create VMHBA Ports in ESXi 33
Displaying vmnic and vmrdma Interfaces 35
NVMe Fabrics and Namespace Discovery 36
Deleting the ESXi RoCEv2 Interface Using UCS Manager 37

CHAPTER 5 **Using the UCS Manager CLI to Configure the RoCEv2 Interface 39**
Configure Windows SMBDirect RoCEv2 Interface using UCS Manager CLI 39
Configuring the Linux RoCEv2 Interface Using the UCS Manager CLI 40
Deleting the Windows RoCEv2 Interface Using the CLI for UCS Manager 41
Deleting the Linux RoCEv2 Interface Using the UCS Manager CLI 42
Configuring the RoCEv2 VMware ESXi Interface Using the UCS Manager CLI 43
Deleting the ESXi RoCEv2 Interface Using the UCS Manager CLI 44

CHAPTER 6 **Known Issues in RoCEv2 47**
Known-Issues in RoCEv2 47

Preface

- Audience, on page v
- Conventions, on page v
- Related Cisco UCS Documentation, on page vii
- Documentation Feedback, on page vii

Audience

This guide is intended primarily for data center administrators with responsibilities and expertise in one or more of the following:

- Server administration
- Storage administration
- Network administration
- Network security

Conventions

Text Type	Indication
GUI elements	GUI elements such as tab titles, area names, and field labels appear in **this font**. Main titles such as window, dialog box, and wizard titles appear in **this font**.
Document titles	Document titles appear in *this font*.
TUI elements	In a Text-based User Interface, text the system displays appears in `this font`.
System output	Terminal sessions and information that the system displays appear in `this font`.
CLI commands	CLI command keywords appear in **this font**. Variables in a CLI command appear in *this font*.
[ ]	Elements in square brackets are optional.

Text Type	Indication
{x | y | z}	Required alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
< >	Nonprinting characters such as passwords are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the document.

Tip Means *the following information will help you solve a problem*. The tips information might not be troubleshooting or even an action, but could be useful information, similar to a Timesaver.

Timesaver Means *the described action saves time*. You can save time by performing the action described in the paragraph.

Caution Means *reader be careful*. In this situation, you might perform an action that could result in equipment damage or loss of data.

Warning IMPORTANT SAFETY INSTRUCTIONS

This warning symbol means danger. You are in a situation that could cause bodily injury. Before you work on any equipment, be aware of the hazards involved with electrical circuitry and be familiar with standard practices for preventing accidents. Use the statement number provided at the end of each warning to locate its translation in the translated safety warnings that accompanied this device.

SAVE THESE INSTRUCTIONS

Related Cisco UCS Documentation

Documentation Roadmaps

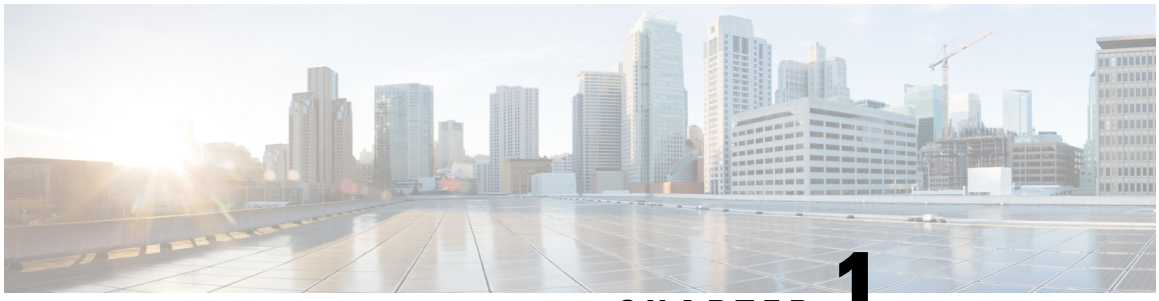
For a complete list of all B-Series documentation, see the *Cisco UCS B-Series Servers Documentation Roadmap* available at the following URL: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/overview/guide/UCS_roadmap.html

For a complete list of all C-Series documentation, see the *Cisco UCS C-Series Servers Documentation Roadmap* available at the following URL: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/overview/guide/ucs_rack_roadmap.html.

For information on supported firmware versions and supported UCS Manager versions for the rack servers that are integrated with the UCS Manager for management, refer to Release Bundle Contents for Cisco UCS Software.

Documentation Feedback

To provide technical feedback on this document, or to report an error or omission, please send your comments to ucs-docfeedback@external.cisco.com. We appreciate your feedback.



CHAPTER 1

RDMA Over Converged Ethernet (RoCE) version 2

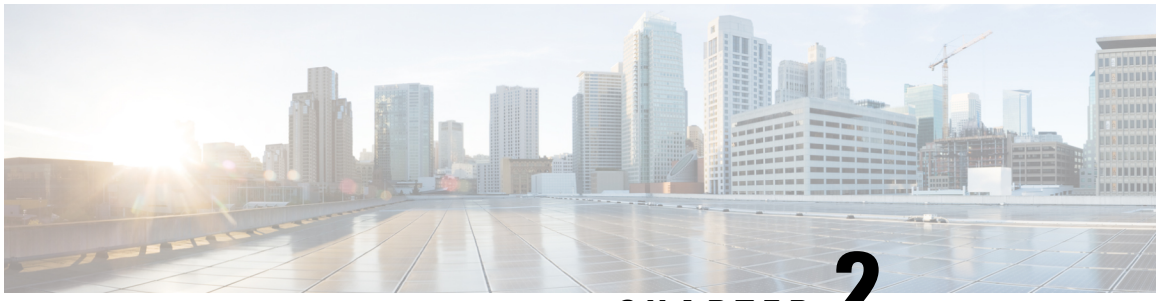
• RDMA Over Converged Ethernet (RoCE) v2 , on page 1

RDMA Over Converged Ethernet (RoCE) v2

RDMA over Converged Ethernet version 2 (RoCEv2) is an *internet layer* protocol, which means that RoCEv2 packets can be routed. RoCEv2 allows direct memory access over the network by encapsulating an Infiniband (IB) transport packet over Ethernet.

The RoCEv2 protocol exists on top of either the UDP/IPv4 or the UDP/IPv6 protocol. The UDP destination port number 4791 has been reserved for RoCEv2. Since RoCEv2 packets are routable, the RoCEv2 protocol is sometimes called Routable RoCE.

RoCEv2 is supported on the Windows, Linux, and ESXi Operating Systems.



CHAPTER 2

Configuring SMB Direct with RoCEv2 in Windows

- Guidelines for Using SMB Direct support on Windows using RDMA over converged Ethernet (RoCE) v2, on page 3
- Overview of Configuring RoCEv2 Modes 1 and 2 in Windows, on page 5
- Windows Requirements, on page 5
- Configuring SMB Direct Mode 1 on UCS Manager, on page 6
- Configuring SMB Direct Mode 1 on the Host System, on page 8
- Configuring Mode 2 on UCS Manager, on page 11
- Configuring Mode 2 on the Host System, on page 14

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

General Guidelines and Limitations:

- Cisco UCS Manager release 4.1.x and later releases support Microsoft SMB Direct with RoCEv2 on Microsoft Windows Server 2019 and later. Cisco recommends that you have all KB updates from Microsoft for your Windows Server release.

Note RoCEv2 is not supported on Microsoft Windows Server 2016.

- Cisco recommends you check UCS Hardware and Software Compatibility specific to your UCS Manager release to determine support for Microsoft SMB Direct with RoCEv2 on Microsoft Windows.
- Microsoft SMB Direct with RoCEv2 is supported only with Cisco UCS VIC 1400 Series, 14000 Series, and 15000 Series adapters. It is not supported with UCS VIC 1200 Series and 1300 Series adapters. SMB Direct with RoCEv2 is supported on all UCS Fabric Interconnects.

Note RoCEv1 is not supported with Cisco UCS VIC 1400 Series, Cisco UCS VIC 14000 Series, and Cisco UCS VIC 15000 Series.

- RoCEv2 configuration is supported only between Cisco adapters. Interoperability between Cisco adapters and third party adapters is not supported.
- RoCEv2 supports two RoCEv2 enabled vNIC per adapter and four virtual ports per adapter interface, independent of SET switch configuration.
- RoCEv2 cannot be used on the same vNIC interface as NVGRE, NetFlow, and VMQ features.
- RoCEv2 cannot be used with usNIC.
- RoCEv2-enabled vNIC interfaces must have the no-drop QoS system class enabled in UCS Manager.
- The RoCE Properties queue pairs setting must for be a minimum of 4 queue pairs.
- 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.
- UCS Manager does not support fabric failover for vNICs with RoCEv2 enabled.
- SMB Direct with RoCEv2 is supported on both IPv4 and IPv6.
- RoCEv2 cannot be used with GENEVE offload.

MTU Properties:

- In older versions of the VIC driver, the MTU was derived from either a UCS Manager service profile or from the Cisco IMC vNIC MTU setting in non-cluster setup. This behavior changes on Cisco UCS VIC 1400 Series and later adapters, where MTU is controlled from the Windows OS Jumbo Packet advanced property. A value configured from UCS Manager or Cisco IMC has no effect.
- The RoCEv2 MTU value is always power-of-two and its maximum limit is 4096.
- RoCEv2 MTU is derived from the Ethernet MTU.
- RoCEv2 MTU is the highest power-of-two that is less than the Ethernet MTU. For example:
  - if the Ethernet value is 1500, then the RoCEv2 MTU value is 1024
  - if the Ethernet value is 4096, then the RoCEv2 MTU value is 4096
  - if the Ethernet value is 9000, then the RoCEv2 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 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. Cisco does not support NDPKI Mode 3 operation.
- The recommended default adapter policy for RoCEv2 Mode 1 is Win-HPN-SMBd .
- The recommended default adapter policy for RoCEv2 Mode 2 is MQ-SMBd.
- RoCEv2 enabled vNICs for Mode2 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 RoCEv2 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 RoCEv2 properties.

Downgrade Limitations: Cisco recommends you remove the RoCEv2 configuration before downgrading to any non-supported RoCEv2 release. If the configuration is not removed or disabled, downgrade will fail.

Overview of Configuring RoCEv2 Modes 1 and 2 in Windows

Configuration of RoCEv2 on the Windows platform requires first configuring RoCEv2 Mode 1, then configuring RoCEv2 Mode 2. Modes 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.

To configure RoCEv2 mode 1, you will:

- Configure a no-drop class in CoS System Class. By default, Platinum with CoS 5 is a default in UCS Manager.
- Configure an Ethernet adapter policy for Mode 1 in UCS Manager.
- Configure Mode 1 on the host system.

RoCEv2 Mode 1 must be configured before configuring Mode 2.

To configure RoCEv2 mode 2, you will:

- Either create an Ethernet VMQ connection policy for RoCEv2 or use the UCS Manager MQ-SMBd policy.

Windows Requirements

Configuration and use of RDMA over Converged Ethernet for RoCEv2 in Windows Server requires the following:

- Windows 2019 or Windows Server 2022 with latest Microsoft updates
- UCS Manager release 4.1.1 or later
- VIC Driver version 5.4.0.x or later
- UCS M5 B-Series or C-Series servers with VIC 1400 Series adapters: only Cisco UCS VIC 1400 Series or VIC 15000 series adapters are supported.

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

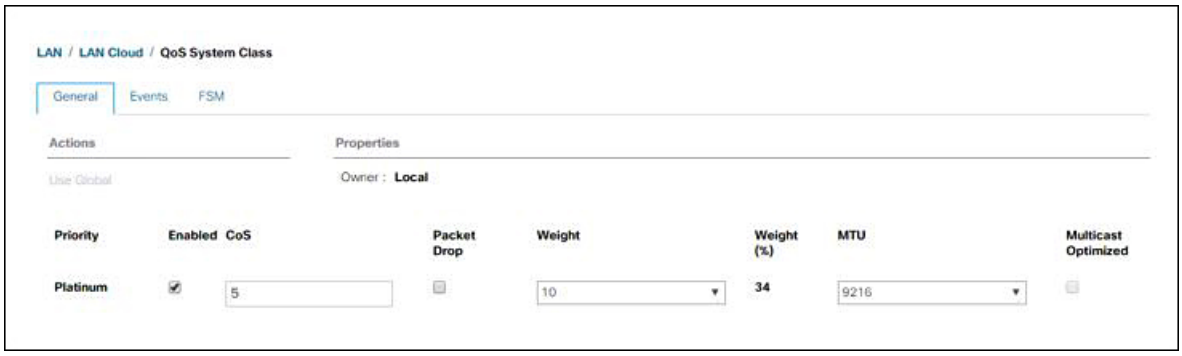
All Powershell commands or advanced property configurations are common across Windows 2019 and 2022 unless explicitly mentioned.

Configuring SMB Direct Mode 1 on UCS Manager

To avoid possible RDMA packet drops, make sure same no-drop COS is configured across the network.

Before you begin

Configure a no-drop class in UCSM QoS Policies and use it for RDMA supported interfaces. Go to **LAN** > **LAN Cloud** > **QoS System Class** and enable **Priority** Platinum with CoS 5.



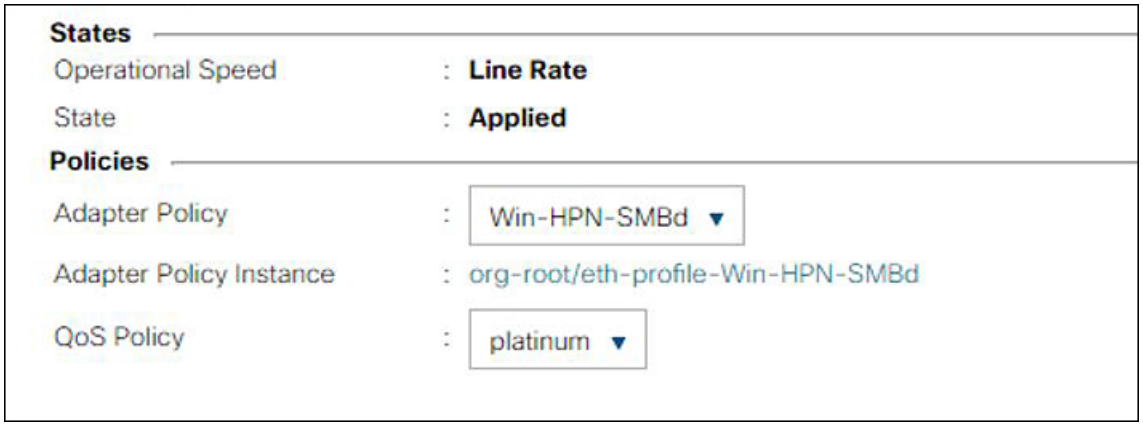
Step 1 In the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Policies**.

Step 3 Expand the node for the organization where you want to create the policy.

If the system does not include multitenancy, expand the **root** node.

Step 4 Expand **Adapter Policies** and choose the existing adapter policy for Win-HPN-SMBd.

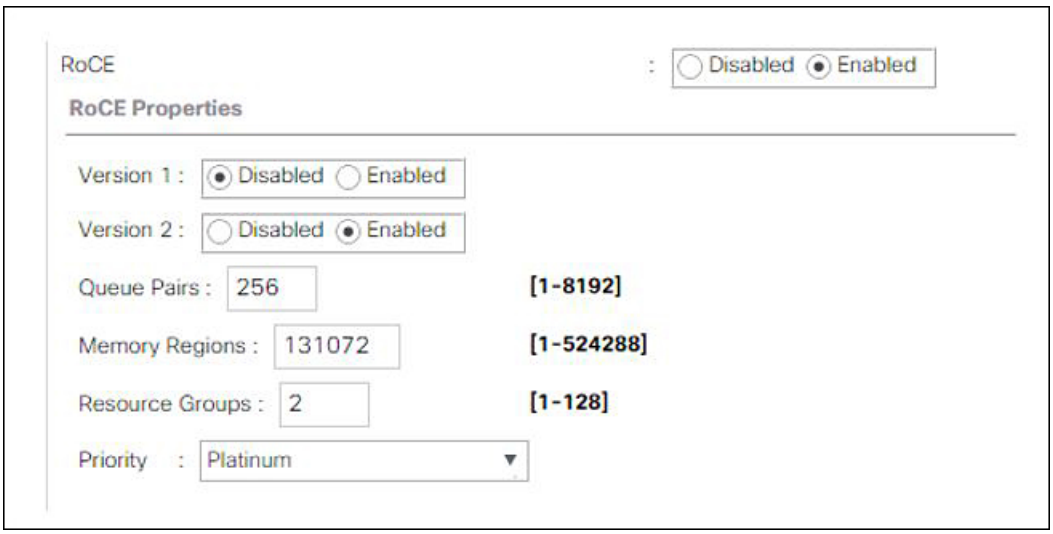


If using a user-defined adapter policy, use the configuration steps below.

a) On the **General** tab, scroll down to **RoCE** and click the **Enabled** radio button.
b) In the **RoCE Properties** field, under **Version 1**, click the **Disabled** radio button. For **Version 2**, click the **Enabled** radio button.
c) For **Queue Pairs**, enter `256`.
d) For **Memory Regions**, enter `131072`.

e) For **Resource Groups**, enter 2.
f) For **Priority**, choose **Platinum No-Drop COS.** from the dropdown.

This setting assumes you are using the default No-Drop policy.



g) Click **Save Change**s.

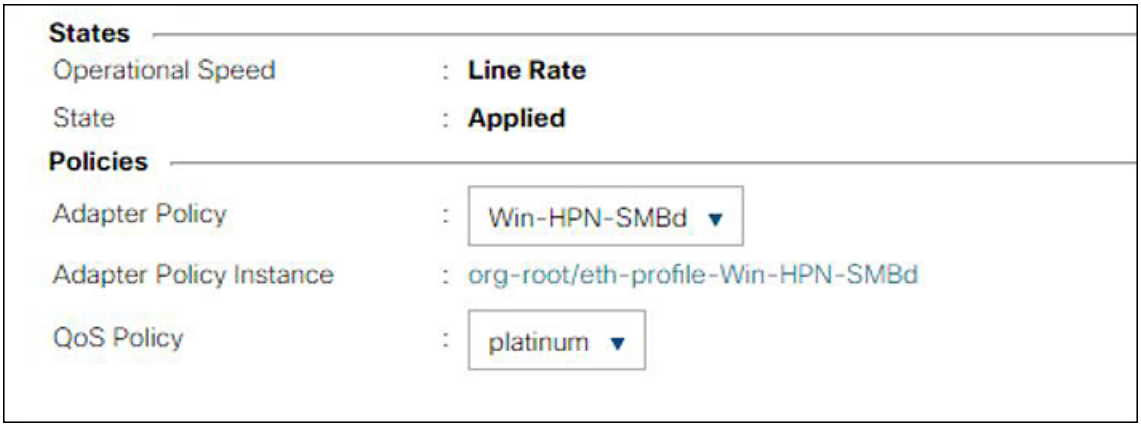
Step 5 Next, create an Ethernet Adapter Policy. In the Navigation pane, click **LAN**.

Step 6 Expand **LAN** > **Policies**.

Step 7 Right-click the **vNIC Templates** node and choose **Create vNIC Template**.

Step 8 Go to **vNIC Properties** under the General tab and modify the vNIC policy settings as follows:

a) Set **MTU** to **`1500`** or **`4096`**.
b) For the Adapter Policy, select **`Win-HPN-SMBd`**
c) For the **QoS policy**, specify **`Platinum`**.



Step 9 Click **Save Changes**.

Step 10 After you save the changes, UCS Manager will prompt you to reboot. Reboot the system.

What to do next

When the server comes back up, configure RoCEv2 mode 1 on the Host.

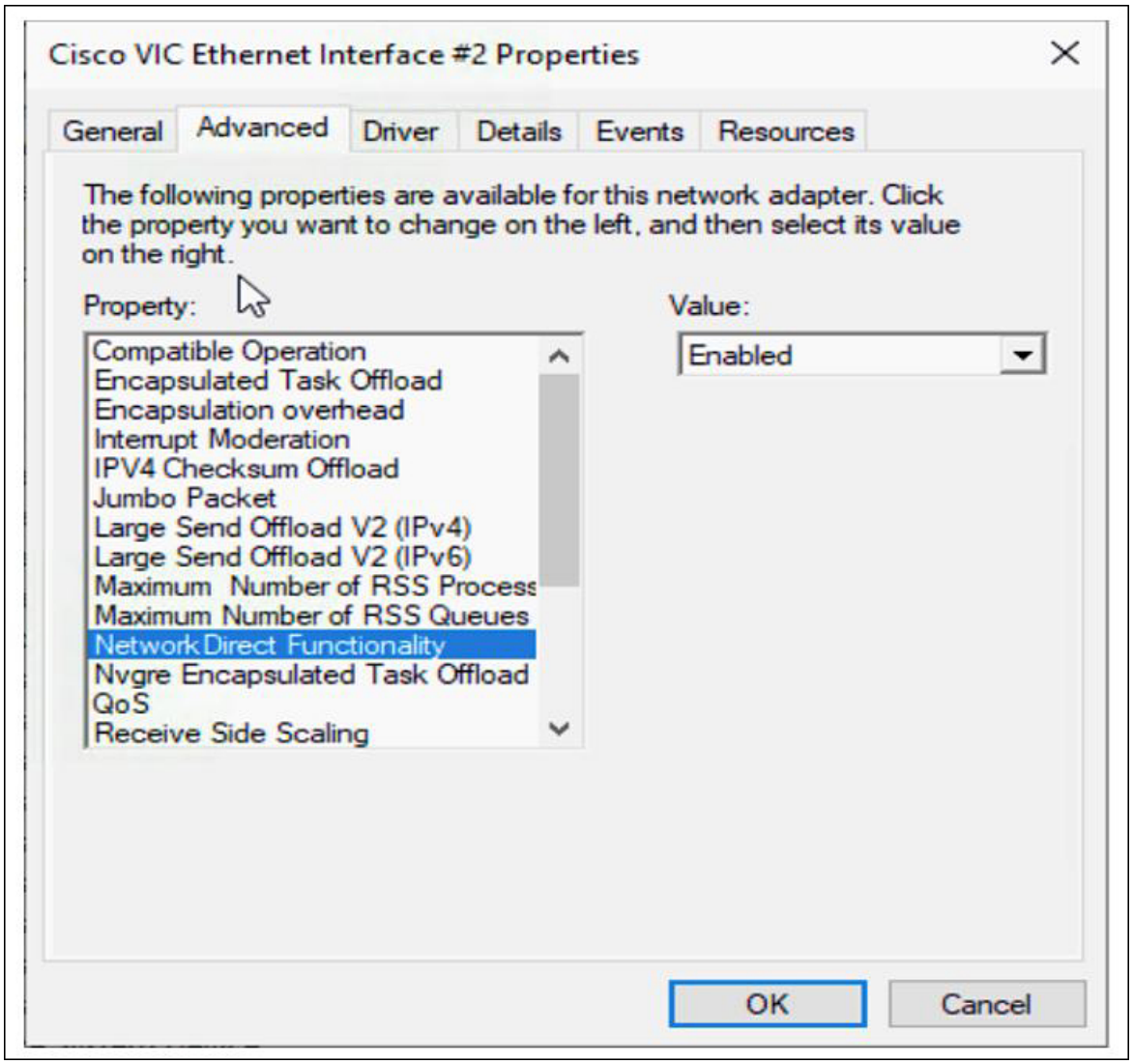
Configuring SMB Direct Mode 1 on the Host System

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 RoCEv2-enabled vNIC as described below.

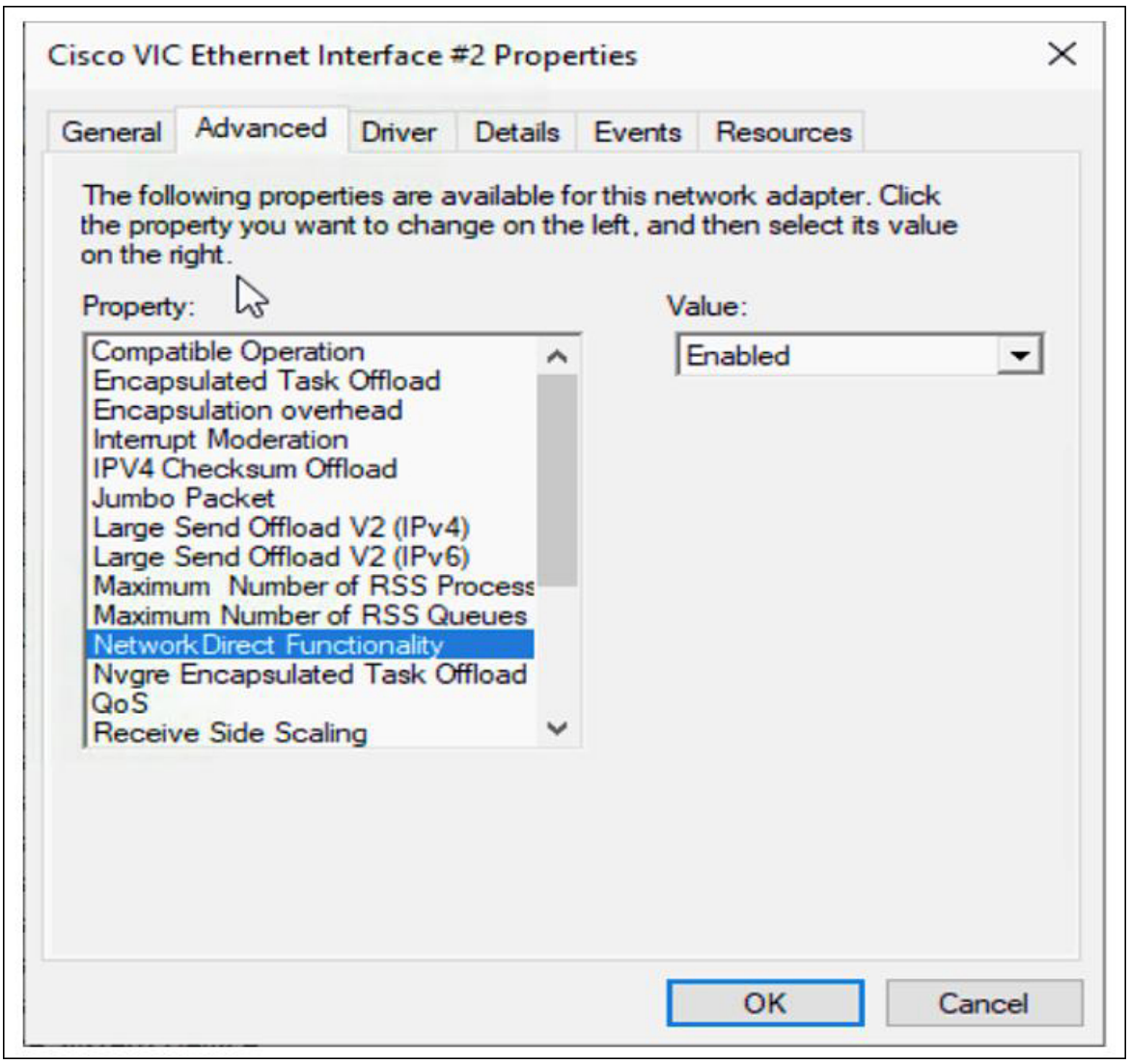
Before you begin

Configure RoCEv2 for Mode 1 in UCS Manager.

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 (vp1)
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"]**

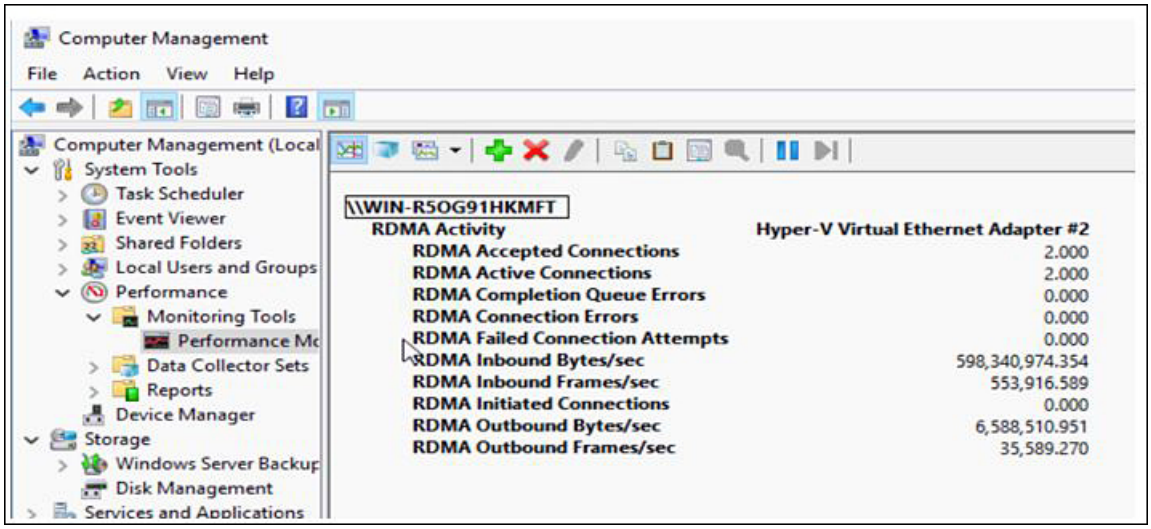
Step 6 Verify the overall RoCEv2 Mode 1 configuration at the Host as follows:

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

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) 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 RoCEv2 mode1 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
```

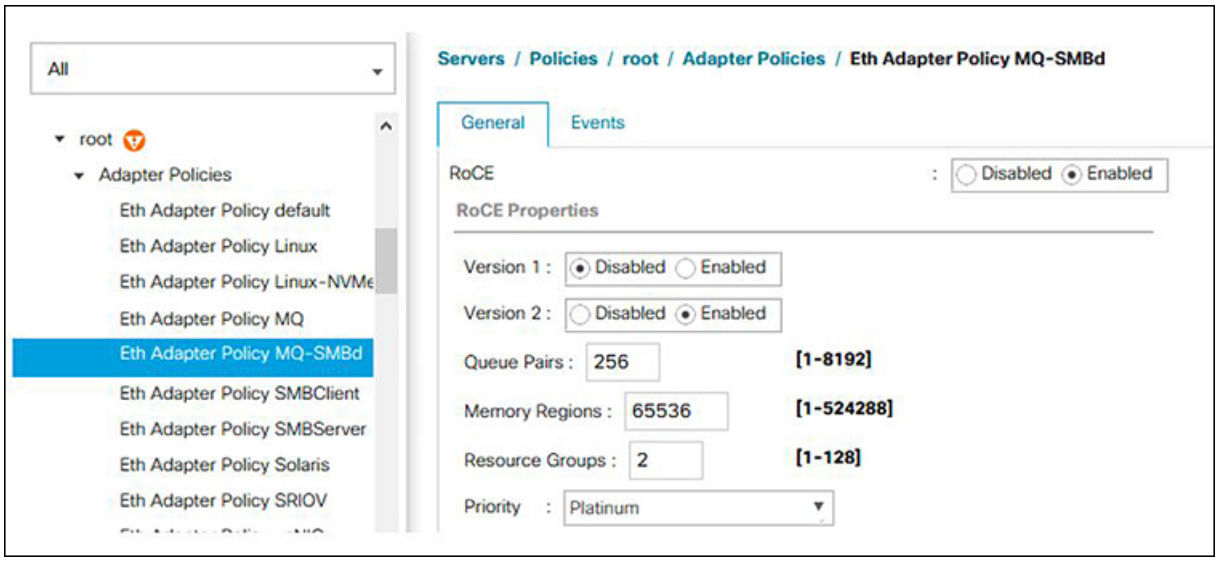
Configuring Mode 2 on UCS Manager

You will apply the VMQ Connection Policy as vmmq.

Before you begin

Configure RoCEv2 Policies in Mode 1.

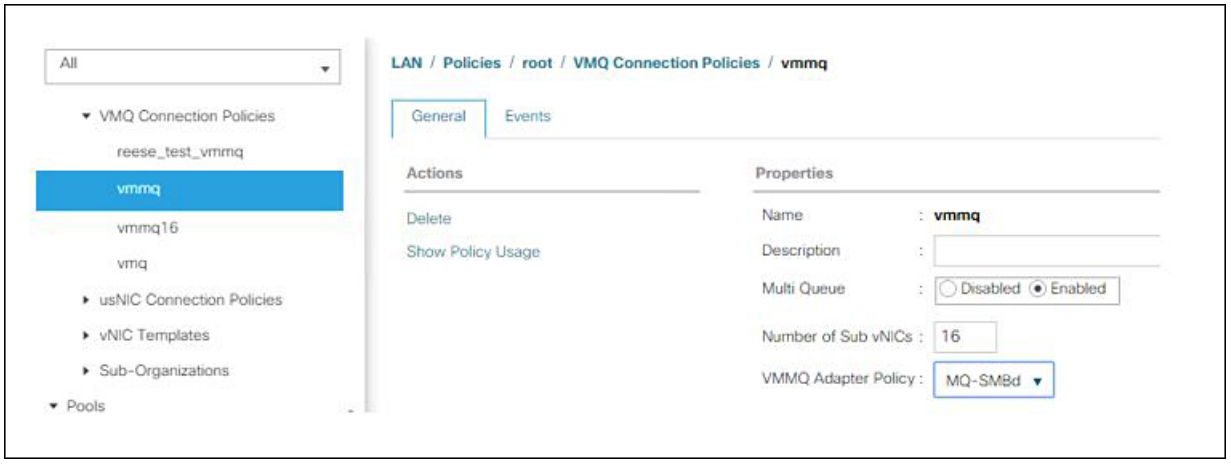
Use the pre-defined default adapter policy “MQ-SMBd”, or configure a user-defined Ethernet adapter policy with the following recommended RoCE-specific parameters:



- RoCE: Enabled
- Version 1: disabled
- Version 2: enabled
- Queue Pairs: 256
- Memory Regions: 65536
- Resource Groups: 2
- Priority: Platinum

Create a VMQ connection policy with the following values:

- Multi queue : Enabled
- Number of sub-vNIC: 16
- VMMQ adapter policy: MQ-SMBd



Step 1 In the **Navigation** pane, click **Servers**.

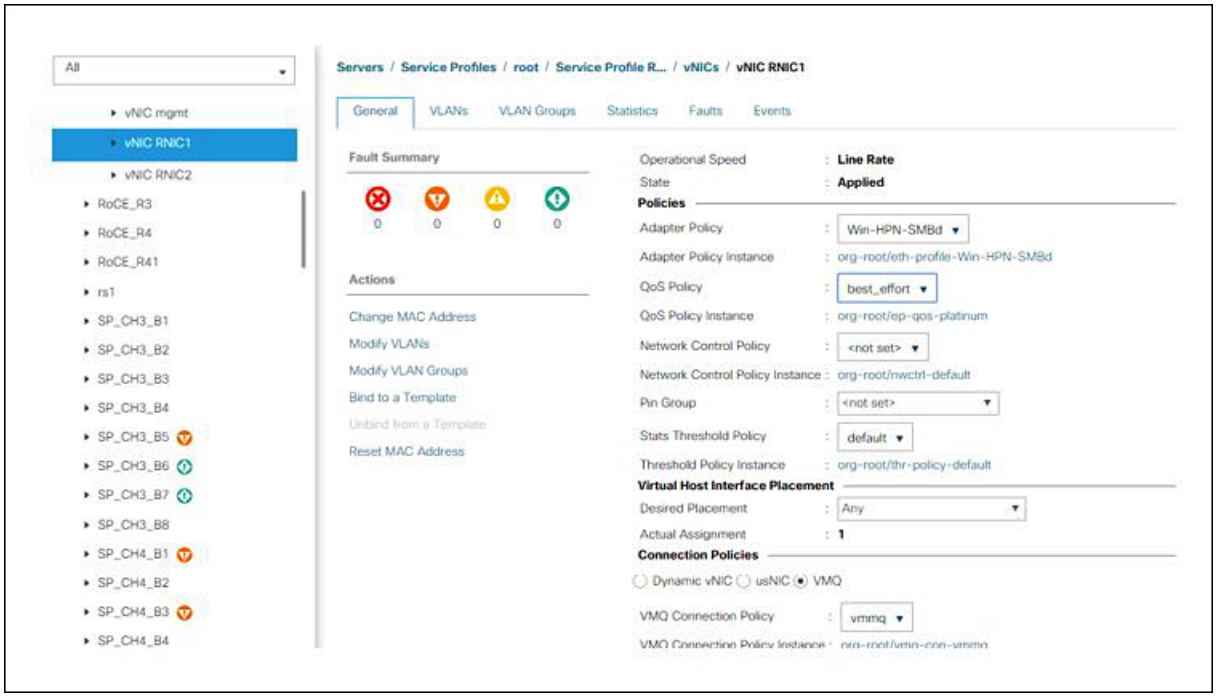
Step 2 Expand **Servers** > **Service Profiles**.

Step 3 Expand **Service Profiles** > **vNICs** and choose the VMQ Connection policy profile to configure.

Step 4 Go to **vNIC Properties** under the General tab and scroll down to the Policies area. Modify the vNIC policy settings as follows:

a) For the Adapter Policy, make sure it uses **`Win-HPN-SMBd`** or the adapter policy configured earlier for Mode 1.

b) For the **QoS policy**, select **`best-effort`**.

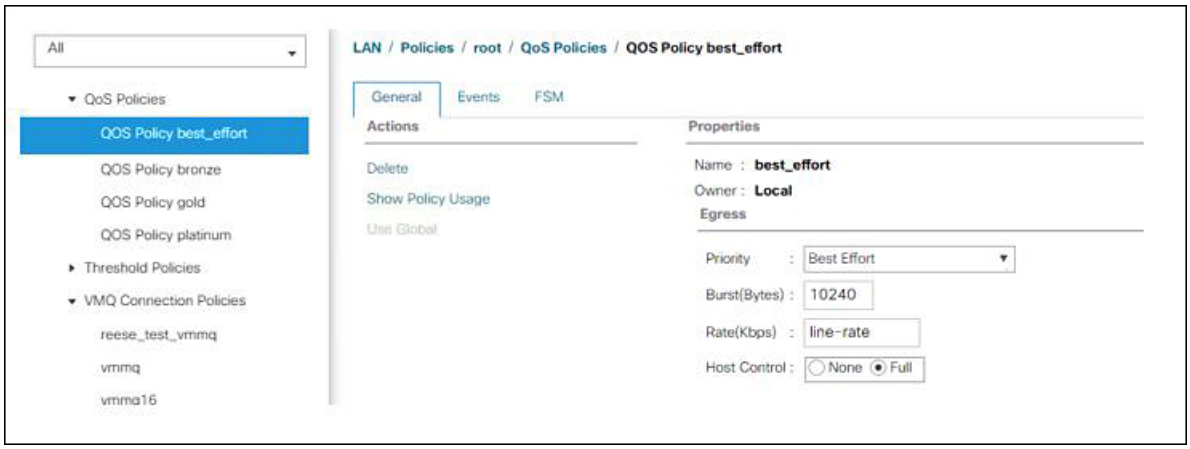


Step 5 Click **Save Changes**.

Step 6 In the Navigation pane, click **LAN**.

Step 7 Expand **LAN** > **Policies** > **QoS Policy Best Effort**.

Step 8 Set **Host Control** to **`Full`**.



Step 9 Click **Save Changes**.

Step 10 After you save the changes, UCS Manager will prompt you to reboot. Reboot the interface.

What to do next

When the server comes back up, configure Mode 2 on the Host.

Configuring Mode 2 on the Host System

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

Before you begin

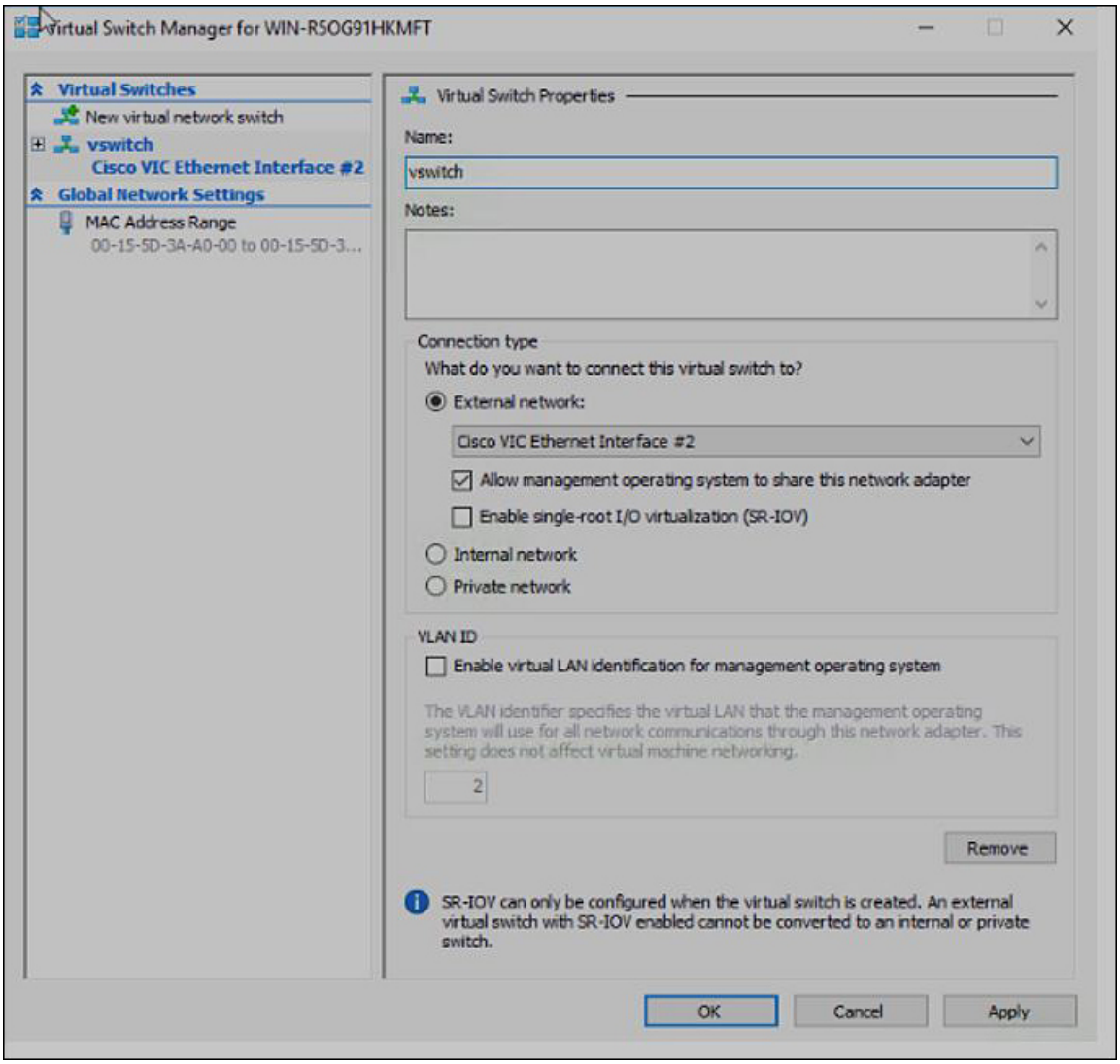
- Configure and confirm the connection for Mode 1 for both the UCS Manager and Host.
- Configure Mode 2 in UCS Manager.

Step 1 Go the the Hyper-V switch manager.

Step 2 Create a new Virtual Network Switch (vswitch) for theRoCEv2-enabled Ethernet interface.

a) Choose **External Network** and select **VIC Ethernet Interface 2** and **Allow management operating system to share this network adapter**.

b) Click **OK** to create the 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 setswtch -name svp1
```

You will see the new vport when you again enter

```
get-netadapterrdma
```

b) Add a vport.

```
add-vmNetworkAdapter -switchname setswtch -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.

a) For smb-client and smb-server in the host system, configure the RoCEv2-enabled vNIC as described above.

b) 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.

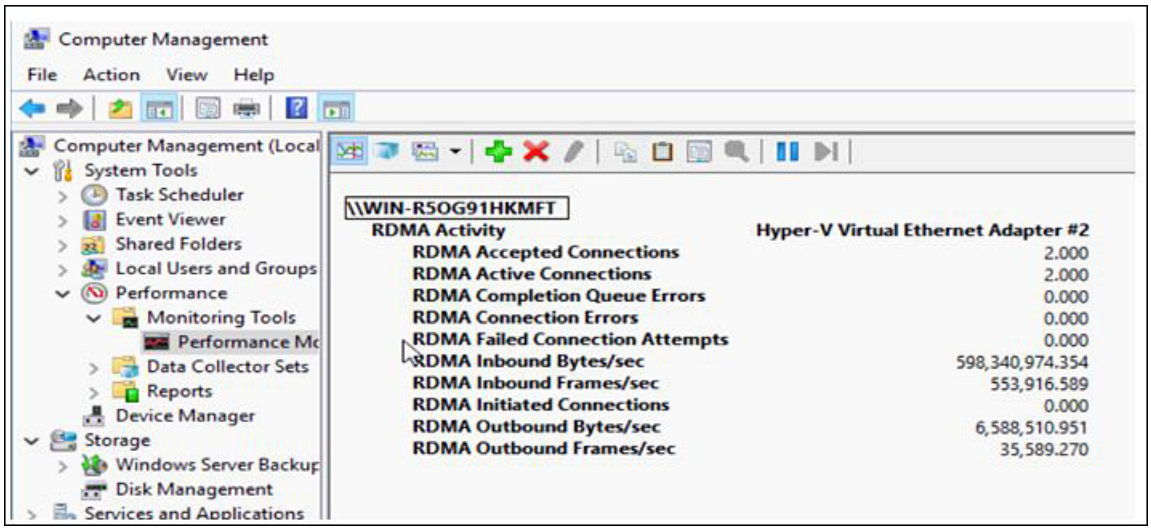
c) Create a share in smb-server and map the share in the smb-client.

Step 6 Finally, 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.

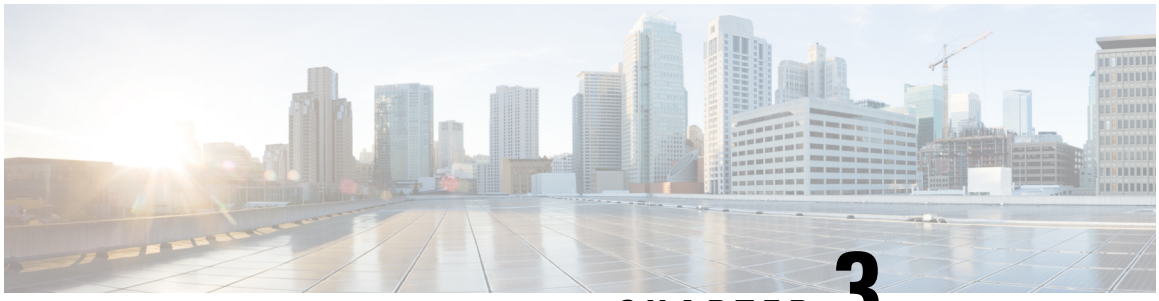


c) Issue the **netstat -xan** command again and check for the connection entries to verify they are displayed.

```
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.



CHAPTER 3

Configuring NVMe Over Fabrics (NVMeoF) with RoCEv2 in Linux

- Guidelines for using NVMe over Fabrics (NVMeoF) with RoCEv2 on Linux, on page 19
- Linux Requirements, on page 20
- Configuring RoCEv2 for NVMeoF on UCS Manager, on page 21
- Configuring RoCEv2 for NVMeoF on the Host System, on page 22
- Setting Up Device Mapper Multipath, on page 25
- Deleting the RoCEv2 Interface Using UCS Manager, on page 26

Guidelines for using NVMe over Fabrics (NVMeoF) with RoCEv2 on Linux

General Guidelines and Limitations:

- Cisco recommends you check UCS Hardware and Software Compatibility specific to your UCS Manager release to determine support for NVMeoF. NVMeoF is supported on UCS M5 and later B-Series and C-Series servers.
- NVMe over RDMA with RoCEv2 is supported with the fourth generation Cisco UCS VIC 1400 Series UCS VIC 14000, and UCS VIC 15000 Series and UCS VIC 15000 Series adapters. NVMe over RDMA is not supported on UCS 6324 Fabric Interconnects or on UCS VIC 1200 Series and 1300 Series adapters.
- When creating RoCEv2 interfaces, use Cisco UCS Manager provided Linux-NVMe-RoCE adapter policy.

Note Do not use the default Linux Adapter policy with RoCEv2; RoCEv2 interfaces will not be created in the OS.

- When configuring RoCEv2 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.
- RoCEv2 supports maximum two RoCEv2 enabled interfaces per adapter.
- Booting from an NVMeoF namespace is not supported.

- Layer 3 routing is not supported.
- RoCEv2 does not support bonding.
- Saving a crashdump to an NVMeoF namespace during a system crash is not supported.
- NVMeoF cannot be used with usNIC, VMFEX, VxLAN, VMQ, VMMQ, NVGRE, GENEVE Offload, and DPDK features.
- Netflow monitoring is not supported on RoCEv2 interfaces.
- In the Linux-NVMe-RoCE 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.
- 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.
- Set MTU size correctly on the VLANs and QoS policy on 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.
- UCS Manager does not support fabric failover for vNICs with RoCEv2 enabled.

Interrupts

- Linux RoCEv2 interface supports only MSIx interrupt mode. Cisco recommends avoiding changing interrupt mode when the interface is configured with RoCEv2 properties.
- The minimum interrupt count for using RoCEv2 with Linux is 8.

Downgrade Limitations:

- Cisco recommends you remove the RoCEv2 configuration before downgrading to any non-supported RoCEv2 release.

Linux Requirements

Configuration and use of RoCEv2 in Linux requires the following:

- InfiniBand kernel API module ib_core
- UCS Manager release 4.1.1 or later
- Minimum VIC firmware 5.1(1x) for IPv4 support and 5.1(2x) for IPv6 support
- UCS M5 and later B or C-series servers with Cisco UCS VIC 1400 or 15000 Series adapters
- eNIC driver version 4.0.0.6-802-21 or later provided with the 4.1.1 release package
- enic_rdma driver version 1.0.0.6-802-21 or later provided with the 4.1.1 release package

Note Use eNIC driver version 4.0.0.10-802.34 or later and enic_rdma driver version 1.0.0.10-802.34 or later for IPv6 support.

- A storage array that supports NVMeoF connection

Configuring RoCEv2 for NVMeoF on UCS Manager

Use these steps to configure the RoCEv2 interface on UCS Manager.

Step 1 In the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Service Profiles**.

Step 3 Expand the node for the organization where you want to create the policy.

If the system does not include multitenancy, expand the **root** node.

Step 4 Click on **vNICs** and go to the **Network** tab in the work area.

Modify the vNIC policy, according to the steps below.

a) On the **Network** tab, scroll down to the desired vNIC and click on it, then click **Modify**.
b) A popup dialog box will appear. Scroll down to the **Adapter Performance Profile** area, and click on the dropdown area for the Adapter Policy. Choose **Linux-NVMe-RoCE** from the drop-down list.
c) Click **OK**.

Step 5 Click **Save Changes**.

Step 6 Select **Reboot**.

Enabling an SRIOV BIOS Policy

Use these steps to configure the server's service profile with the RoCE v2 vNIC and enable the SRIOV BIOS policy before enabling the IOMMU driver in the Linux kernel.

Step 1 In the the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Service Profiles**

Step 3 Expand the node for the organization where you want to create the policy.

If the system does not include multitenancy, expand the **root** node.

Step 4 Select the service profile node where you want to enable SRIOV.

Step 5 In the Work pane, select **Policies** tab.

Step 6 In the Policies Area, expand **BIOS Policy**.

Step 7 Choose the default SRIOV policy from the **BIOS Policy** drop-down list.

Step 8 Click **Save Changes**.

Configuring RoCEv2 for NVMeoF on the Host System

Before you begin

Configure the server's service profile with RoCEv2 vNIC and the SRIOV enabled BIOS policy.

Step 1 Open the `/etc/default/grub` file for editing.

Step 2 Add `intel_iommu=on` to the end of the line for `GRUB_CMDLINE_LINUX`as shown in the sample file below.

```
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, run the following command to 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 that the server booted with the `intel_iommu=on` option by checking the output file.

```
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=/vmlinuz-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.

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:

```
# dmesg | grep enic_rdma
[    4.025979] enic_rdma: Cisco VIC Ethernet NIC RDMA Driver, ver 1.0.0.6-802.21 init
[    4.052792] enic 0000:62:00.1 eth1: enic_rdma: IPv4 RoCEv2 enabled
[    4.081032] enic 0000:62:00.2 eth2: enic_rdma: IPv4 RoCEv2 enabled
```

Step 4 Load the vme-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, please go to the section for Setting Up Device Mapper Multipath.

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.

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

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 RoCEv2 Interface Using UCS Manager

Use these steps to remove the RoCE v2 interface.

Step 1 In the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Service Profiles**.

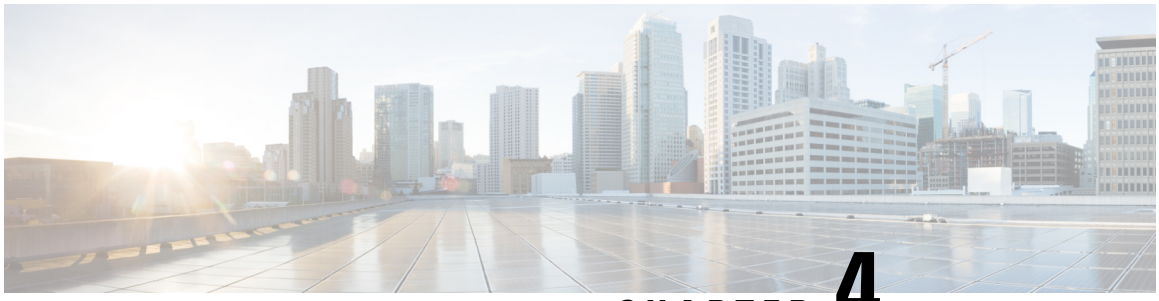
Step 3 Expand the node for the organization where you want to create the policy. If the system does not include multitenancy, expand the **root** node

Step 4 Click on **vNICs** and go to the **Network** tab in the work area.

Modify the vNIC policy, according to the steps below.

a) On the **Network** tab, scroll down to the desired vNIC and click on it, then click **Modify**.
b) A popup dialog box will appear. Scroll down to the **Adapter Performance Profile** area, and click on the dropdown area for the Adapter Policy. Choose **Linux** from the drop-down list.
c) Click **OK**.

Step 5 Click **Save Changes**.



CHAPTER 4

Configuring NVMe with ROCE V2 in ESXi

- Guidelines for using RoCEv2 Protocol in the Native ENIC driver on ESXi, on page 27
- ESXi nENIC RDMA Requirements, on page 28
- Installing NENIC Driver, on page 28
- Configuring and Enabling RoCEv2 on UCS Manager, on page 29
- Configuring RoCEv2 for VMware NVMeoF on UCS Manager, on page 29
- ESXi NVMe RDMA Host Side Configuration, on page 30
- NENIC RDMA Functionality, on page 30
- Create Network Connectivity Switches, on page 31
- Create VMHBA Ports in ESXi, on page 33
- Displaying vmnic and vmrdma Interfaces, on page 35
- NVMe Fabrics and Namespace Discovery, on page 36
- Deleting the ESXi RoCEv2 Interface Using UCS Manager, on page 37

Guidelines for using RoCEv2 Protocol in the Native ENIC driver on ESXi

General Guidelines and Limitations:

- Cisco UCS Manager release 4.2(3b) supports RoCEv2 only on ESXi 7.0 U3.
- Cisco recommends you check UCS Hardware and Software Compatibility specific to your UCS Manager release to determine support for ESXi. RoCEv2 on ESXi is supported on UCS B-Series and C-Series servers with Cisco UCS VIC 15000 Series and later adapters.
- RoCEv2 on ESXi is not supported on UCS VIC 1200, 1300 and 1400 Series adapters.
- RDMA on ESXi nENIC currently supports only ESXi NVME that is part of the ESXi kernel. The current implementation does not support the ESXi user space RDMA application.
- Multiple mac addresses and multiple VLANs are supported only on VIC 15000 Series adapters.
- RoCEv2 supports maximum two RoCEv2 enabled interfaces per adapter.
- Pvrdma, VSAN over RDMA, and iSER are not supported.
- The COS setting is not supported on UCS Manager.

Downgrade Limitations:

- Cisco recommends you remove the RoCEv2 configuration before downgrading to any non-supported RoCEv2 release.

ESXi nENIC RDMA Requirements

Configuration and use of RoCEv2 in ESXi requires the following:

- VMWare ESXi version 7.0 U3.
- UCS Manager release 4.2.3 or later
- Nenic-2.0.4.0-1OEM.700.1.0.15843807.x86_64.vib provides both standard eNIC and RDMA support.
- A storage array that supports NVMeoF connection. Currently, tested and supported on Pure Storage with Cisco Nexus 9300 Series switches.

Downgrade Limitations:

- Cisco recommends you remove the RoCEv2 configuration before downgrading to any non-supported RoCEv2 release.

Installing NENIC Driver

The enic drivers, which contain the rdma driver, are available as a combined package. Download and use the enic driver on cisco.com.

These steps assume this is a new installation.

Note While this example uses the /tmp location, you can place the file anywhere that is accessible to the ESX console shell.

Step 1 Copy the enic VIB or offline bundle to the ESX server. The example below uses the Linux **scp** utility to copy the file from a local system to an ESX server located at 10.10.10.10: and uses the location /tmp.

```
scp nenic-2.0.4.0-1OEM.700.1.0.15843807.x86_64.vib root@10.10.10.10:/tmp
```

Step 2 Specifying the full path, issue the command shown below.

```
esxcli software vib install -v {VIBFILE}
```

or

```
esxcli software vib install -d {OFFLINE_BUNDLE}
```

Here is an example:

```
esxcli software vib install -v /tmp/nenic-2.0.4.0-1OEM.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 being installed, you may need to put ESX into maintenance mode. This can be done through the VI Client, or by adding the `--maintenance-mode` option to the above `esxcli` command.

Upgrading NENIC Driver

a. To upgrade NENIC driver, enter the command:

```
esxcli software vib update -v {VIBFILE}
```

or

```
esxcli software vib update -d {OFFLINE_BUNDLE}
```

b. Copy the enic VIB or offline bundle to the ESX server using Step 1 given above.

What to do next

Create and configure the Adapter Policy for ESXi NVMe RDMA in UCS Manager.

Configuring and Enabling RoCEv2 on UCS Manager

Configuring RoCEv2 for VMware NVMeoF on UCS Manager

UCS Manager contains a default adapter policy that is prepopulated with operational parameters, so you do not need to manually create the adapter policy. However, you do need to create the RoCEv2 interface.

Use these steps to configure the RoCEv2 interface on UCS Manager.

Step 1 In the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Service Profiles**.

Step 3 Expand the node for the organization where you want to create the policy.

If the system does not include multitenancy, expand the **root** node.

Step 4 Click on a RDMA service profile you created and expand the service profile.

Step 5 Right-click on **vNICs** and choose **Create vNIC** to create a new vNIC.

Step 6 Click on a RDMA service profile you created with the Service Policy and scroll down to **vNICs**. Right-click and choose **Create** to create a new vNIC.

The **Create VNIC** pop-up menu is displayed.

Perform the below steps to modify the vNIC policy:

a) Name the new VNIC.
b) On the **MAC address** dropdown, select the desired address or use the default in the dropdown.
c) Select which VLAN you want use use from the list.
d) In the Adapter Performance Profile, select the default adapter policy named `VMWareNVMeRoCEv2`.

e) Click **OK**. The interface is now configured for one port.

Step 7 Click **Save Changes**.

Step 8 Select **Reboot**.

What to do next

Configure the Host side for ESXi NVMe RDMA.

ESXi NVMe RDMA Host Side Configuration

NENIC RDMA Functionality

One major difference exists between the use case for RDMA on Linux and ESXi.

- In ESXi, the physical interface (vmnic) MAC is not used fo 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 that you have downloaded and installed the enic drivers.

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 trafic.
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.

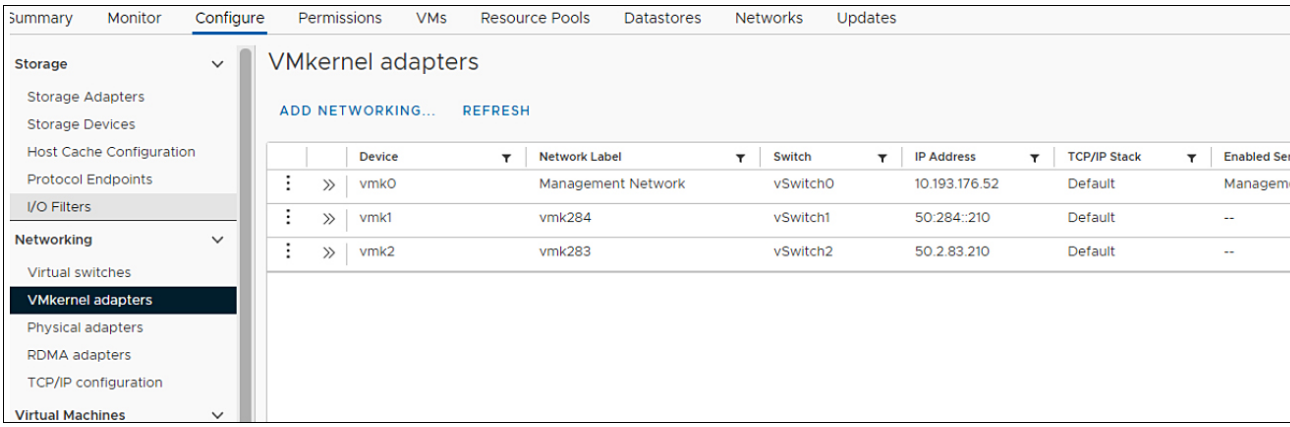
f) Configure IPV4 and/or IPV6 settings.

Step 8 On the **Ready to Complete** page, click **Finish**.

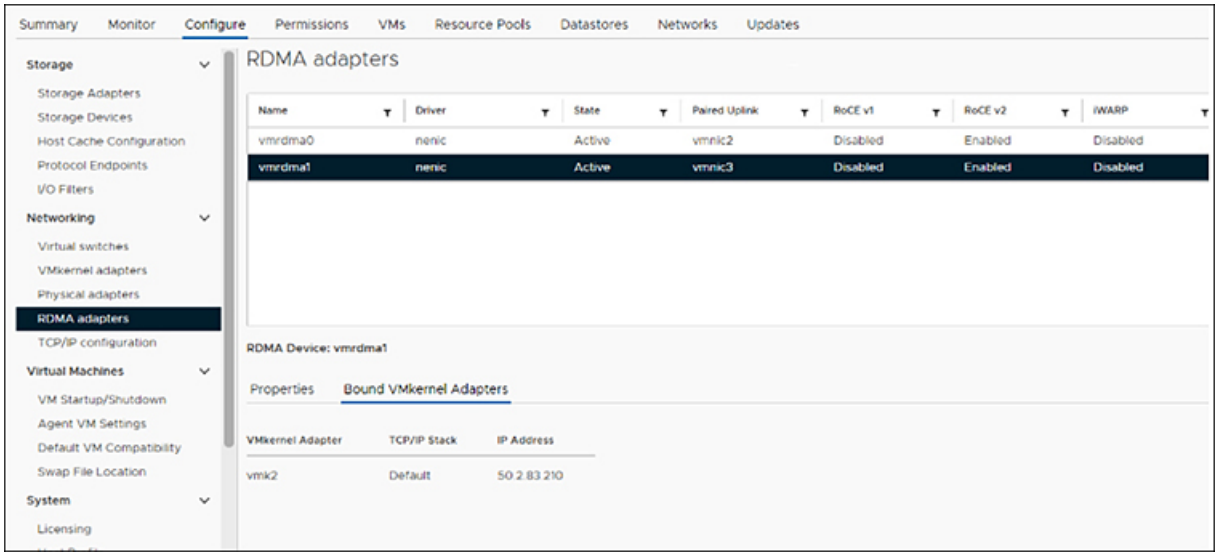
Step 9 Check the VMkernel ports for the VM Adapters or port groups with NVMe RDMA in the vSphere client, as shown in the Results below.

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

Example



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



What to do next

Create vmhba ports on top of vmrdma ports.

Create VMHBA Ports in ESXi

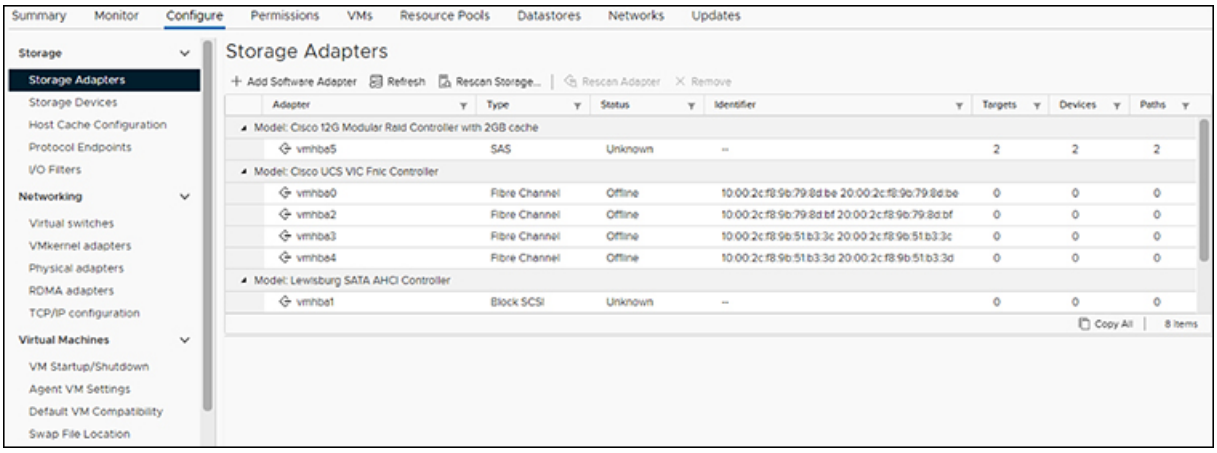
Use the following steps for creating vmhba ports on top of the vmrdma adapter ports.

Before you begin

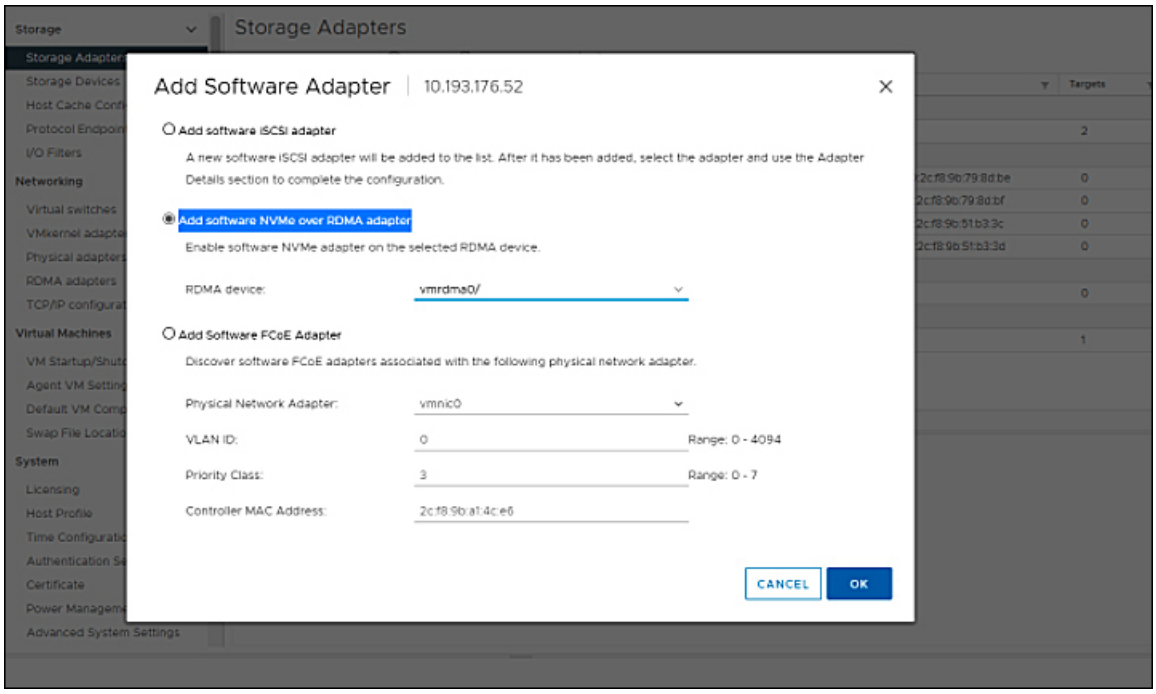
Create the adapter ports for storage connectivity.

Step 1 Go to vCenter where your ESXi host is connected.

Step 2 Click on **Host**>**Configure**>**Storage adapters**.



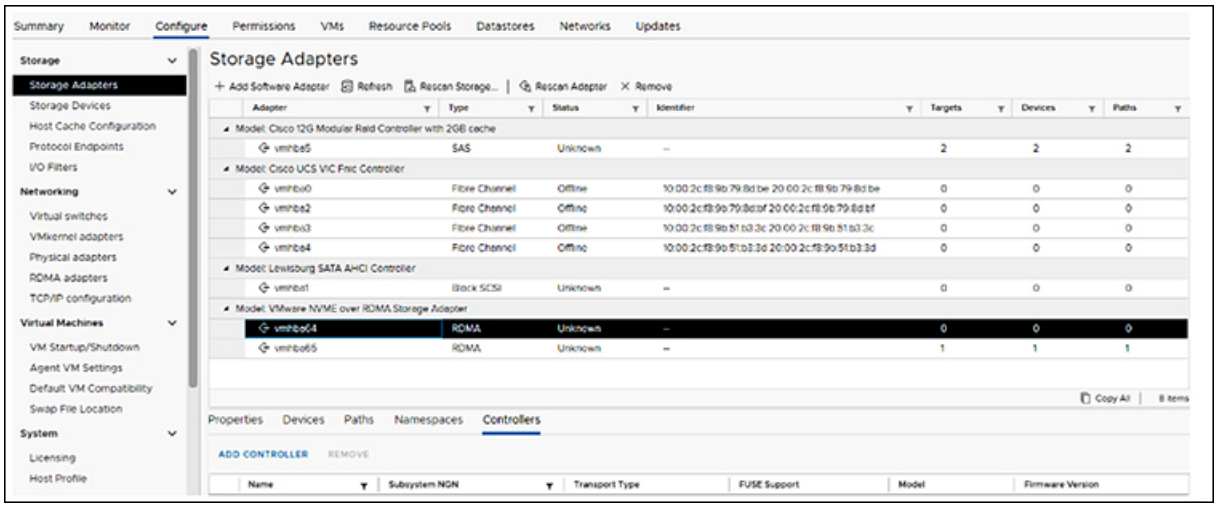
Step 3 Click **+Add Software Adapter**. The following dialog box will appear.



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



What to do next

Configure NVME.

Displaying vmnic and vmrdma Interfaces

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

Before you begin

Create Network Adapters and VHBA ports.

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   50000Mbps  Full   2c:f8:9b:79:8d:bc 1500   Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic3  0000:1d:00.1 nenic       Up   50000Mbps  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
```

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.

Step 3 Use **esxcli rdma device list** to list the vmrdma devices.

```
[root@StockholmRackServer:~] esxcli rdma device list
Name     Driver  State   MTU  Speed    Paired Uplink  Description
-------  ------  ------  ----  -------  -------------  -----------
vmrdma0  nenic   Active  4096  50 Gbps  vmnic1         Cisco UCS VIC 15XXX (A0)
vmrdma1  nenic   Active  4096  50 Gbps  vmnic2         Cisco UCS VIC 15XXX (A0)
[root@StockholmRackServer:~] esxcli rdma device vmknic list
Device   Vmknic  NetStack
-------  ------  --------
vmrdma0  vmk1    defaultTcpipStack
vmrdma1  vmk2    defaultTcpipStack
```

Step 4 Use **esxcli rdma device list** to check the protocols supported by the vmrdma interface.

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

Step 5 Use **esxcli rdma device protocol list** to check the protocols supported by the vmrdma interface.

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

```
[root@ESXi7U3Bodega:~] esxcli rdma device protocol list
Device   RoCE v1  RoCE v2  iWARP
-------  -------  -------  -----
vmrdma0    false     true  false
vmrdma1    false     true  false
[root@ESXi7U3Bodega:~]
```

Step 6 Use **esxcli nvme adapter list** to list the NVMe adapters and the vmrdma and vmnic interfaces it is configured on.

```
[root@ESXi7U3Bodega:~] esxcli nvme adapter list
Adapter  Adapter Qualified Name         Transport Type  Driver    Associated Devices
-------  -----------------------------  --------------  --------  ------------------
vmhba64  aqn:nvmerdma:2c-f8-9b-79-8d-bc  RDMA            nvmerdma  vmrdma0, vmnic2
vmhba65  aqn:nvmerdma:2c-f8-9b-79-8d-bd  RDMA            nvmerdma  vmrdma1, vmnic3
[root@ESXi7U3Bodega:~]
```

Step 7 All vmhbas in the system can be listed using **esxcli storage core adapter list**.

```
[root@ESXi7U3Bodega:~] esxcli storage core adapter list
HBA Name  Driver    Link State  UID                                   Capabilities         Description
--------  --------  ----------  ------------------------------------  -------------------  -----------
vmhba0    nfnic     link-down   fc.10002cf89b798dbe:20002cf89b798dbe  Second Level Lun ID  (0000:1d:00.2) Cisco Corporation Cisco UCS VIC Fnic Controller
vmhba1    vmw_ahci  link-n/a    sata.vmhba1                                                (0000:00:11.5) Intel Corporation Lewisburg SATA AHCI Controller
vmhba2    nfnic     link-down   fc.10002cf89b798dbf:20002cf89b798dbf  Second Level Lun ID  (0000:1d:00.3) Cisco Corporation Cisco UCS VIC Fnic Controller
vmhba3    nfnic     link-down   fc.10002cf89b51b33c:20002cf89b51b33c  Second Level Lun ID  (0000:63:00.2) Cisco Corporation Cisco UCS VIC Fnic Controller
vmhba4    nfnic     link-down   fc.10002cf89b51b33d:20002cf89b51b33d  Second Level Lun ID  (0000:63:00.3) Cisco Corporation Cisco UCS VIC Fnic Controller
vmhba5    lsi_mr3   link-n/a    sas.5cc167e9732f9b00                                       (0000:3c:00.0) Broadcom Cisco 12G Modular Raid Controller with 2GB cache
vmhba64   nvmerdma  link-n/a    rdma.vmnic2:2c:f8:9b:79:8d:bc                              VMware NVMe over RDMA Storage Adapter on vmrdma0
vmhba65   nvmerdma  link-n/a    rdma.vmnic3:2c:f8:9b:79:8d:bd                              VMware NVMe over RDMA Storage Adapter on vmrdma1
[root@ESXi7U3Bodega:~]
```

What to do next

Configure NVME.

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.

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.

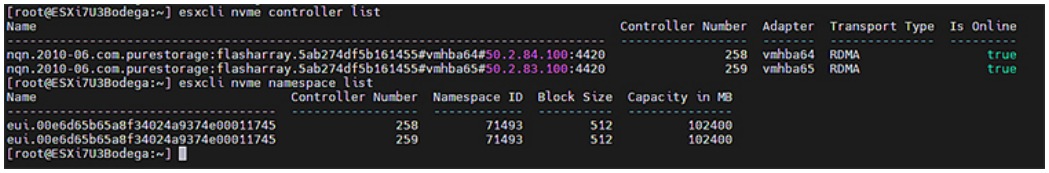
```
esxcli nvme fabrics discover -a vmhba64 -l transport_address p Transport Service ID -s Subsystem NQN
```

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
```



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 ESXi RoCEv2 Interface Using UCS Manager

Use these steps to remove the RoCE v2 interface for a specific port.

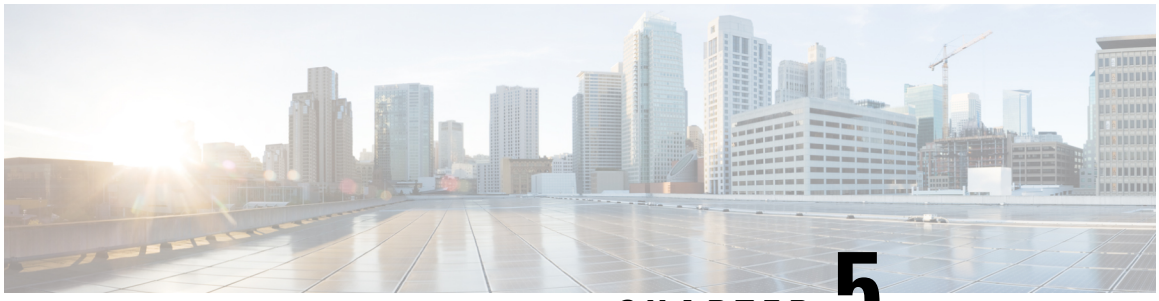
Step 1 In the **Navigation** pane, click **Servers**.

Step 2 Expand **Servers** > **Service Profiles**.

Step 3 Expand the node for the profile to delete.

Step 4 Click on **vNICs** and select the desired interface. Right click and select **Delete** from the dropdown.

Step 5 Click **Save Changes**.



CHAPTER 5

Using the UCS Manager CLI to Configure the RoCEv2 Interface

- Configure Windows SMBDirect RoCEv2 Interface using UCS Manager CLI, on page 39
- Configuring the Linux RoCEv2 Interface Using the UCS Manager CLI, on page 40
- Deleting the Windows RoCEv2 Interface Using the CLI for UCS Manager, on page 41
- Deleting the Linux RoCEv2 Interface Using the UCS Manager CLI, on page 42
- Configuring the RoCEv2 VMware ESXi Interface Using the UCS Manager CLI, on page 43
- Deleting the ESXi RoCEv2 Interface Using the UCS Manager CLI, on page 44

Configure Windows SMBDirect RoCEv2 Interface using UCS Manager CLI

Use the following steps to configure the RoCEv2 interface in the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	To configure Windows SMBDirect RoCEv2 Mode 1: **Example:**	Specifies a Windows SMBDirect RoCEv2 adapter policy for RoCEv2 Mode 1.

	Command or Action	Purpose
	UCS-A /org/service-profile/vnic # set adapter-policy Win-HPN-SMBd	
Step 5	To configure Windows SMBDirect RoCEv2 Mode 2: **Example:** UCS-A# scope org UCS-A /org # create vmq-conn-policy policy name UCS-A /org/vmq-conn-policy* # set multi-queue enabled UCS-A /org/vmq-conn-policy* # set vmmq-sub-vnic-count 64 UCS-A /org/vmq-conn-policy* # set vmmq-adaptor-profile-name MQ-SMBd UCS-A /org/vmq-conn-policy* # commit-buffer UCS-A /org/vmq-conn-policy #	Configures Windows Mode 2, after creating a VMQ connection policy and assigning the adapter policy **MQ-SMBd**:
Step 6	**Example:** UCS-A /org/service-profile/vnic* # commit-buffer	Commit the transaction to the system configuration.

This example shows how to configure the RoCEv2 Win-HPN-SMBd adapter policy:

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:

Name    Fabric ID        Dynamic MAC Addr  Virtualization Preference
--------   -------------        ----------------------------
-----------------------------------
eth00          A B            00:25:B5:3A:84:00          NONE
eth01          A               00:25:B5:3A:84:01           NONE
eth02          B               00:25:B5:3A:84:02            NONE


UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy Win-HPN-SMBd
UCS-A /org/service-profile/vnic* # commit-buffer
UCS-A /org/service-profile/vnic #
```

Configuring the Linux RoCEv2 Interface Using the UCS Manager CLI

Use the following steps to configure the RoCEv2 interface for Linux in the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	**Example:** `UCS-A /org/service-profile/vnic # set adapter-policy Linux-NVMe-RoCE`	Specify Linux-NVMe-RoCE as the adapter policy for the vNIC that you want to use for NVMeoF.
Step 5	**Example:** `UCS-A /org/service-profile/vnic* # commit-buffer`	Commit the transaction to the system configuration.

This example shows how to configure the RoCEv2 Linux adapter policy on the eth01 vNIC:

Example

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:
    Name               Fabric ID Dynamic MAC Addr   Virtualization Preference
    ------------------ --------- ------------------ -------------------------
    eth00              A B       00:25:B5:3A:84:00  NONE
    eth01              A         00:25:B5:3A:84:01  NONE
    eth02              B         00:25:B5:3A:84:02  NONE
UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy Linux-NVMe-RoCE
UCS-A /org/service-profile/vnic* # commit-buffer
UCS-A /org/service-profile/vnic #
```

Deleting the Windows RoCEv2 Interface Using the CLI for UCS Manager

Use the following steps to delete the Windows RoCEv2 interface in the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	**Example:** `UCS-A /org/service-profile/vnic # set adapter-policy Windows`	Removes the Windows RoCEv2 adapter policy by setting the default Windows adapter policy.
Step 5	**Example:** `UCS-A /org/service-profile/vnic* # commit-buffer`	Commit the transaction to the system configuration.

What to do next

This example shows how to remove the RoCEv2 interface on the eth01 vNIC on Windows.

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:

Name    Fabric ID        Dynamic MAC Addr  Virtualization Preference
--------   -------------       ----------------------------
-----------------------------------
eth00          A B            00:25:B5:3A:84:00          NONE
eth01          A               00:25:B5:3A:84:01           NONE
eth02          B               00:25:B5:3A:84:02            NONE


UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy Windows
UCS-A /org/service-profile/vnic* # commit-buffer
UCS-A /org/service-profile/vnic #
```

Deleting the Linux RoCEv2 Interface Using the UCS Manager CLI

Use the following steps to delete the Linux RoCEv2 interface in the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	**Example:** `UCS-A /org/service-profile/vnic # set adapter-policy Linux`	Removes Linux-NVMe-RoCE policy by setting the default Linux adapter policy.
Step 5	**Example:** `UCS-A /org/service-profile/vnic* # commit-buffer`	Commit the transaction to the system configuration.

This example shows how to remove the RoCEv2 interface on the eth01 vNIC on Linux.

Example

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:
    Name               Fabric ID Dynamic MAC Addr   Virtualization Preference
    ------------------ --------- ------------------ -------------------------
    eth00              A B       00:25:B5:3A:84:00  NONE
    eth01              A         00:25:B5:3A:84:01  NONE
    eth02              B         00:25:B5:3A:84:02  NONE
UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy Linux
UCS-A /org/service-profile/vnic* # commit-buffer
```

Configuring the RoCEv2 VMware ESXi Interface Using the UCS Manager CLI

Use the following steps to configure the RoCEv2 interface for VMware ESXi in the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	**Example:** `UCS-A /org/service-profile/vnic # set adapter-policy VMWareNVMeRoCEv2`	Specify VMWareNVMeRoCEv2 as the adapter policy for the vNIC that you want to use for NVMeoF.
Step 5	**Example:** `UCS-A /org/service-profile/vnic* # commit-buffer`	Commit the transaction to the system configuration.

This example shows how to configure the RoCEv2 VMware adapter policy on the eth01 vNIC:

Example

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:
    Name               Fabric ID Dynamic MAC Addr   Virtualization Preference
    ------------------ --------- ------------------ -------------------------
    eth00              A B       00:25:B5:3A:84:00  NONE
    eth01              A         00:25:B5:3A:84:01  NONE
    eth02              B         00:25:B5:3A:84:02  NONE
UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy VMWareNVMeRoCEv2
UCS-A /org/service-profile/vnic* # commit-buffer
UCS-A /org/service-profile/vnic #
```

Deleting the ESXi RoCEv2 Interface Using the UCS Manager CLI

Use the following steps to delete the ESXi RoCEv2 interface using the Cisco UCS Manager CLI.

Before you begin

You must log in with admin privileges.

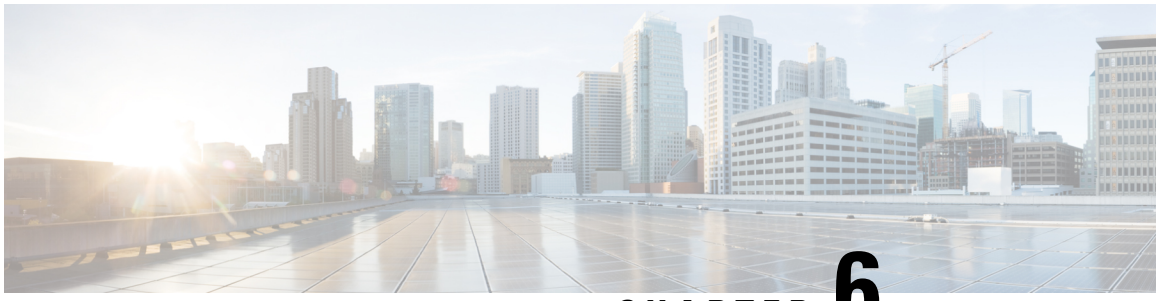
Procedure

	Command or Action	Purpose
Step 1	**Example:** `UCS-A # scope service-profile server chassis-id / blade-id or rack_server-id`	Enter the service profile for the specified chassis, blade or UCS managed rack server ID.
Step 2	**Example:** `UCS-A /org/service-profile # show vnic`	Display the vNICs available on the server.
Step 3	**Example:** `UCS-A /org/service-profile # scope vnic vnic name`	Enter the vnic mode for the specified vNIC.
Step 4	**Example:** `UCS-A /org/service-profile/vnic # set adapter-policy VMWare`	Removes `VMWareNVMeRoCEv2` policy by setting the default ESXi adapter policy.
Step 5	**Example:** `UCS-A /org/service-profile/vnic* # commit-buffer`	Commit the transaction to the system configuration.

This example shows how to remove the RoCEv2 interface on the eth01 vNIC on ESXi.

Example

```
UCS-A# scope service-profile server 1/1
UCS-A /org/service-profile # show vnic

vNIC:
    Name               Fabric ID Dynamic MAC Addr   Virtualization Preference
    ------------------ --------- ------------------ -------------------------
    eth00              A B       00:25:B5:3A:84:00  NONE
    eth01              A         00:25:B5:3A:84:01  NONE
    eth02              B         00:25:B5:3A:84:02  NONE
UCS-A /org/service-profile # scope vnic eth01
UCS-A /org/service-profile/vnic # set adapter-policy ESXi
UCS-A /org/service-profile/vnic* # commit-buffer
```



CHAPTER 6

Known Issues in RoCEv2

- Known-Issues in RoCEv2, on page 47

Known-Issues in RoCEv2

The following known issues are present in the RoCEv2 release.

Symptom	Conditions	Workaround

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 `Peak_cell` or `PeakQos` value for the port reaches more than 1000.	The NVMe traffic will drop.	To recover the switch from this error mode.
		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:
		`# clear counters`
		`# clear counter buffers module 1`
		`# clear qos statistics`
		4. Run **`no shutdown`** on the port that was shut down.

Symptom	Conditions	Workaround
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.	Case 1 : Installing Windows 2019 nenic driver on Windows 2016 succeeds-but on Windows 2016 RDMA is not supported. 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.	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/upgraded.