# cisco.



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

First Published: 2022-07-18

Last Modified: 2024-06-26

### **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 © 2022-2024 Cisco Systems, Inc. All rights reserved.

# **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.



# **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 a network protocol that allows for Remote Direct Memory Access (RDMA) over Ethernet networks. It enables low-latency and high-bandwidth communication between servers or storage systems by leveraging the benefits of RDMA technology. RoCEv2 eliminates the need for traditional TCP/IP networking stack overhead, resulting in improved performance and reduced latency. It allows for efficient data transfers and enables applications to directly access remote memory, enhancing overall network efficiency and scalability. RoCEv2 is often used in data centers and high-performance computing environments to optimize network performance and accelerate data-intensive workloads.

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



CHAPTER

# 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 10
- Configuring Mode 2 on Cisco Intersight, on page 13
- Configuring Mode 2 on the Host System, on page 16
- Deleting the RoCE v2 Interface Using Cisco Intersight, on page 19

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

**General Guidelines and Limitations:** 

 Cisco Intersight Managed Mode 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 Managed Mode domain profile.
- 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. Implementation of Network Direct Kernel Provider Interface (NDKPI) differs in Mode 1 and Mode 2 of operation: 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 both MSI & MSIx interrupts mode. Default interrupt mode is MSIx. 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 firmware 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.

- 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 domain profile.
  - 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 operating 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 operating 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.

• Cisco UCS M8 C-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 Interisght.

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

eisco	Intersight	ç Infrastructure Service 🗸		Q Search	0	¢1 1) Q 💷 (	A © 🕬
<ul><li>Ø Overvie</li><li>Ø Operation</li></ul>		Policies > System QoS					
Serveri Chassi Fabric I HyperF	s nterconnects lex Clusters	General     Policy Details	Policy Details Add policy details This policy is applicable only for Configure Priorities	r UCS Domains			
<ul> <li>Configu</li> <li>Profiles</li> <li>Templa</li> <li>Policies</li> </ul>	are ^		Platinum CoS 5	S ⇔ 10 D I	C o Allow Packer	MTU 8 <sub>0</sub> 9000	0 × 1000 - 8010
Pools	mand Palette		Silver  Silver  Bronze  Best Effort Co5 Any	Weight © 5	C o 💼 Alter Packs	MTU 0 1500	5 e
to Help > Com	mend Palette		Channel 3	o s	0 + 30 Diago Contraction 0 + 30 Diago Contraction Diago Contractio	мты о 2240	9500 - 9738 0 1900 - 9738
			< Cancel			[	Back Create

L

### Step 3 Click Create

**Step 4** Associate the System QoS policy to the Domain Profile and deploy.

	UCS Domain Configuration	
General	Select the compute and management policies to be associated with the fabric interconnect.	
UCS Domain Assignment	Show Attached Policies (1)	
VLAN & VSAN Configuration	^ Management 0 of 4 Policies Configured	
Ports Configuration		
5 UCS Domain Configuration	NTP	Select Policy 🗐
6 Summary	Syslog	Select Policy 🗐
0	Network Connectivity	Select Policy 🗐
	SNMP	Select Policy
	Network 1 of 2 Policies Configured	
	System QoS *	×   👁   🖉   Domain-QoS-5GFI 🗐
	Switch Control	Select Policy 🗐

**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

Navigate to <b>CONFIGURE &gt; Policies</b> . Click <b>Create Policy</b> , select <b>UCS Server</b> platform type, search or choose <b>LAN Connectivity</b> policy, and click <b>Start</b> .
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.
In the <b>Policy Details</b> page, click <b>Add vNIC</b> to create a new vNIC.
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 <b>Consistent Device Naming (CDN)</b> section of the Standalone server or the <b>Failover</b> section of EL-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

**Note** This property is available only on Standalone servers.

• Slide to Enable Trust Host CoS toggle button.

Note This property is available only on Intersight Managed Mode servers.

$\equiv \frac{-d_{\rm color}}{c_{ m soco}}$ Intersight	\$€ Infrastructure Service ∨		Q s	learch	ା ତା କଣ 🔹	013 (15)	ତ   ନ
線 Overview ⑥ Operate	Policies > Ethernet QoS Create						
Servers Chassis Fabric Interconnects HyperFlex Clusters	Ceneral     Policy Details	Policy Details Add policy details GoS Settings		V Al Padom	uCS Server (Standalo	ne)   UCS Server B	1-Attached)
Profiles Templates Policies		MTU, Bytes 9000 Class of Service	₿ • 1500 - 9000	Rate Limit, Mbps 0 Burst			₿ ● 100000
Pools           New         Command Palette         ************************************		5 Priority Platnum	0-6 v 0	10240			1000000
Navigate Interviget with Citri-K or go to Help > Command Palette		C Enable Trust Host CoS o					
	¢	Cancel				Back	Create

- Click **Select Policy** link below the **Ethernet Adapter**. Follow on to click 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 under Ethernet Adapter Default Configuration click Select Default Configuration to search and select Win-HPN-SMBd, the pre-defined Ethernet Adapter Default Configuration. Click Next and then Create.

= dealer Intersight 🗦	Infrastructure Service 🗸		Q Search	0 41 Q 🚥 💶 0	۹
12 Oversiew	Policies > Ethernet Adapter			Select Default Configuration	×
	Create			Ethernet Adapter Default Configuration 16	
(U. Operate				Q, Search	
Servers		General		( MQ-SMBJ	۲
Chassis	① General	Add a name, description and tag for the policy.		<ul> <li>SM8Server</li> </ul>	۲
Fabric Interconnects	2 Policy Details			SMBClent	•
HyperFlex Clusters		default ~		E Solaria	۲
🦉 Configure				( vsNCOracleRAC	۲
Profiles		Eth0		() WN-AzureStack	۲
To and the				@ Win-HPN-SMDd	•
Templates		Set Tans		( Wo-HPN	۰
Policies				Einux-NVMe-RoCE	•
Pools				(I MQ	•
		<= 1024		(] usNC	•
New Command Palette		Ethernet Adapter Default Configuration $\circ$		(] SROV	•
Navigate Intersight with Ctrl+K or go		Select Default Configuration 1		UMWarePassThru	•
to Help P Command Palette				① VMWare	•
				Windows	۰
				① Linux	•
	<	Cancel			

- Configure RoCE Settings in the policy: Click Create New to create a new policy. In the General page, enter the name of the policy. Under Policy Details page on right pane, use the following property settings, then 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

= the Intersight	💥 Infrastructure Service 🗸				Q, Search	9 4 9 0 00 00 0 A
<ul> <li>B. Overview</li> <li>B. Operate</li> <li>A.</li> </ul>	Policies > LANConnectivity > Create Create Ethernet A	Adapter				
Servers Chassie Fabric Warrowwetts HyperFiles Chasters	🕑 General	RoCE Settings Course ROMA over Converges Ethernet © Course Firm 286	0.0	Morony Regions * 131072	3 + 2 1 + 0.000	4 4 1 10
Politics Tergation		Version 2 Interrupt Settings	~ 0	5 Branupt Mode	<ul> <li>0</li> <li>interrupt Time, us</li> </ul>	
Point		512 minnup: Costoocing Type Me	0 0 - 0	Mär	v 0 13	4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Terrigent Interview And David or get		Receive Rentry Done Court 4	0 + (-1000	Receiver Ring Star 192	2 a 14 mile	
		Transmit Tommit Queve Count 1	\$ 4 1-00	Transmit Alling Sca 236	2 e 11 - 11 11	
		Completion Dampleton Count S	¢ •	Completion Reg Score 1	1	
		c Caced	0.0			Beth Diese

• Click Add to add and save the new vNIC settings.

- **Note** All the fields with \* are mandatory for creating LAN Connectivity Policy. Ensure they are 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 and deploy.
```

**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.

The following properties are available for this network adapter. Click the property you want to change on the left, and then select its value on the right. Property:  Yelue:  Compatible Operation Encapsulated Task Offload Encapsulation overhead Internupt Moderation IPV4 Checksum Offload V2 (IPv4) Large Send Offload V2 (IPv4) Large Send Offload V2 (IPv6) Maximum Number of RSS Process Maximum Number of RSS Queues Network Direct Functionality Nvgre Encapsulated Task Offload QoS Receive Side Scaling	General	Advanced	Driver	Details	Events	Resources		
Compatible Operation Encapsulated Task Offload Encapsulation overhead Interrupt Moderation IPV4 Checksum Offload Jumbo Packet Large Send Offload V2 (IPv4) Large Send Offload V2 (IPv6) Maximum Number of RSS Process Maximum Number of RSS Queues Network Direct Functionality Nvgre Encapsulated Task Offload QoS Receive Side Scaling	The foll the prop on the r	owing proper perty you war ight. y: 🕞	ties are a nt to char	nyailable fo nge on the	or this net e left, and Va	work adapter. then select it: ilue:	. Click s value	
	Encap Encap Interruj IPV4 C Jumbo Large Maximi Maximi Netwo Nvgre QoS Receiv	sulated Task sulation overf pt Moderation Checksum Off Packet Send Offload Send Offload um Number of rk Direct Fund Encapsulated ve Side Scalir	Offload head V2 (IPv4 V2 (IPv6 of RSS P f RSS Qu tionality d Task C	4) 5) rocess ueues	JE	21aDied		•

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

 $The {\tt Get-NetOffloadGlobalSetting}\ command\ shows\ NetworkDirect\ is\ enabled.$ 

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

Interface Inde>	RSS Capable	RDMA Capable	Speed	IpAddresses	Friendly Name
14	Тсне	Falce	40 Ghos	(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

#### 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 hos	st; listenei	s will be show	n in the command output.		
PS C:\Use PS C:\Use	rs∖Admir rs∖Admir	nistrator> nistrator> r	netstat -xan		
Active Ne	tworkDir	rect Connect	ions, Listeners, Shared	Endpoints	
Mode	IfIndex	Туре	Local Address	Foreign Address	PID
Kernel	9	Listener	50.37.61.23:445	NA	0
Kernel PS C:\Use	26 rs\Admir	Listener histrator>	10.37.60.158:445	NA	0

- 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.

⊨ 🐟   🙇 📷   🖾   👄	1	
<ul> <li>Computer Management (Local</li> <li>System Tools</li> <li>Task Scheduler</li> <li>Event Viewer</li> <li>Shared Folders</li> <li>Local Users and Groups</li> <li>Performance</li> <li>Monitoring Tools</li> <li>Performance Mc</li> <li>Data Collector Sets</li> <li>Device Manager</li> <li>Storage</li> <li>Windows Server Backup</li> <li>Disk Management</li> </ul>	Image: Solution of the system         Image: Solution of the system </th <th>Hyper-V Virtual Ethernet Adapter #2 2.000 2.000 0.000 0.000 0.000 598,340,974.354 553,916.589 0.000 6,588,510.951 35,589.270</th>	Hyper-V Virtual Ethernet Adapter #2 2.000 2.000 0.000 0.000 0.000 598,340,974.354 553,916.589 0.000 6,588,510.951 35,589.270

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 mode1 connections are correctly established between client and server.

ctive No	etworkDi	rect Connectio	ons, Listeners, Shared	Endpoints	
Mode	IfIndex	Туре	Local Address	Foreign Address	PIC
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



**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 <b>CONFIGURE &gt; Policies</b> . Click <b>Create Policy</b> , select <b>UCS Server</b> 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

and the state and the state of	🔆 Infrastructure Service 🗸		Q	Search	) © ¢10	0 60 60	۲	۹
two Intersight     transight     very Intersight     oreniam     oreniam     chasts     rates Interconnects     hyperfiles Clusters     configure     Paties     Tampietes     Tempietes     Pools	<ul> <li>Material control of a control of a</li></ul>	Policy Details Asspoky attass CoS Settings MTU, Byres too Class of Service 0 Policy Policy	0 0 1981-400 9 19	Search Rates Linit, Otgas 0 Bonst 10240	O 41 1	C III (III)	Image: State of the state o	A
General Command Patets Resplay time girls and Carlos (ar to https://Command Patets		Cruide That Host CoS - a				Back	Create	

- 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

esco Intersight	At Infrastructure Service ~				Q Search	Ø ¢10	0 00 00 0
Overview Operate	Policies > LAN Connectivity > Cre Create Ethernet	t Adapter					
Servers		RoCE Settings					
Channin	General	Enable RDMA over Converged I	Ethernet 0				
Courses	O denie	Queoe Pairs		Memory Regions *		Resource Groups	
Fabric Interconnects	Policy Details	256	Q 0	65536	č •	2	Ç e
HyperFlex Clusters	and the second sec	1					
		Version		Class Of Service			
Configure	^	Version 2	× 0	5	× 0		
Profiles							
		Interrupt Settings					
Templates		Interrupts		Interrupt Mode		Interrupt Timer, us	
Policies		132	2 0	MSIx	× 0	125	<u> </u>
P0015		Interrupt Coalescing Type					
		Min	~ 0				
Command Palette		Receive					
wigate Intersight with Ctrl+K or g Help > Command Palette	90	Receive Queue Count		Receive Ring Size			
		512	đ 0	512	Ç e		
		Transmit					
		Transmit Queue Count		Tranumit Ring Size			
		64	3 0	256	5 e		
		Completion					
		Company Company Company					
		576	0.5	Completion long size	5 o		
			1-2000		1.20		
		Unlink Failback Timeout (seconds)					
		5	5.0				
							(
		Cancel					Back Creat

- In the **Connection** section, use the following property setting for VMQ Connection and to create VMMQ Adapter policy:
  - For connection, select VMQ.
  - Enable Virtual Machine Multi-Queue using slider button.
  - 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.
      - Attention Do not modify the pre-defined parameters under Policy Details page, retain the default settings.
    - Click Next and then Create.

= duals Intersight 2	🔹 Infrastructure Service 🗸		0 41 Q 🚥 🚥 0	
) Dverview 16. Operate A	Posicies > LAN Connectivity > Create Create Ethernet Ad	apter	Select Default Configuration Policies 18 Q <sub>1</sub> Search	
Servers Chasis Fabric Interconnects Higherflux Clasters Portice Temptos Potos P	Decent     Policy Details	General       Add a name, desception and tag for the policy.       Organization *       andow       Market       Bitl       Set Tags       Description       Ethernet Adapter Default Configuration •       Linkst Default Configuration •	VO-2008     VO-2008     VO-2008     VO-2008     SoftConcelled     SoftConcelled     VO-2008     V	
	<	Cancel		

- 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 on the host operating system.

### **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.

Follow the below procedure for the host operating system configuration for RoCEv2 Mode 2.

### 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
  - 2 Create a new Virtual Network Switch (vswitch) for the RoCE v2-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.

Virtual Switches	X. Virtual Switch Properties
Cisco VIC Ethernet Interface #2	vswitch
Global Network Settings	Notes:
00-15-5D-3A-A0-00 to 00-15-5D-3	
	Connection type
	What do you want to connect this virtual switch to?
	External network:
	Cisco VIC Ethernet Interface #2
	Allow management operating system to share this network adapter
	Enable single-root I/O virtualization (SR-IOV)
	O Internal network
	O Private network
	VIAN ID
	Enable virtual LAN identification for management operating system
	The VLAN identifier specifies the virtual LAN that the management operating system will use for all network communications through this network adapter. This setting does not affect virtual machine networking.
	2
	Remove
	SR-IOV can only be configured when the virtual switch is created. An external virtual switch with SR-IOV enabled cannot be converted to an internal or private 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)"



Active N	etworkDi	rect Connect	tions, Listeners, Shared	Endpoints	
Mode	IfIndex	Туре	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:\Us	ers\Admin	nistrator>			

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:\Use PS C:\Use	rs\Admir rs\Admir	histrator> histrator> ne	tstat -xan		
Active Ne	tworkDir	rect Connecti	ons, Listeners, Share	dEndpoints	
Mode	IfIndex	Туре	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:\Use	rs\Admir	nistrator>			

#### 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.

≡	cisco Intersight	ho Infrastructure Service $$			Q POLIC	⊗ ⊘ ಧ 🚥	<b>41</b>	Q
:@: :©_	Overview Analyze	Smart Licensing registration failed. Register	er your license with a valid Product	instance Registration Token. Go to	Licensing		>	×
	Operate ^	Policies					Create Policy	•
	Servers Chassis Fabric Interconnects	* All Policies  +	× Add Filter	×	Export 19 items found	6 ∨ per page 🔣 🔇 1	of 4 >>>	
,c	HyperFlex Clusters Configure  Profiles	UCS Server 19	• Used 10 • Not Used 9					
	Policies	Name : Pl	atform Type	'ype	Usage	Last Update	: 4	
	Pools		CS Server L	AN Connectivity	3 🐻	May 29, 2021 4:36 AM		
		lcp_policy1_2021 UC	CS Server L	AN Connectivity	1 🕞	May 13, 2021 4:15 AM		
			CS Server L	AN Connectivity	1 🔯	May 12, 2021 5:31 AM		
Nev	Command Palette		CS Server L	AN Connectivity	1 🕼	Feb 12, 2021 12:12 PM		
Navi	ate Intersight with Ctrl+K or go	LCPra UC	CS Server L	AN Connectivity	0 🗇	Feb 12, 2021 12:12 PM		
to He	Ip > Command Palette		CS Server L	AN Connectivity	1 🗟	Feb 12, 2021 12:11 PM		
		Carl Selected 1 of 19 Show Selected	cted Unselect All			⊠ < 1	of 4 > >	•

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



# Configuring NVMeoF with RoCEv2 in Linux

- Guidelines for using NVMe over Fabrics (NVMeoF) with RoCE v2 on Linux, on page 21
- Linux Requirements, on page 22
- Configuring RoCE v2 for NVMeoF on Cisco Intersight, on page 22
- Configuring RoCE v2 for NVMeoF on the Host System, on page 26
- Setting Up Device Mapper Multipath, on page 29
- Deleting the RoCE v2 Interface Using Cisco Intersight, on page 30

# 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 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.

≡	uludu Intersight	»؛ ا	nfrastructure Service 🗸		Q s	iearch	⊘ ⊈1	Q 🚥	A 15 🧿	۶
:@:	Overview		Dark theme is now available in Intersight	It. To switch the theme go to the Us	er Settings					×
0	Operate Servers Chassis	^	Policies > System QoS Create							
	Fabric Interconnects			Add policy details						
	HyperFlex Clusters		General	• This policy is applicable	only for UCS Domains					
۰,	Configure	^	2 Policy Details	Configure Priorities						1
	Profiles Templates			Platinum CoS 5	() 0 - 6	eight 0 o Allo 0 - 10	ow Packet ops	MTU 9216	0 o	
	Policies			Gold						
	Pools			Silver						
N	ew Command Palette	×		Bronze						
Na to I	vigate Intersight with Ctrl+K or Help > Command Palette	90		Best Effort CoS Any	Wei ○ 5	eight	ow Packet o	MTU 1500	۰ و	
				Fibre CoS Channel 3	Wei ○ 5	ight	ow Packet o	1 MTU 2240	500 - 9216 ©	
			<	Cancel					Back Cre	eate

### Step 3 Click Create.

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

	UCS Domain Configuration	
General	Select the compute and management policies to be associated with the fabric intercon	nect.
UCS Domain Assignment	Show Attached Policies (1)	
VLAN & VSAN Configuration	Management 0 of 4 Policies Configured	
Ports Configuration	······································	
5 UCS Domain Configuration	NTP	Select Policy 🗐
<ol> <li>Cummuni</li> </ol>	Syslog	Select Policy 🗐
6 Summary	Network Connectivity	Select Policy 🗐
	SNMP	Select Policy 🗐
	Network 1 of 2 Policies Configured	
	System QoS *	×   👁   🖉   Domain-QoS-5GFI 🗐
	Switch Control	Select Policy 🗐

**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**.

= 🖞 esce Intersight 🏃 Infrastructure Service 🗸		0 41 Q 🚥 🌆 0 R
Policies > Ethernet Adapter		Select Default Configuration
Create		Ethernet Adapter Default Configuration 18
Operate A		Q. Search
Servers		E No-subs an
Chassis O General	General Add a name, description and tag for the policy.	SMBServer (D)
Fabric Interconnects (2) Policy Details		g Milleri al-
HyperFlex Clusters	detautt -	🗐 Sooria 👘
🖉 Configure 🔿		🗇 wANDOWCHAD 🗠 🗠
Profiles	Eth2	T wei-Azurdiack (D)
Termistes		🗇 Wo-HEN-SM63 👘
	Set Tags	S Wo-HPN (D)
Postors		C Linux-Miller RoCE (D)
Pools	Description	@ MO @
	er tak	E whit of
New Command Palette	Ethernet Adapter Default Configuration	e seov ao
Navigala Intercipte with CB1+K or go to Histo > Command Palatte	Select Default Configuration II	🗇 VMWarePassThru 🗠
		1 VMMee (D)
		🗑 Windows 🔅
		🗇 Livas 🗠
5	Cancel	

- 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.

≡ diada Intersight 🗼	Infrastructure Service 🗸		Q Sea	irch G	) \$1	Q (13) (A15)	0
<ul> <li>Overview</li> <li>Operate</li> </ul>	Policies > BIOS						
Servers Chassis	General	+ LOM And PCIe Slots					
Fabric Interconnects HyperFlex Clusters	2 Policy Details	+ Main					
🖉 Configure 🔨		- Memory					
Profiles Templates		Enhanced Memory Test platform-default	BM v © plat	tE DMA Mitigation tform-default		÷	0
Policies		CPU SMEE platform-default	√ ○ ena	имu abled		v	0
New Command Palatte		Bank Group Swap platform-default	⊂hi √ ⊙ plat	ipset Interleave tform-default		v	0
Navigate Intersight with Ctrl+K or go to Help > Command Palette		SNP Memory Coverage platform-default	SNa v © plat	P Memory Size to Cover in MiB * tform-default			0
		NUMA Nodes per Socket platform-default	AM v 0_ plat	tD Memory Interleaving tform-default		v	0
		AMD Memory Interleaving Size platform-default	SE\ ~ 0 plat	V-SNP Support tform-default		v	0
	c	Cancel				Back	Crea

- 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_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=/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.

### 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@localhos1	t~]#	dmesg	grep	enic_	rdma					
[	3.137083]	enic	_rdma:	Cisco \	/IC Eth	hernet NIC	RDMA Dr	iver,	ver	1.2.0.	28-877.2
2	init										
[	3.242663]	enic	0000:1	b:00.1	eno6:	<pre>enic_rdma:</pre>	FW v3	RoCEv2	2 ena	bled	
[	3.284856]	enic	0000:1	b:00.4	eno9:	enic_rdma:	FW v3	RoCEv2	2 ena	bled	
[	16.441662]	enic	0000:1	b:00.1	eno6:	<pre>enic_rdma:</pre>	Link l	JP on (	enic_	rdma_C	)
[	16.458754]	enic	0000:1	b:00.4	eno9:	enic_rdma:	Link L	JP 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** -1 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 29.

### **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 nyme 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.

#### **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	09A703295	EE2954E	Pure Storage	FlashArray	72656
4.29 GB /	4.29 GB	512 в+	0 в 99.9.9		
/dev/nvme0n2 5.37 GB /	09A703295 5.37 GB	EE2954E 512 B +	Pure Storage 0 B 99.9.9	FlashArray	72657

## **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

def

- 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 :

aults	5 {	
	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:
	<pre># multipath -v2</pre>
Step 7	Check the multipath status:

# multipath -11

# **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.

	their Intersight	💡 Infrastructure Service 🗸		Q POLIC	⊗ ⊘ ಧ 🚥	<u>(41)</u> ③ へ
:©: 	Overview Analyze	Smart Licensing registration failed. Register your liv	ense with a valid Product instance Registration Token.	Go to Licensing		×
0	Operate ^	Policies				Create Policy
	Servers Chassis Fabric Interconnects HyperFlex Clusters	× All Policies      +      ✓      ✓      ✓      ✓      ✓      ✓	Filter	× C Export 19 Items found	6 v per page 🔣 <	1 of 4 > >
,e	Configure ^	(19) • Used • Not Us	0 di 9			
	Policies	Name Platform T	уре Туре	C Usage	Last Update	- <i>P</i>
	Pools	UCS Serve	r LAN Connectivity	3 🐻	May 29, 2021 4:36 AM	
		Lcp_policy1_2021 UCS Serve	r LAN Connectivity	1 🗟	May 13, 2021 4:15 AM	
		UCS Serve	r LAN Connectivity	1 🔯	May 12, 2021 5:31 AM	
Nev	W Command Palette ×	LCP UCS Serve	LAN Connectivity	1 🐻	Feb 12, 2021 12:12 PM	
Navi	gate Intersight with Ctrl+K or go	LCP. it in UCS Serve	LAN Connectivity	0 🐻	Feb 12, 2021 12:12 PM	
to He	elp > Command Palette	UCS Serve	LAN Connectivity	1 🐻	Feb 12, 2021 12:11 PM	
		🖉 🖉 📋 Selected 1 of 19 Show Selected	Unselect All			1 of 4 > >

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



# **Configuring NVMeoF with RoCEv2 in ESXi**

- Guidelines for using NVMe over Fabrics (NVMeoF) with RoCE v2 on ESXi, on page 31
- ESXi Requirements, on page 32
- Configuring RoCE v2 for NVMeoF on Cisco Intersight, on page 32
- NENIC Driver Installation, on page 36
- ESXi NVMe RDMA Host Side Configuration, on page 36
- Deleting the RoCE v2 Interface Using Cisco Intersight, on page 43

# 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-2.0.4.0-10EM.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.

≡	ີ Intersight	<b>.</b> ** 1	Infrastructure Service 🗸		Q	Search	⊘ ⊄11	Q 013 🔺 15	0	Q
:@:	Overview		Dark theme is now available in Intersight	It. To switch the theme go to the User Setting	S					×
0	Operate Servers Chassis	^	Policies > System QoS Create							
	Fabric Interconnects		General	Add policy details  This policy is applicable only for UC	S Domains	s				1
.0	Configure	^	2 Policy Details	Configure Priorities						
	Profiles Templates			Platinum CoS 5	€ 0 1 0 - 6	Weight 10 🗍 © 🗌 Allo 0 - 10	w Packet <sub>©</sub> ps	MTU 9216 ©	© 9216	
	Policies Pools			Gold						
				Silver						
Nan to H	Command Palette	x go		Best Effort	v	Weight	uu De sket	MTU		
				Fibre CoS	© 5 V	5 C O O O O O O O O O O O O O O O O O O	ps ©	1500 () 1500 - 1 MTU	9216	
				Channel 3	0-6	0 - 10	ps ©	2240	9216	
			<	Cancel				Back	Creat	te

### Step 3 Click Create.

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

	UCS Domain C	onfiguration	
General	Select the compute	and management policies to be associated with the fabric interconnect.	
UCS Domain Assignment		Show Attached Policies (1)	
VLAN & VSAN Configuration	M	Anagement 0 of 4 Policies Configured	
Ports Configuration			
5 UCS Domain Configuration		NTP	Select Policy 🗐
6 Summary		Syslog	Select Policy 🗐
U U U U U U U U U U U U U U U U U U U		Network Connectivity	Select Policy 🗐
		SNMP	Select Policy
	^ N	letwork 1 of 2 Policies Configured	
		System QoS *	×   👁   🖉   Domain-QoS-5GFI 🗐
		Switch Control	Select Policy 🗐

**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 <b>CONFIGURE &gt; Policies</b> . Click <b>Create Policy</b> , select <b>UCS Server</b> platform type, search or choose <b>LAN Connectivity policy</b> , and click <b>Start</b> .
Step 2	In the policy <b>General</b> page, enter the policy name, select the Target Platform as <b>UCS Server (Standalone)</b> or <b>UCS Server (FI-Attached)</b> , and click <b>Next</b> .
Step 3	In the <b>Policy Details</b> page, click <b>Add vNIC</b> to create a new vNIC.
Step 4	In the Add vNIC page, follow the configuration parameters to enable the RoCE v2 vNIC:
	<ul> <li>a) In the General section, provide a name for virtual ethernet interface.</li> <li>b) Incase of a Standalone server, click the Consistent Device Naming (CDN) or click the Failover of a FI-attached server , and do the following:</li> </ul>
	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 <b>Policy Details</b> page of the Ethernet Adapter Policy, modify the following property setting:
	RoCE Settings
	• For <b>Enable RDMA over Converged Ethernet</b> , slide to enable and set the RoCE on this virtual interface.
	• For Queue Pairs, select or enter 1024
F	• For Memory Regions, select or enter 131072
	• For <b>Resource Groups</b> , select or enter <b>8</b>
b) Inc FI-	• 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.

≡ <sup>-diada</sup> Intersight	å≋ In	ifrastructure Service 🗸			Q Search		⊘ ⊄11	<b>Д 013 615</b>	0		
<ul> <li>Overview</li> <li>Operate</li> <li>Servers</li> </ul>	^	Dark theme is now available in Intersight Policies > LAN Connectivity > Create Create Ethernet Adap	To switch the theme go to the User Setti	ngs							
Chassis Fabric Interconnects HyperFlex Clusters Configure Profiles Templates	^	General     Policy Details	Enable Virtual Extensible L/     Enable Network Virtualizati     Enable Accelerated Receive     Enable Precision Time Proto	LN ⊙ musing Gen Flow Steerin col ⊙	eric Routing Encapsulation g ♀	All Platforms UC	S Server (Standalor	e) UCS Server (FI-Attache	ed)		
Policies Pools New Command Palette	×		Endle Advancel Filter     Endle Advancel Filter     Stable Interrupt Schlag     O     Endle GENEVE Official     O     RocCE Settings     Concerned Etherer								
Navigato Intersight with CBrHK or g	o to	K	Queue Pairs 1024 Version Version 2		Memory Regions * 131072 Class Of Service 5	<ul> <li> <sup>(1)</sup> • 524288      </li> </ul>	Resource Grou 8	ps	© 28 reate		

- 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
  TS0 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.

Summary	Monitor	Configur	e Pe	rmissions	VMs	Resour	ce Pools	Datastores	Net	works	Update	s					
Storage		~ 1	VMk	ernel a	dapte	rs											
Storage Storage	Adapters Devices	- 1	ADD	NETWORKI	NG	REFRESH											
Host Cad	che Configuratio	on		Device		т	Network La	bel	т	Switch	т	IP Ad	dress	т	TCP/IP Stack	т	Enabled Services
Protocol	I Endpoints		:	>> vmk0			Managem	ent Network		vSwitch	0	10.19	3.176.52		Default		Management
I/O Filter	rs	_	:	>> vmk1			vmk284			vSwitch	1	50:2	84::210		Default		
Networkin	ng	~	:	>> vmk2			vmk283			vSwitch:	2	50.2	83.210		Default		
Virtual s	witches		<u> </u>														
VMkerne	el adapters																
Physical	adapters																
RDMA ad	dapters																
TCP/IP c	configuration	- 1															
Virtual Ma	chines	~															

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

Summary Monitor	Configur	e Permissions	VMs	Resource	Pools (	Datastores	Net	works	Updates						
Storage	~	RDMA adap	ters												
Storage Adapters Storage Devices	- 1	Name	т	Driver	٠	State	٣	Paired Uplin	nk T	RoCE v1	۲	RoCE v2	۲	IWARP	۲
Host Cache Configuration	n	vmrdma0		nenic		Active		vmnic2		Disabled		Enabled		Disabled	
Protocol Endpoints	- 1	vmrdma1		nenic		Active		vmnic3		Disabled		Enabled		Disabled	
VO Filters	- 1														
Networking	~														
Virtual switches	- 1														
VMkernel adapters	- 1														
Physical adapters															
RDMA adapters															
TCP/IP configuration	- 1	RDMA Device: vmr	fma1												
Virtual Machines	~														
VM Startup/Shutdown		Properties Bo	und VM	kernel Adapte	ers										
Agent VM Settings	- 1														
Default VM Compatibility		VMkernel Adapter	TC	VIP Stack	IP Address										
Swap File Location		vmk2	De	ault	50 2 83 210										
System	~														
Licensing															

### 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.

Summary Monitor	Configur	re Permissions VMs Re	source Pool	s Datastor	es	Networks	Up	dates							
Storage	~	Storage Adapters													
Storage Adapters		+ Add Software Adapter 😸 Refre	sh 🖾 Resce	an Storage	G Res	can Adapter X	Rer	nove							
Storage Devices	- 1	Adapter	¥	Type	Ŧ	Status	¥	Identifier	¥	Targets	Ψ	Devices	Ŧ	Paths	Ŧ
Host Cache Configuration		<ul> <li>Model: Cisco 12G Modular Raid</li> </ul>	controller with	2GB cache											1
Protocol Endpoints	- 1	♦ vmhba5		SAS		Unknown				2		2		2	_
VO Filters	- 1	Model: Cisco UCS VIC Fnic Cont	oller												
Networking	~	G- vmhba0		<b>Fibre Channel</b>		Offline		10.00.2c#8.9b79.8d.be 20.00.2c#8.9b79.8d.be	e-	0		0		0	_
Virtual switches	- 1	vmhba2		<b>Fibre Channel</b>		Offline		10.00.2c18.9b79.8d.bl 20.00.2c18.9b79.8d.bl		0		0		0	_
VMkernel artanters	- 1	winba3		<b>Fibre Channel</b>		Offine		10.00.2c#8.9b:51.b3:3c 20.00.2c#8.9b:51.b3:3c		0		0		0	
Dhusical adapters	- 1	vmhba4		Rbre Channel		Offline		10.00.2c18.9b.51b3.3d 20.00.2c18.9b.51b3.3d		0		0		0	
Physical adapters	- 1	<ul> <li>Model: Lewisburg SATA AHCI Co</li> </ul>	ntroller												
TCD/D configuration	- 1			Block SCSI		Unknown		-		0		0		0	
Torrir contryaction												00	py All	8	items
Virtual Machines	×														
VM Startup/Shutdown	- 1														
Agent VM Settings	- 1														
Default VM Compatibility															
Swap File Location															

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

Storage	Storage Adapt	ers						
Storage Adapters				_				
Storage Devices	Add Software Adapte	r 10.193.176.52		×		٣	Targets	
Protocol Endooin							2	
I/O Filters	A new software (CCC) educies will	he added to the lat. After it has been	en added salect the adapter and use the Adapter				-	
Networking	Details section to complete the co	infiguration.	in added, select the adapter and use the Adapter		2c18.9b:79.8d.be		0	
Mintual caritebas					2c18:90:79:8d:bf		0	
VMkernel adapter	Add software NVMe over RDMA ad	lapter			2c:f8:96:51:b3:3c		0	
Physical adapters	Enable software NVMe adapter or	n the selected RDMA device.			2c:18:9b:51:b3:3d		0	
RDMA adapters								
TCP/IP configurat	RDMA device:	vmrdma0/	~				0	
Virtual Machines	O Add Software FCoE Adapter							
VM Startup/Shute	Discover software FCoE adapters	associated with the following physic	al network adapter.					
Agent VM Setting								
Default VM Comp	Physical Network Adapter:	vmnic0	~					
Swap File Locatio	VLAN ID:	0	Range: 0 - 4094					
System								
Licensing	Priority Class:	3	Range: 0 - 7					
Host Profile	Controller MAC Address:	2c;f8:9b:a1:4c:e6						
Time Configuratio								
Authentication Se				_				
Certificate			CANCEL OK					
Power Manageme								
Advanced System	Settings							

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

Summary Monitor	Configu	re Permissions VMs Resour	ce Pools Datastores	Networks 0	Updates					
Storage	<u>~</u> ]	Storage Adapters								
Storage Adapters		+ Add Software Adapter 😨 Refresh	🖞 Rescen Storage   🕲 A	Rescan Adapter 🛛 🛛	temove					
Storage Devices	- 1	Adapter	y Type y	Status y	klentifier	<ul> <li>Targets</li> </ul>	Y Devices	Y Paths Y		
Host Cache Configuration	Host Cache Configuration   Model, Cauco 12G Modular Red Controller with 2GB cache									
Protocol Endpoints		(+ vmhbe5	SAS	Unknown	-	2	2	2		
VO Filters		Model: Cisco UCS VIC Fric Centroller								
Networking	~	G vmhba0	Fibre Channel	Offine	10 00 2c f8 9b 79 8d be 20 00 2c f8 9b 79 8d be	0	0	0		
Michael excitence		G- vmhbe2	Fibre Channel	offine	10:00:2c:f8:9b:79:8d:bf 20:00:2c:f8:9b:79:8d:bf	0	0	0		
Virtual switches		Gr vmhbi3	Fibre Channel	Offine	10:00 2c #8.9b 51 b3 3c 20:00 2c #8.9b 51 b3 3c	0	0	0		
VMkernel adapters		G vmhbe4	Fibre Channel	Offine	10:00:2cf8:9b;51:b3:3d 20:00:2cf8:9b;51:b3:3d	0	0	0		
Physical adapters		<ul> <li>Model: Lewisburg SATA AHCI Controll</li> </ul>	er							
RDMA adapters		G vention	Block SCSI	Unknown	-	0	0	0		
TCP/IP configuration		<ul> <li>Model: VMware NVME over RDMA Sto</li> </ul>	rage Adapter							
Virtual Machines	~	G vmbbo64	ROMA	Unknown	-	0	0	0		
VM Startup/Shutdown			RDMA	Unknown	-	1	1	1		
Agent VM Settings										
Default VM Compatibility										
Swap File Location								Copy At   8 tems		
System	~	Properties Devices Paths N	mespaces Controllers							
Licensing		ADD CONTROLLER REMOVE								
Host Profile		Name y Sub	system NON	Transport Type	FUSE Support	Model	Firmware Ver	sion		

### **Displaying vmnic and vmrdma Interfaces**

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

L

### 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 I	Link	Speed	Duplex	MAC Address	MTU	Description
vmnic0	0000:3b:00.0	ixgben (	Down	OMbps	Half	2c:f8:9b:a1:4c:e6	1500	Intel(R) Ethernet Controller X550
vmnic1	0000:3b:00.1	ixgben I	Up	1000Mbps	Full	2c:f8:9b:a1:4c:e7	1500	Intel(R) Ethernet Controller X550
vmnic2	0000:1d:00.0	nenic I	Up	50000Mbps	Full	2c:f8:9b:79:8d:bc	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic3	0000:1d:00.1	nenic I	Up	50000Mbps	Full	2c:f8:9b:79:8d:bd	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic4	0000:63:00.0	nenic I	Down	OMbps	Half	2c:f8:9b:51:b3:3a	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic5	0000:63:00.1	nenic (	Down	OMbos	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				Half		1500	Intel(R) Ethernet Controller X550
vmnic1	0000:3b:00.1	ixgben	Up		1000	Full	2c:f8:9b:a1:4c:e7	1500	Intel(R) Ethernet Controller X550
vmnic2	0000:1d:00.0	nenic	Up		50000	Full	2c:f8:9b:79:8d:bc	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic3	0000:1d:00.1	nenic			50000	Full	2c:f8:9b:79:8d:bd	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic4	0000:63:00.0	nenic	Up			Half	2c:f8:9b:51:b3:3a	1500	Cisco Systems Inc Cisco VIC Ethernet NIC
vmnic5	0000:63:00.1	nenic	Up			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@S	R	ackServe	r:~] e	sxcli rdm	a device list	
lame	Driver	State	MTU	Speed	Paired Uplink	Description
/mrdma0	nenic	Active	4096	50 Gbps	vmnic1	Cisco UCS VIC 15XXX (A0)
/mrdma1	nenic	Active	4096	50 Gbps	vmnic2	Cisco UCS VIC 15XXX (A0)
root@S	R	ackServe	r:~] e	sxcli rdm	a device vmknic	list
)evice	Vmknic	NetStac	k			
/mrdma0	vmk1	default	TcpipS	tack		
/mrdma1	vmk2	default	TcpipS	tack		

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



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

[root@ESXi7U3	er list		
Adapter Adapter Qualified Name	Transport Type	Driver	Associated Devices
vmhba64 aqn:nvmerdma:2c-f8-9b-79-8d-bc vmhba65 aqn:nvmerdma:2c-f8-9b-79-8d-bd [root@ESXi7U3	RDMA RDMA	nvmerdma nvmerdma	vmrdma0, vmnic2 vmrdma1, vmnic3

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

[root@ESX	i703	:~] esxcli s	torage 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@ESX	i7U3	:~]			

**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 -1 transport address

figure with esxcli nvme fabrics discover -a vmhba64 -1 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 <b>]</b>	ntroller list			Controller Number	Adapter	Transport Type	Is Online
ngn.2010-06.com.purestorage:flasharra	v.5ab274df5b1614554	wmhba64#50.2.8	34.100:4420	258	vmhba64	RDMA	true
ngn.2010-06.com.purestorage:flasharra	y.Sab274df5b161455#	wmhba65#50.2.8	33.100:4420	259	vmhba65	RDMA	
Name	Controller Number	Namespace ID	Block Size	Capacity in MB			
eui.00e6d65b65a8f34024a9374e00011745	258	71493	512	102400			
eui.00e6d65b65a8f34024a9374e00011745	259	71493	512	102400			

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 -1 50.2.84.100
Transport Type Address Family Subsystem Type Controller ID Admin Queue Max Size Transport
Address Transport Service ID Subsystem NQN
IPV4
                            NVM
RDMA
                                          65535
                                                        31
50.2.84.100
             4420
nq.210-06.com.purestorage:flasharray:2dp1239anjk1484
[root@ESXiUCSA:~] esxcli nvme fabrics discover -a vmhba64 -1 50.2.84.100 p 4420 -s
nq.210-06.com.purestorage:flasharray:2dp1239anjk1484
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.

$\equiv$ "dially intersight	ູ່ ຈະ ເຈົ້າການ ເປັນ ເພື່ອ ເປັນ ເປັນ ເປັນ ເປັນ ເປັນ ເປັນ ເປັນ ເປັນ	
:@: Overview ⓒ Analyze	Smart Licensing registration failed. Register your license with a valid Product instance Registration To	oken. Go to Licensing
() Operate	Policies	Create Policy
Servers Chassis Fabric Interconnects HyperFlex Clusters Configure Profiles		× G Export 19 Items found 6 ∨ per page ≤ 1 of 4 2 x
Policies	Name Platform Type Type	: Usage Last Update : §
Bask	UCS Server LAN Connectivity	3 🔯 May 29, 2021 4:36 AM
Pools	LAN Connectivity	1 🕞 May 13, 2021 4:15 AM ····
	UCS Server LAN Connectivity	1 🗋 May 12, 2021 5:31 AM
New Command Palette	× LCP LCP LAN Connectivity	1 🕞 Feb 12, 2021 12:12 PM
Navigate Intersight with Ctrl+K of to Help > Command Palette	go LCP LAN Connectivity	0 [b] Feb 12, 2021 12:12 PM
	Selected 1 of 19     Show Selected     Unselect All	K € 1 of 4 ∑ 2

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



# Known Issues

- Windows, on page 45
- Linux, on page 46
- ESXi, on page 46

# Windows

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.

I

Linux

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.	<ol> <li>To recover the switch from this error mode.</li> <li>Log into the switch.</li> <li>Locate the port that connected to the storage and shut down the port using "shutdown" command</li> <li>Execute the following commands one by one:         <ul> <li># clear counters</li> <li># clear counter buffers module 1</li> <li># clear qos statistics</li> </ul> </li> <li>Run no shutdown on the port that was shut down.</li> </ol>

# ESXi

Symptom	Conditions	Workaround
When using the command <b>esxcli storage</b> <b>core adapter list</b> 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> .	This is a known issue in ESXi 7.0 Update 3.	None
Note VMware Developer Center Partner Network (DCPN) Case ID - 00113157		