



Cisco Intersight 構成ガイド for RDMA over Converged Ethernet (RoCE) バージョン 2

最終更新：2024年11月19日

シスコシステムズ合同会社

〒107-6227 東京都港区赤坂9-7-1 ミッドタウン・タワー

<http://www.cisco.com/jp>

お問い合わせ先：シスコ コンタクトセンター

0120-092-255 (フリーコール、携帯・PHS含む)

電話受付時間：平日 10:00～12:00、13:00～17:00

<http://www.cisco.com/jp/go/contactcenter/>

通信、サービス、偏向のない言語、およびその他の情報

- シスコからタイムリーな関連情報を受け取るには、[Cisco Profile Manager](#) でサインアップしてください。
- 重要な技術によりビジネスに必要な影響を与えるには、[Cisco Services](#) にアクセスしてください。
- サービス リクエストを送信するには、[Cisco Support](#) にアクセスしてください。
- 安全で検証済みのエンタープライズクラスのアプリケーション、製品、ソリューション、およびサービスを探して参照するには、[Cisco Marketplace](#) にアクセスしてください。
- 一般的なネットワークング、トレーニング、認定関連の出版物を入手するには、[Cisco Press](#) にアクセスしてください。
- 特定の製品または製品ファミリの保証情報を探すには、[Cisco Warranty Finder](#) にアクセスしてください。

マニュアルに関するフィードバック

シスコのテクニカルドキュメントに関するフィードバックを提供するには、それぞれのオンラインドキュメントの右側のペインにあるフィードバックフォームを使用してください。

Cisco バグ検索ツール

[Cisco Bug Search Tool](#) (BST) は、シスコ製品とソフトウェアの障害と脆弱性の包括的なリストを管理する Cisco バグ追跡システムへのゲートウェイとして機能する、Web ベースのツールです。BST は、製品とソフトウェアに関する詳細な障害情報を提供します。

偏向のない言語

この製品のマニュアルセットは、偏向のない言語を使用するように配慮されています。このドキュメントセットでの偏向のない言語とは、年齢、障害、性別、人種的アイデンティティ、民族的アイデンティティ、性的指向、社会経済的地位、およびインターセクショナルリティに基づく差別を意味しない言語として定義されています。製品ソフトウェアのユーザーインターフェイスにハードコードされている言語、基準ドキュメントに基づいて使用されている言語、または参照されているサードパーティ製品で使用されている言語によりドキュメントに例外が存在する場合があります。



第 1 章

コンバージドイーサネット上の RDMA (RoCE) バージョン 2

- [コンバージドイーサネット上の RDMA \(1 ページ\)](#)

コンバージドイーサネット上の RDMA

コンバージドイーサネット上の RDMA バージョン 2 (RoCEv2) は、イーサネットネットワークを介したリモートダイレクトメモリアクセス (RDMA) を可能にするネットワークプロトコルです。RDMA テクノロジーの利点を活用することで、サーバまたはストレージシステム間の低遅延で高帯域幅の通信を可能にします。RoCEv2 では、従来の TCP/IP ネットワーキングスタックのオーバーヘッドが不要になるため、パフォーマンスが向上し、遅延が減少します。これにより、効率的なデータ転送が可能になり、アプリケーションがリモートメモリに直接アクセスできるようになるため、ネットワーク全体の効率と拡張性が向上します。RoCEv2 は、ネットワークパフォーマンスを最適化し、データ集約型のワークロードを高速化するために、データセンターやハイパフォーマンスコンピューティング環境でよく使用されます。

RoCE v2 は、Windows、Linux、および ESXi プラットフォームでサポートされています。



第 2 章

Windows での RoCEv2 を使用した SMB ダイレクトの設定

- [RDMA over コンバージドイーサネット \(RoCE\) v2 を使用して Windows で SMB ダイレクト サポートを使用するためのガイドライン \(3 ページ\)](#)
- [Windows での RoCE v2 モード 1 および 2 の構成の概要 \(5 ページ\)](#)
- [Windows の要件 \(6 ページ\)](#)
- [Cisco Intersight でのモード 1 の構成 \(6 ページ\)](#)
- [ホストシステムでの SMB ダイレクト モード 1 の設定 \(12 ページ\)](#)
- [Cisco Intersight でのモード 2 の設定 \(15 ページ\)](#)
- [ホストシステムでのモード 2 の設定 \(19 ページ\)](#)
- [Cisco Intersight を使用した RoCE v2 インターフェイスの削除 \(22 ページ\)](#)

RDMA over コンバージドイーサネット (RoCE) v2 を使用して Windows で SMB ダイレクト サポートを使用するためのガイドライン

一般的なガイドラインと制限事項

- Cisco Intersight 管理モードは、Microsoft Windows Server 2019 以降で、Microsoft SMB Direct with RoCE v2 をサポートします。Windows Server リリースに対し、Microsoft からのすべての KB 更新を使用することを推奨します。



- (注)
- RoCE v2 は Microsoft Windows サーバ 2016 ではサポートされていません。
 - サポートされている特定のオペレーティングシステム (OS) については、[Windows の要件](#)を参照してください。

- Cisco では、UCS Manager リリースに特有の **UCS ハードウェアおよびソフトウェア互換性**を確認して、Microsoft Windows で RoCE v2 を使用した Microsoft SMB ダイレクトのサポートを決定することをお勧めします。
- Microsoft SMB Direct with RoCE v2は、Cisco UCS VIC 1400 シリーズと VIC 14000、VIC 15000 シリーズアダプタでのみサポートされています。UCS VIC 1200 シリーズおよび VIC 1300 シリーズアダプタではサポートされていません。RoCE v2 を使用した SMB ダイレクトは、すべての UCS ファブリック インターコネクでサポートされています。



(注) RoCE v1 は、Cisco UCS VIC 1400 シリーズ、VIC 14000 シリーズ、および VIC 15000 シリーズ アダプタではサポートされていません。

- Cisco のアダプタ間では、RoCE v2 設定がサポートされています。シスコのアダプタとサードパーティ製のアダプタ間の相互運用性はサポートされていません。
- RoCE v2 は、アダプタごとに 2 個の RoCE v2 対応 vNIC と、アダプタ インターフェイスごとに 4 個の仮想ポートをサポートします。これは、セットスイッチ設定とは無関係です。
- RoCE v2 対応の vNIC インターフェイスでは、Cisco Intersight Managed Mode ドメインプロファイルで no-drop QoS システム クラスが有効になっている必要があります。
- RoCE プロパティのキューペア設定は、少なくとも 4 組のキューペア用である必要があります。アダプタあたりのキューペアの最大数は 2048 です。
- QoS No Drop クラス設定は、Cisco Nexus 9000 シリーズ スイッチなどのアップストリームスイッチで適切に設定する必要があります。QoS の設定は、異なるアップストリームスイッチ間で異なります。
- RNIC インターフェイスあたりのメモリ領域の最大数は 131072 です。
- SMB Direct with RoCE v2 は、IPv4 と IPv6 でサポートされています。
- RoCE v2 は、NVGRE、NetFlow、および VMQ 機能と同じ vNIC インターフェイスでは使用できません。
- RoCE v2 は usNIC では使用できません。
- RoCE v2 は、GENEVE オフロードでは使用できません。

MTU プロパティ :

- VIC ドライバの古いバージョンでは、MTU はスタンドアロンモードの Cisco Intersight サービス プロファイルまたは Cisco IMC vNIC MTU 設定のいずれかから導出されていました。この動作は、Cisco UCS VIC 1400 シリーズ、VIC 14000 シリーズ、および VIC 15000 シリーズでは異なっています。MTU は Windows OS ジャンボ パケットの詳細プロパティから制御されます。
- RoCE v2 の MTU 値は常に 2 の累乗で、最大制限は 4096 です。

- RoCE v2 MTU は、イーサネット MTU から導出されます。
- RoCE v2 MTU は、イーサネット MTU よりも小さい最も高い電力量です。次に例を示します。
 - イーサネット値が 1500 の場合、RoCE v2 の MTU 値は 1024 です。
 - イーサネット値が 4096 の場合、RoCE v2 の MTU 値は 4096 です。
 - イーサネット値が 9000 の場合、RoCE v2 の MTU 値は 4096 です。

Windows NDPKI の動作モード :

- Cisco のネットワーク ダイレクト カーネル プロバイダ インターフェイス (NDPKI) の実装では、モード 1 とモード 2 の 2 つの動作モードがサポートされています。Network Direct Kernel Provider Interface (NDKPI) の実装は、動作がモード 1 かモード 2 かで異なっています。モード 1 はネイティブ RDMA で、モード 2 には RDMA を使用する仮想ポートの構成が関係しています。Cisco は NDPKI Mode 3 の動作をサポートしていません。
- RoCE v2 モード 1 の推奨されるデフォルトのアダプタ ポリシーは、Win-HPN-SMBd です。
- RoCE v2 モード 2 の推奨されるデフォルトのアダプタ ポリシーは、MQ-SMBd です。
- モード 2 操作の RoCE v2 対応 vNIC では、QoS ホスト制御ポリシーがフルに設定されている必要があります。
- モード 2 にはモード 1 が含まれています。モード 2 を動作させるには、モード 1 を有効にする必要があります。
- Windows の場合、RoCE v2 インターフェイスは、MSI および MSIx 割り込みモードを両方サポートします。デフォルトの割り込みモードは MSIx です。RoCE v2 のプロパティを使用してインターフェイスが構成されている場合、Cisco では割り込みモードを変更しないことを推奨します。

ダウングレードに関する制限事項 :

- Cisco では、サポートされていないファームウェア リリースにダウングレードする前に、RoCE v2 の設定を削除することを推奨しています。設定が削除または無効になっていない場合、ダウングレードは失敗します。

Windows での RoCE v2 モード 1 および 2 の構成の概要

Windows プラットフォームで RoCE v2 を構成するには、まず RoCE v2 モード 1 を構成してから、RoCE v2 モード 2 を構成する必要があります。

- RoCE v2 モード 1 を構成するには、次の手順を実行します。
 - システム QoS ポリシーで no-drop クラスを構成します。CoS 5 のプラチナ (Platinum) が、Cisco Intersight ドメイン プロファイルのデフォルト設定です。

- イーサネットアダプタポリシーを作成するか、イーサネットアダプタポリシーのデフォルト（定義済み）構成である *Win-HPN-SMBd* を使用して、Cisco Intersight でモード 1 を構成します。
- ホストオペレーティングシステムでモード 1 を構成します。
- RoCE v2 モード 2 を構成するには、最初に RoCE v2 モード 1 を構成する必要があります。
- VMMQ 接続を使用してイーサネットアダプタポリシーを構成するか、Cisco Intersight のモード 2 のイーサネットアダプタポリシーで、デフォルト（定義済み）構成である *MQ-SMBd* を使用します。
- ホストオペレーティングシステムでモード 2 を構成します。

Windows の要件

Windows サーバで RoCE v2 向けコンバージドイーサネットを介した RDMA の構成と使用には、次のものがが必要です。

- VIC ドライババージョン 5.4.0 以降
- Cisco UCS 1400 シリーズアダプタを搭載した Cisco UCS M5 B シリーズおよび C シリーズ。
- Cisco UCS VIC 1400、VIC 14000、または VIC 15000 シリーズアダプタを搭載した Cisco UCS M6 B シリーズ、C シリーズ、または X シリーズサーバ。
- Cisco UCS VIC 1400、VIC 14000、または VIC 15000 シリーズアダプタを搭載した Cisco UCS M7 C シリーズ、または X シリーズサーバ。
- Cisco UCS VIC 1400、VIC 14000、または VIC 15000 シリーズアダプタを搭載した Cisco UCS M8 C シリーズサーバ。



(注) すべての Powershell コマンドまたは詳細プロパティの構成は、明示的に説明されていない限り、Windows 2019 および 2022 全体で共通です。

Cisco Intersight でのモード 1 の構成

Cisco Intersight で RoCE v2 モード 1 インターフェイスを構成するには、次の手順に従います。

RDMA パケットドロップの可能性を回避するには、ネットワーク全体で同じ非ドロップ COS が構成されていることを確認してください。次の手順に従えば、システム QoS ポリシーで非ドロップクラスを構成して、RDMA でサポートされているインターフェイス用に使用できます。

手順

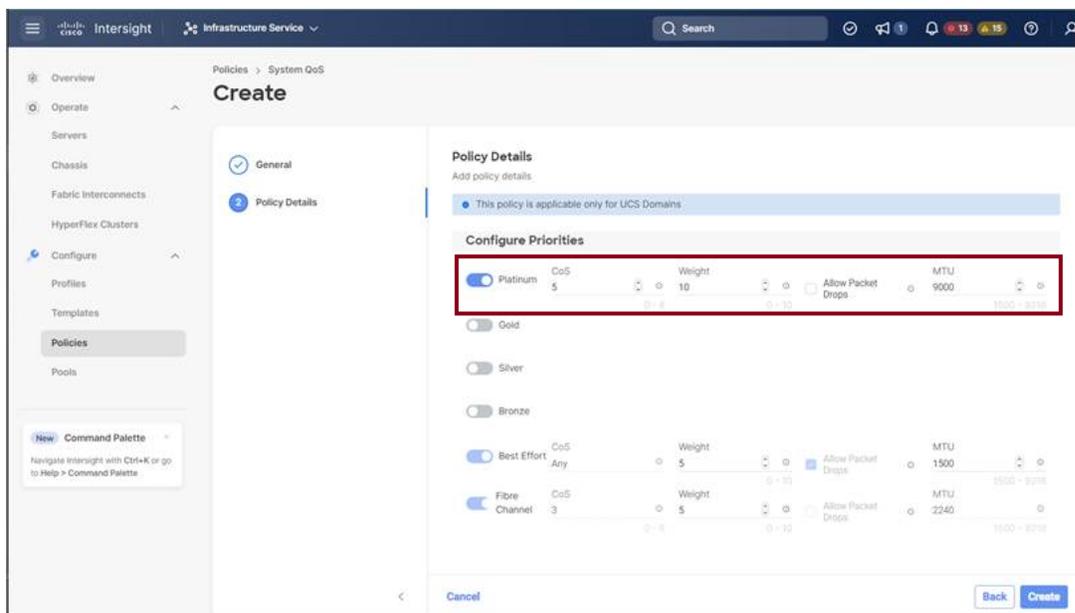
ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS ドメイン (UCS Domain)] プラットフォームタイプを選択し、[システム QoS (System QoS)] を検索または選択して、[Start (開始)] をクリックします。

ステップ 2 [全般 (General)] ページでポリシー名を入力し、[次へ (Next)] をクリックします。次に、[ポリシーの詳細 (Policy Details)] ページで、次のようにシステム QoS ポリシーのプロパティ設定を構成します。

- [優先順位 (Priority)] で、[プラチナ (Platinum)] を選択します。
- [パケットドロップを許可 (Allow Packet Drops)] チェックボックスをオフにします。

(注)

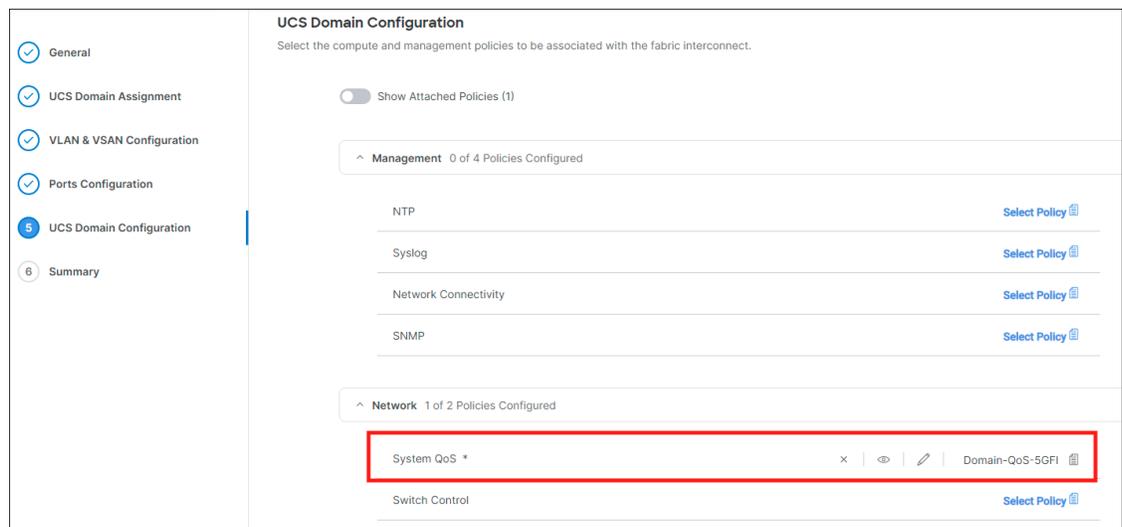
[MTU] フィールドの詳細については、[RDMA over コンバージドイーサネット \(RoCE\) v2 を使用して Windows で SMB ダイレクト サポートを使用するためのガイドライン \(3 ページ\)](#) の MTU のプロパティを参照してください。



ステップ 3 [作成 (Create)] をクリックします。

ステップ 4 システムの QoS ポリシーをドメイン プロファイルに関連付け、展開します。

LAN 接続ポリシーで RoCE 設定を有効化する



(注)
 詳細については、「ドメインポリシーの構成」の「システム QoS ポリシーの作成」および「ドメインプロファイルの構成」を参照してください。

システム QoS ポリシーが正常に作成され、ドメインプロファイルに展開されます。

次のタスク

LAN 接続ポリシーで RoCE v2 vNIC 設定を使用してサーバ プロファイルを構成します。

LAN 接続ポリシーで RoCE 設定を有効化する

モード 1 で RoCE v2 vNIC を構成するには、次の手順を実行します。Cisco Intersight LAN 接続ポリシーでは、次のようにモード 1 構成のイーサネット QoS ポリシーとイーサネットアダプタ ポリシーの RoCE 設定を有効にできます。

手順

- ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS サーバ (UCS Server)] プラットフォーム タイプを選択し、[LAN 接続ポリシー (LAN Connectivity policy)] を検索または選択して、[Start (開始)] をクリックします。
- ステップ 2 ポリシーの [全般 (General)] ページで、ポリシー名を入力し、[ターゲットプラットフォーム (Target Platform)] として [UCS サーバ (スタンドアロン) (UCS Server (Standalone))] または [UCS サーバ (FI アタッチ) (UCS Server (FI-Attached))] を選択し、[次へ (Next)] をクリックします。

ステップ 3 [ポリシーの詳細 (Policy Details)] ページで、[vNIC の追加 (Add vNIC)] をクリックして新しい vNIC を作成します。

ステップ 4 [vNIC の追加 (Add vNIC)] ページで、構成パラメータに従って RoCE vNIC 設定を有効にします。

- [全般 (General)] セクションで、仮想イーサネット インターフェイスの名前を入力します。
- スタンドアロンサーバの [Consistent Device Naming (CDN)] セクションまたは FI アタッチサーバの [フェールオーバー (Failover)] セクションで、次の手順を実行します。
 - [イーサネット QoS (Etehrnet QoS)] の下にある [ポリシーの選択 (Seletct Policy)] リンクをクリックします。[新規作成 (Create New)] ボタンを使用して、次のプロパティ設定で新しいイーサネット QoS ポリシーを作成します。

- [MTU] で、**1500**、**4096**、または **9000** を選択するか、入力します。

- [優先順位 (Priority)] で、[プラチナ (Platinum)] または 任意の **no-drop** を選択します。

- [サービスクラス (Class of Service)] で、**5** を選択するか、入力します。

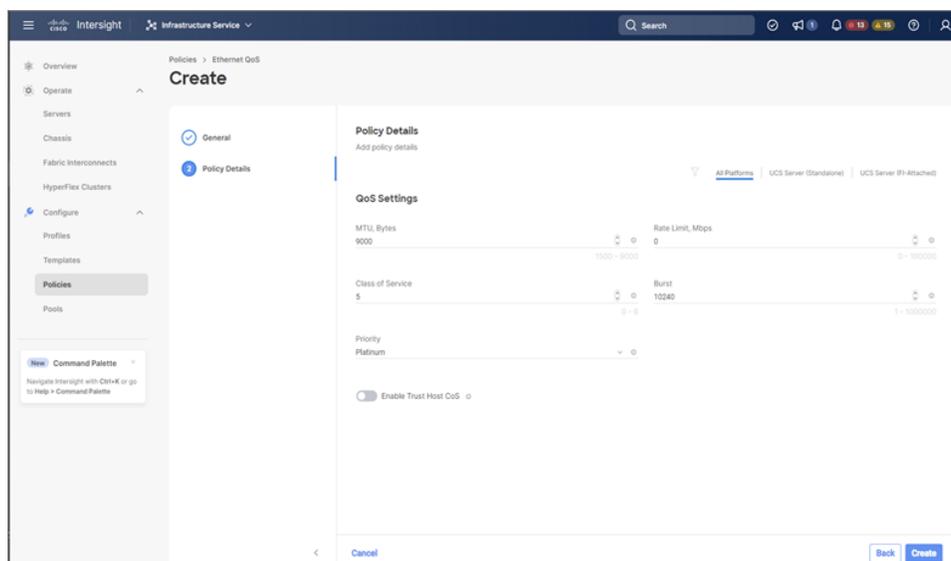
(注)

このプロパティは、スタンドアロンサーバでのみ使用できます。

- [トラスト ホスト CoS を有効にする (Enable Trust Host CoS)] トグルボタンをスライドします。

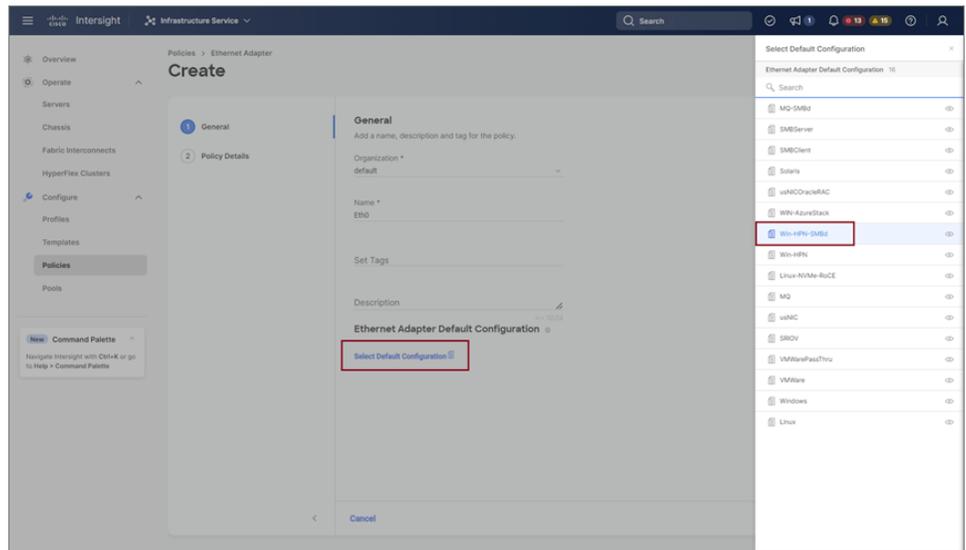
(注)

このプロパティは、Intersight 管理モードのサーバでのみ使用できます。



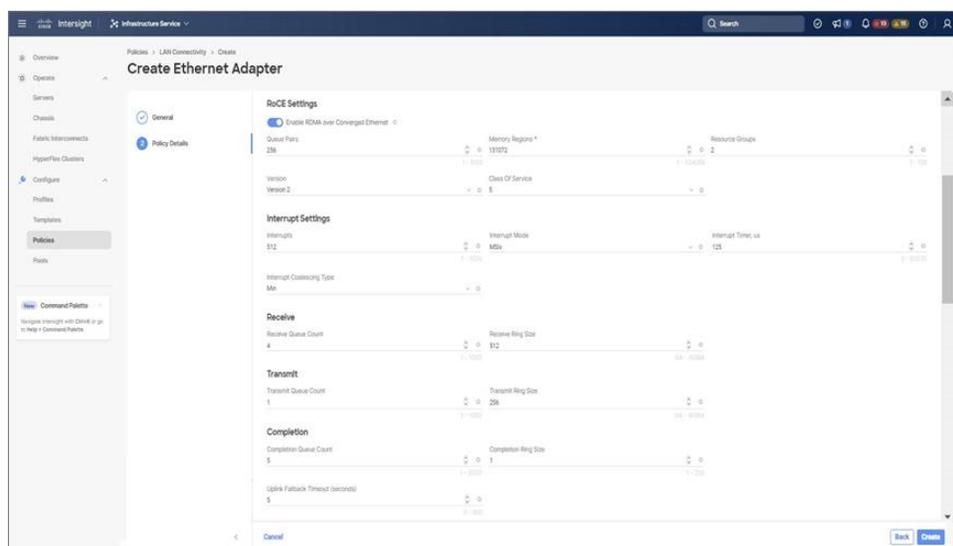
- [イーサネットアダプタ (Ethernet Adapter)] の下の [ポリシーの選択 (Select Policy)] リンクをクリックします。[イーサネットアダプタポリシーの作成 (Create an Ethernet Adapter Policy)] をクリックして、次を実行します。

- [デフォルト設定の活用 (Use the Default Configuration)] : [新規作成 (Create New)] をクリックして、新しいポリシーを作成します。[全般 (General)] ページでポリシーの名前を入力し、[イーサネットアダプタのデフォルト構成 (Ethernet Adapter Default Configuration)] の下で [デフォルト構成の選択 (Select Default Configuration)] をクリックし、事前定義されたイーサネットアダプタのデフォルト構成である [Win-HPN-SMBd] を検索して選択します。[次へ (Next)] をクリックし、[作成 (Create)] をクリックします。



- [ポリシーでの RoCE 設定の構成 (Configure RoCE Settings in the policy)] : [新規作成 (Create New)] をクリックして、新しいポリシーを作成します。[全般 (General)] ページで、ポリシーの名前を入力します。右側のペインの [ポリシーの詳細 (Policy Details)] ページで、次のプロパティ設定を使用し、[次へ (Next)]、[作成 (Create)] の順にクリックします。「」に設定されます。

- [コンバージドイーサネット上の RDMA を有効にする (Enable RDMA over Converged Ethernet)] をスライドして、有効にします。
- [キューペア (Queue Pairs)] で、256 を選択するか、入力します。
- [メモリ領域 (Memory Regions)] で、131072 を選択するか、入力します。
- [リソースグループ (Resource Groups)] で、2 を選択するか、入力します。
- [バージョン (Version)] で、[バージョン 2 (Version 2)] を選択します。



- **[追加 (Add)]** をクリックして新しい vNIC 設定を追加し、保存します。

(注)

LAN接続ポリシーを作成するには、*が付いたすべてのフィールドが必須です。それらのフィールドが入力されていること、または適切なポリシーが選択されていることを確認します。

ステップ 5 **[作成 (Create)]** をクリックし、RoCE v2 プロパティ設定によって LAN 接続ポリシーを完成させます。

ステップ 6 LAN 接続ポリシーをサーバプロファイルに関連付け、展開します。

(注)

詳細については、[UCS サーバポリシー](#)および[UCS サーバプロファイルの LAN 接続ポリシー](#)、[イーサネット QoS ポリシー](#)の作成、および[イーサネットアダプタポリシー](#)の作成を参照してください。

イーサネット QoS ポリシーとイーサネットアダプタポリシーの vNIC 設定を含む LAN 接続ポリシーが正常に作成され、RoCE v2構成を有効にするためのサーバプロファイルが展開されます。

次のタスク

RoCE v2 のポリシー構成が完了したら、サーバを再起動し、ホストの RoCE v2 モード 1 構成を続行します。

ホストシステムでの SMB ダイレクト モード 1 の設定

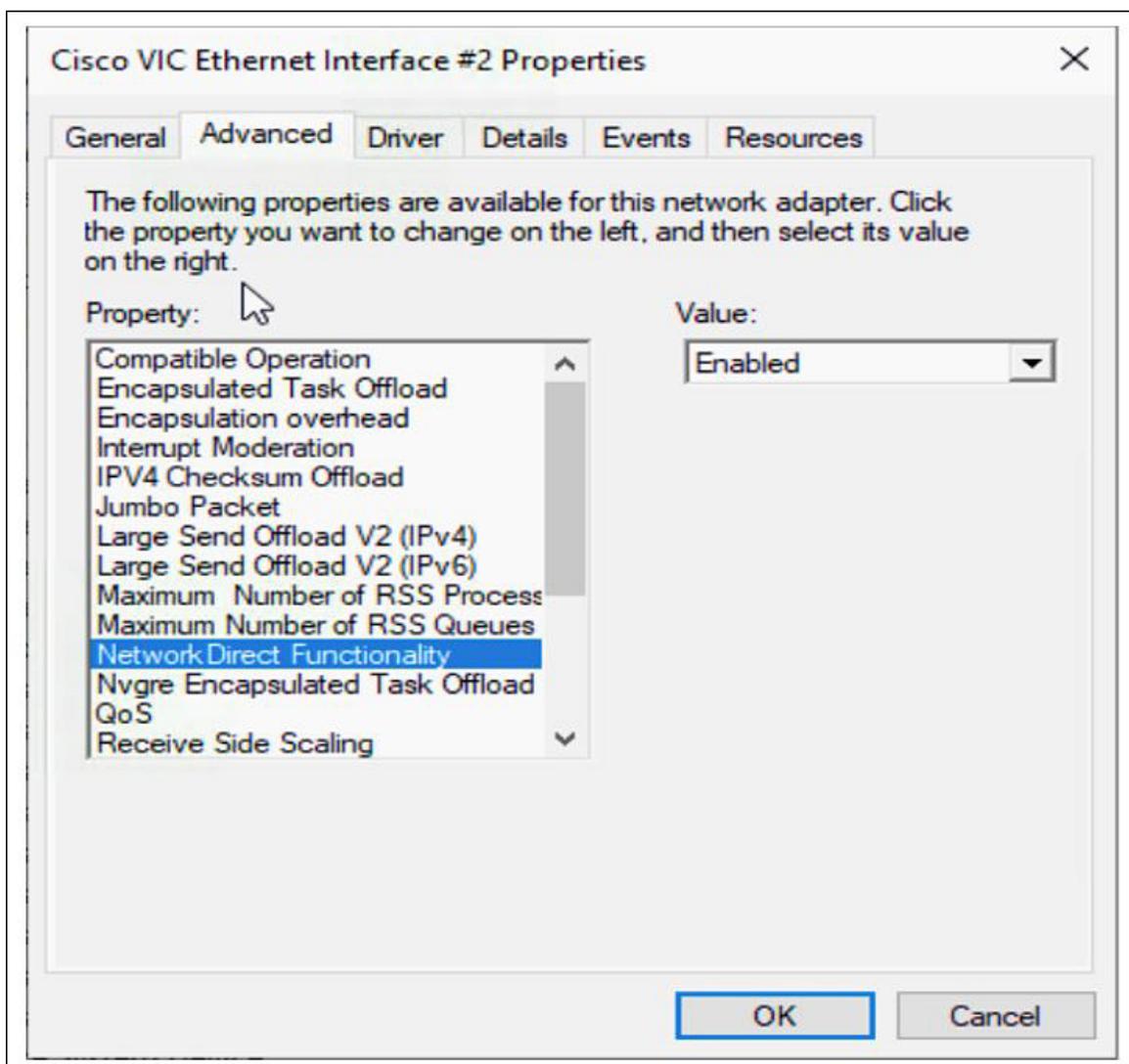
2 個のホスト インターフェイスで smb クライアントと smb サーバ間の接続を設定します。これらのサーバのそれぞれについて、smb クライアントおよび smb サーバで、次の説明に従って RoCE v2 対応 vNIC を設定します。

始める前に

Cisco Intersight で RoCE v2 をモード 1 に構成します。

手順

-
- ステップ 1 Windows ホストで、[デバイス マネージャ (Device Manager)] に移動し、適切な Cisco VIC インターネット インターフェイスを選択します。
 - ステップ 2 [ツール (Tools)] > [コンピュータ管理 (Computer Management)] > [デバイス マネージャ (Device Manager)] > [ネットワーク アダプタ (Network Adapter)] > [VIC ネットワーク アダプタ (VIC Network Adapter)] > [プロパティ (Properties)] > [アドバンスド (Advanced)] > [ネットワーク ダイレクト機能 (Network Direct Functionality)] に移動します。smb サーバと smb クライアント両方の vNICs に対してこの操作を実行します。



ステップ 3 PowerShell を使用して、ホスト オペレーティング システムで RoCE が有効になっていることを確認します。

Get-NetOffloadGlobalSetting コマンドは、NetworkDirect が有効になっていることを示します。

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

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

(注)

NetworkDirect 設定が無効と表示されている場合は、コマンド Set-NetOffloadGlobalSetting -NetworkDirect enabled を使用して有効にします。

ステップ 4 Powershell を起動し、次のコマンドを入力します。

```
get-SmbClientNetworkInterface
```

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

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

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

ステップ 5 `enable - netadapterrdma [-name] ["Ethernetname"]` と入力します

ステップ 6 次の手順に従って、ホストで全体的な RoCE v2 モード 1 の構成を確認します。

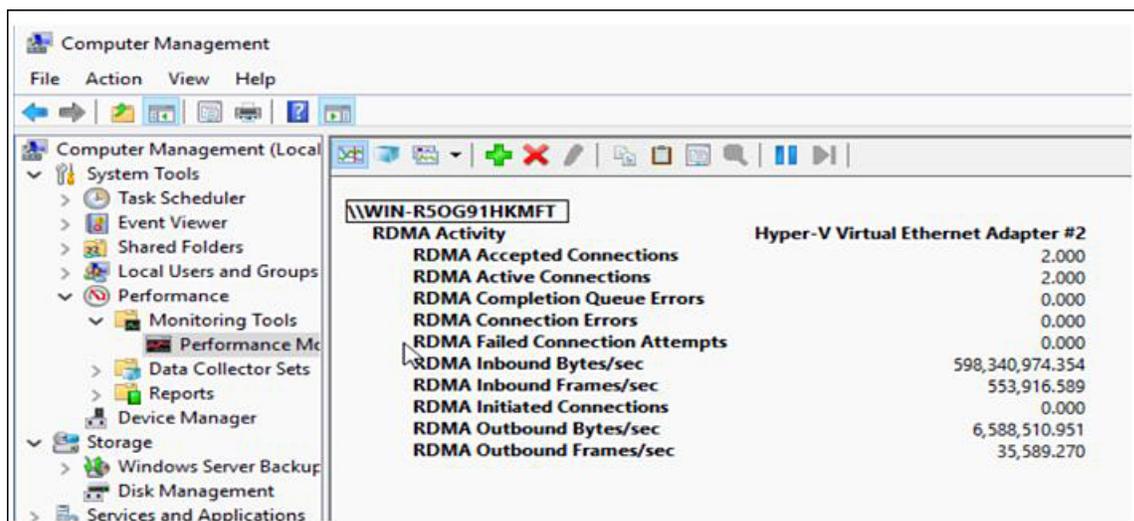
- a) Powershell コマンド `netstat -xan` を使用して、smb クライアントと smb サーバ Windows ホストの両方のリスナーを確認します。リスナーはコマンド出力に表示されます。

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

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

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

- b) smb-client サーバ ファイル共有に移動し、I/O 操作を開始します。
- c) パフォーマンス モニタに移動し、RDMA アクティビティが表示されていることを確認します。



ステップ 7 Powershell コマンドウィンドウで、`netstat -xan` 出力コマンドを使用して接続エントリをチェックして、表示されていることを確認します。コマンドプロンプトから `netstat -xan` を実行することもできます。netstat-xan 出力に接続エントリが表示されている場合は、クライアントとサーバの間で RoCE v2 モード 1 接続が正しく確立されています。

```
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
```

(注)

IP 値は代表のみです。

ステップ 8 デフォルトでは、Microsoft の SMB ダイレクトは RDMA インターフェイスごとに 2 個の RDMA 接続を確立します。RDMA インターフェイスごとに RDMA 接続数を 1 個または複数の接続数に変更できます。

たとえば、RDMA 接続の数を 4 個に増やすには、PowerShell で次のコマンドを入力します。

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

Cisco Intersight でのモード 2 の設定

モード 2 で RoCE v2 ポリシーを設定するには、次の手順に従います。Cisco Intersight LAN 接続ポリシーでは、次のようにして、イーサネット QoS ポリシーとイーサネットアダプタポリシーの RoCE 設定、およびモード 2 構成の VMMQ アダプタポリシーを有効にできます。

始める前に

モード 1 で RoCE v2 ポリシーを構成します。

手順

- ステップ 1** [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS サーバ (UCS Server)] プラットフォームタイプを選択し、[LAN 接続ポリシー (LAN Connectivity policy)] を検索または選択して、[Start (開始)] をクリックします。
- ステップ 2** ポリシーの [全般 (General)] ページで、ポリシー名を入力し、[ターゲットプラットフォーム (Target Platform)] として [UCS サーバ (スタンドアロン) (UCS Server (Standalone))] または [UCS サーバ (FI アタッチ) (UCS Server (FI-Attached))] を選択し、[次へ (Next)] をクリックします。

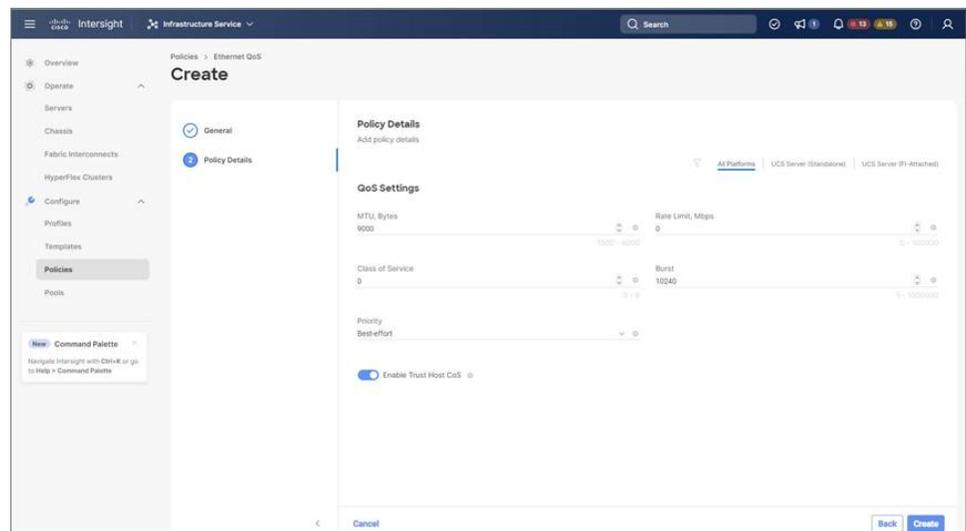
ステップ 3 [ポリシーの詳細 (Policy Details)] ページで、[vNIC の追加 (Add vNIC)] をクリックして新しい vNIC を作成します。

ステップ 4 [vNIC の追加 (Add vNIC)] ページで、構成パラメータに従って RoCE vNIC 設定を有効にします。

- [全般 (General)] セクションで、仮想イーサネット インターフェイスの名前を入力します。
- スタンドアロン サーバの [Consistent Device Naming (CDN)] セクションまたは FI アタッチ サーバの [フェールオーバー (Failover)] セクションで、次の手順を実行します。

- [イーサネット QoS (Ethernet QoS)] の下にある [ポリシーの選択 (Select Policy)] リンクをクリックします。[新規作成 (Create New)] ボタンを使用して、次のプロパティ設定で新しいイーサネット QoS ポリシーを作成します。

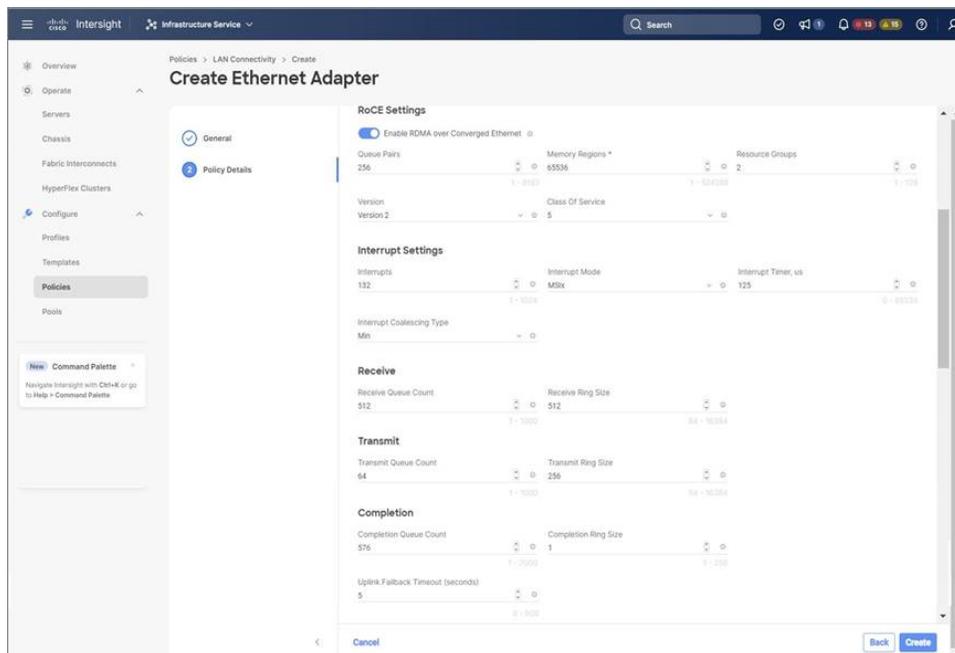
- [MTU] で、**1500**、**4096**、または **9000** を選択するか、入力します。
- [優先順位 (Priority)] で、[ベストエフォート (Best-effort)] を選択するか、入力します。
- [トラストホスト CoS を有効にする (Enable Trust Host CoS)] をスライドして、有効にします。



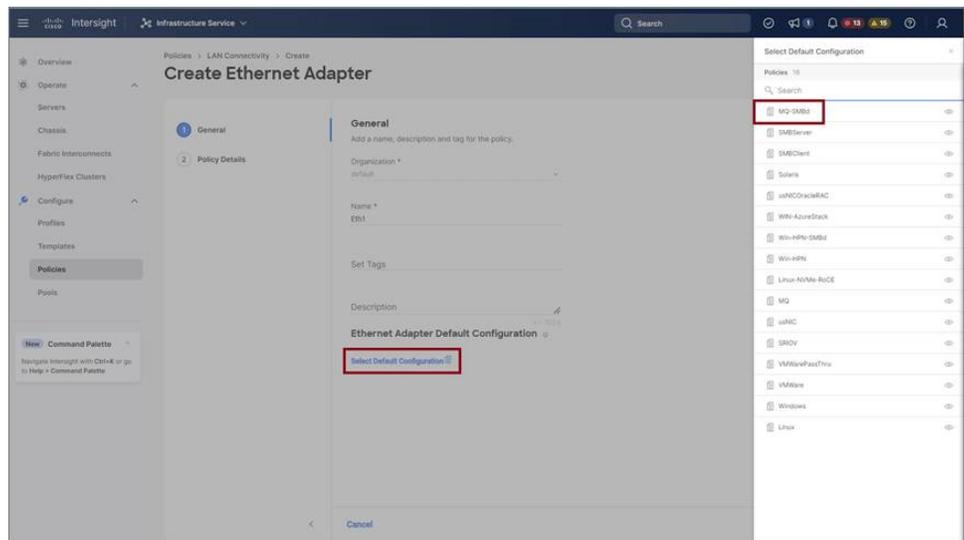
- [イーサネット アダプタ (Ethernet Adapter)] の下の [ポリシーの選択 (Select Policy)] リンクをクリックします。[新規作成 (Create New)] ボタンを使用して、次のプロパティ設定で新しいイーサネット アダプタ ポリシーを作成します。

- [コンバージド イーサネット上の RDMA を有効にする (Enable RDMA over Converged Ethernet)] をスライドして、有効にします。
- [キュー ペア (Queue Pairs)] で、**256** を選択するか、入力します。
- [メモリー レジヨン (Memory Regions)] で、**65536** を選択するか、入力します。
- [リソース グループ (Resource Groups)] で、**2** を選択するか、入力します。

- [バージョン (Version)] で、[バージョン 2 (Version 2)] を選択します。
- [サービスクラス (Class of Service)] で、5 を選択するか、入力します。



- [接続 (Connection)] セクションで、VMQ 接続の次のプロパティ設定を使用し、VMMQ アダプタ ポリシーを作成します。
 - 接続については、[VMQ] を選択します。
 - スライドボタンで、[仮想マシン マルチキューを有効にする (Enable Virtual Machine Multi-Queue)] を有効にします。
 - [サブ vNIC 数 (Number of Sub vNICs)] で、4 を選択するか、入力します。
 - [VMMQ アダプタ ポリシー (VMMQ Adapter Policy)] については、[VMMQ アダプタ ポリシー (VMMQ Adapter Policy)] の下にある、[ポリシーの選択 (Select Policy)] リンクをクリックし、次の手順を実行します。
 - 新しい SIG ポリシーを作成するには、[新規作成 (Create New)] をクリックします。[全般 (General)] ページで、ポリシーの名前を入力して [デフォルト構成の選択 (Select Default Configuration)] をクリックして検索し、事前定義された VMMQ アダプタのデフォルト構成である [MQ-SMBd] を選択します。
- 注目**
- [ポリシーの詳細 (Policy Details)] ページでは、デフォルト設定を保持します。事前定義されたパラメータを変更しないでください。
- [次へ (Next)] をクリックし、[作成 (Create)] をクリックします。



- [追加 (Add)] をクリックして新しい vNIC 設定を追加し、保存します。

(注)

* が付いているすべてのフィールドは必須です。適切なポリシーに従って入力または選択されていることを確認してください。

ステップ 5 [作成 (Create)] をクリックし、RoCE v2 プロパティ設定によって LAN 接続ポリシーを完成させます。

ステップ 6 LAN 接続ポリシーをサーバプロファイルに関連付けます。

(注)

イーサネット QoS の作成、イーサネットアダプタポリシー、および VMMQ アダプタポリシーの詳細については、[UCS サーバポリシーの構成](#) および [UCS サーバプロファイルの構成](#) を参照してください。

イーサネット QoS ポリシー、イーサネットアダプタポリシー、および VMMQ アダプタポリシーを使用した LAN 接続ポリシーが正常に作成および展開され、RoCE v2 構成が有効になります。

次のタスク

RoCE v2 のポリシー構成が完了したら、サーバを再起動し、ホストオペレーティングシステムで RoCE v2 モード 2 の構成を続行します。

ホスト システムでのモード 2 の設定

このタスクでは、Windows Server 2019 および Windows Server 2022 と互換性のある Hyper-V 仮想化ソフトウェアを使用します。

RoCEv2 モード 2 用にホスト オペレーティング システムを構成するには、次の手順に従います。

始める前に

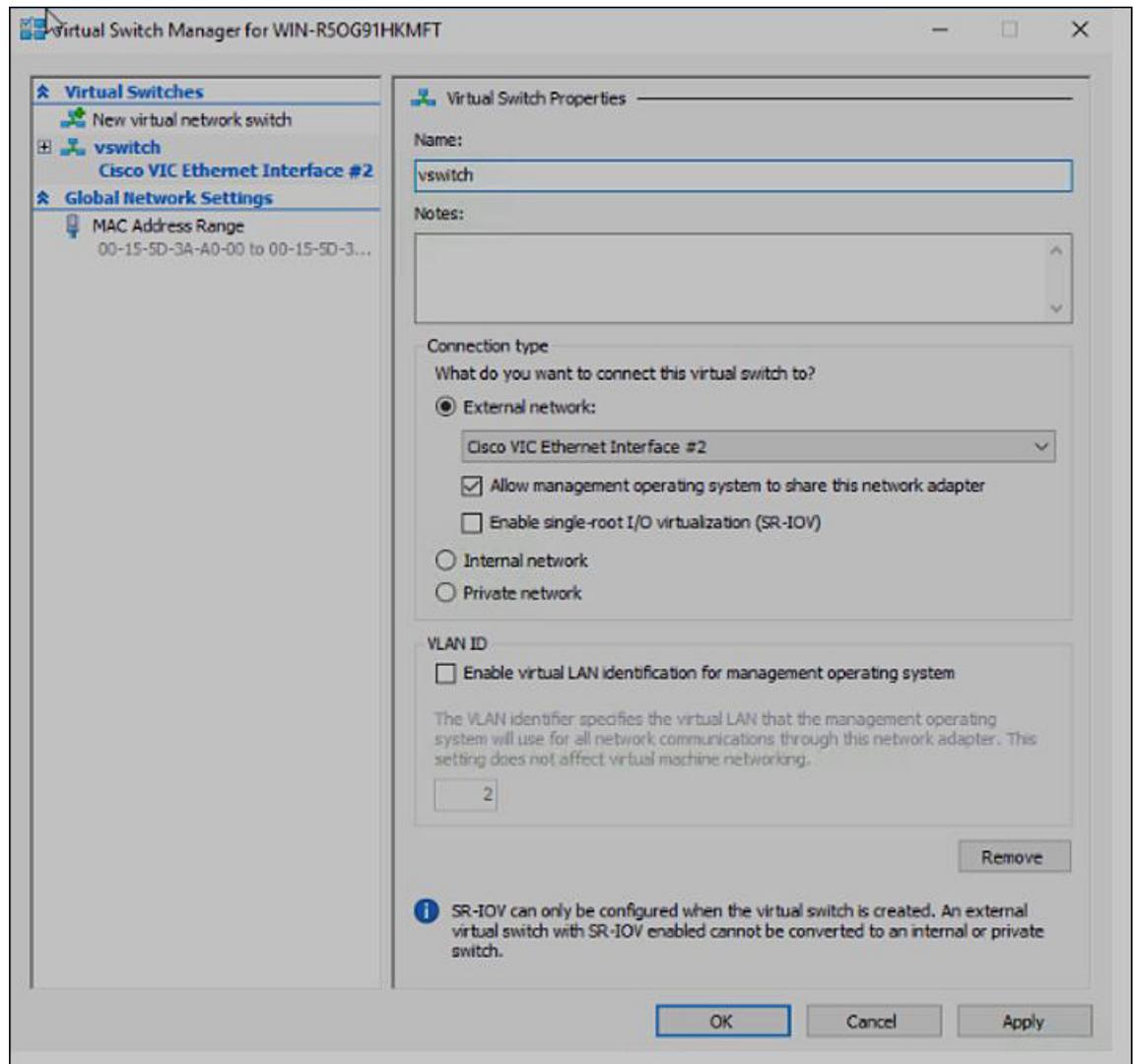
- Cisco Intersight とホストの両方に対して、モード 1 の接続を構成して確認します。
- Cisco Intersight でモード 2 を構成します。

手順

ステップ 1 Hyper-V スイッチ マネージャに移動します。

ステップ 2 RoCE v2 対応イーサネット インターフェイスの新しい仮想ネットワーク スイッチ (vswitch) を作成します。

- a) [外部ネットワーク (External Network)] を選択し、[VIC イーサネット インターフェイス 2 (VIC Ethernet Interface 2)] および [管理オペレーティング システムでこのネットワーク アダプタの共有を許可する (Allow management operating system to share this network adapter)] を選択します。
- b) [OK] をクリックして、仮想スイッチを作成します。



Powershell インターフェイスを起動します。

ステップ 3 デフォルト以外の vPort を設定し、次の Powershell コマンドを使用して RDMA を有効にします。

```
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) 次の Powershell コマンドを使用して、設定スイッチを設定します。

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

これにより、スイッチが作成されます。インターフェイスを表示するには、次を使用します。

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

再度入力すると、新しい vport が表示されます。

```
get-netadapterrdma
```

- b) vport を追加します。

```
add-vmNetworkAdapter -switchname setswtch -name svp1
```

再度入力すると、新しい vport が表示されます。

```
get-netadapterrdma
```

- c) vport で RDMA を有効にします。

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

ステップ 4 両方のサーバの RDMA 対応 vport で IPV4 アドレスを設定します。

ステップ 5 smb サーバで共有を作成し、smb クライアントで共有をマッピングします。

- ホスト システムの smb クライアントおよび smb サーバ用に、前述の方法で RoCE v2 対応 vNIC を構成します。
- 両方のサーバに同じ IP サブネットと同じ固有の vlan を使用して、両方のサーバでプライマリ ファブリックとサブ vNICs の IPV4 アドレスを設定します。
- smb サーバで共有を作成し、smb クライアントで共有をマッピングします。

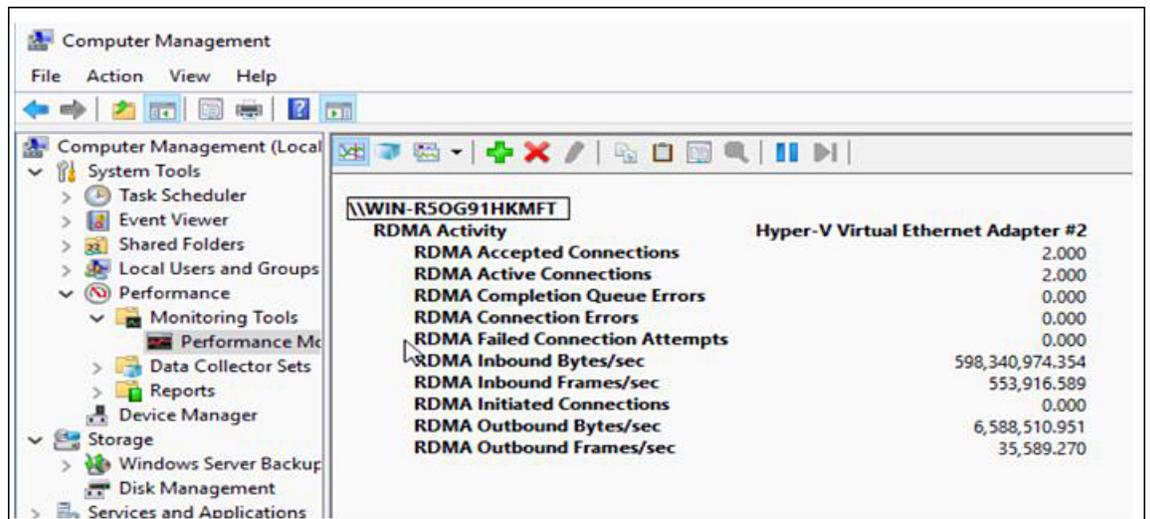
ステップ 6 モード 2 設定を確認します。

- Powershell コマンド `netstat -xan` を使用して、リスナーとそれらに関連付けられている IP アドレスを表示します。

```
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>
```

- smb クライアントのファイル共有で RDMA I/O を開始します。



c) `Netstat-xan` コマンドを再度発行し、接続エントリが表示されていることを確認します。

```
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>
```

次のタスク

必要に応じて、すべての項目のトラブルシューティングを行います。

Cisco Intersight を使用した RoCE v2 インターフェイスの削除

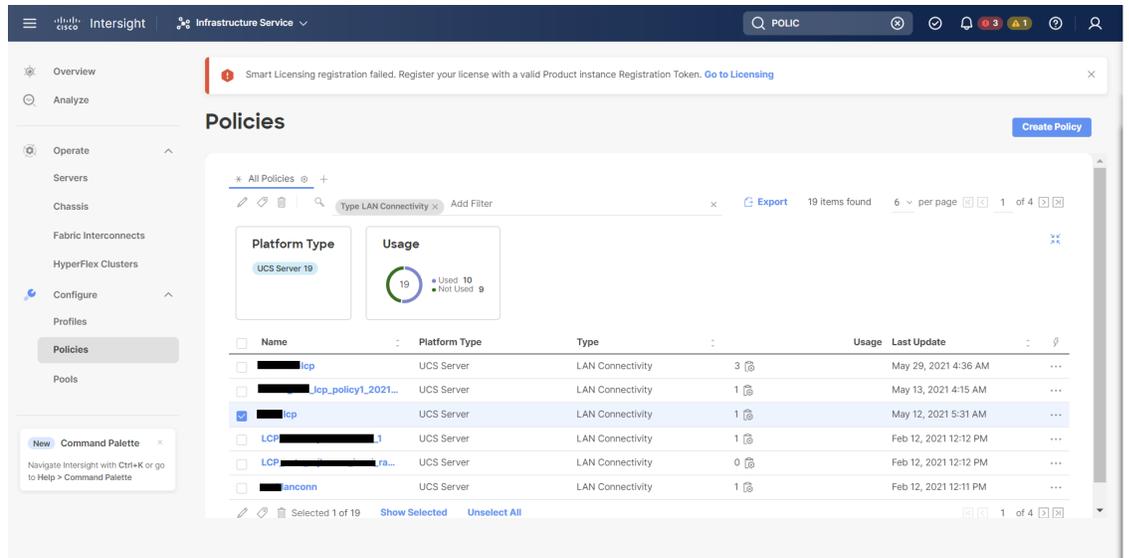
RoCE v2 インターフェイスを削除するには、次の手順を実行します。

手順

ステップ 1 [構成 (CONFIGURE)]>[ポリシー (Policies)]に移動します。[フィルタの追加 (Add Filter)] フィールドで、[タイプ: LAN 接続 (Type: LAN Connectivity)]を選択します。

ステップ 2 RoCE V2 構成用に作成された適切な LAN 接続ポリシーを選択し、ポリシー リストの上部または下部にある削除アイコンを使用します。

ステップ 3 ポリシーを削除するには、[削除 (Delete)] をクリックします。



The screenshot shows the Cisco Intersight interface for managing policies. The main content area is titled "Policies" and displays a list of policies for UCS Servers. The table below shows the details of the policies.

Name	Platform Type	Type	Usage	Last Update
[redacted] lcp	UCS Server	LAN Connectivity	3	May 29, 2021 4:36 AM
[redacted] lcp_policy1_2021...	UCS Server	LAN Connectivity	1	May 13, 2021 4:15 AM
[redacted] lcp	UCS Server	LAN Connectivity	1	May 12, 2021 5:31 AM
[redacted] LCP [redacted]	UCS Server	LAN Connectivity	1	Feb 12, 2021 12:12 PM
[redacted] LCP [redacted]	UCS Server	LAN Connectivity	0	Feb 12, 2021 12:12 PM
[redacted] anconn	UCS Server	LAN Connectivity	1	Feb 12, 2021 12:11 PM

ステップ 4 RoCE v2 構成を削除したら、サーバプロファイルを再展開し、サーバを再起動します。



第 3 章

Linux での RoCEv2 を使用した NVMeoF の構成

- [Linux 上で RoCE v2 を使用するファブリック \(NVMeoF\) を介して NVMe を使用する際のガイドライン \(25 ページ\)](#)
- [Linux の要件 \(26 ページ\)](#)
- [Cisco Intersight での RoCE v2 for NVMeoF の構成 \(26 ページ\)](#)
- [ホストシステムでの NVMeoF の RoCE v2 の構成 \(31 ページ\)](#)
- [デバイス マッパー マルチパスの設定 \(35 ページ\)](#)
- [Cisco Intersight を使用した RoCE v2 インターフェイスの削除 \(36 ページ\)](#)

Linux 上で RoCE v2 を使用するファブリック (NVMeoF) を介して NVMe を使用する際のガイドライン

一般的なガイドラインと制限事項

- Cisco では、[UCS ハードウェアとソフトウェアの互換性](#)をチェックして、NVMeoF のサポートを判断することを推奨します。NVMeoF は、Cisco UCS B シリーズ、C シリーズ、および X シリーズのサーバでサポートされています。
- RoCE v2 を使用した RDMA 上の NVMe は、Cisco UCS VIC 1400、VIC 14000、および VIC 15000 シリーズのアダプタでサポートされています。
- RoCE v2 インターフェイスを作成する際には、Cisco Intersight が提供する Linux-NVMe-RoCE アダプタ ポリシーを使用します。
- Ethernet Adapter ポリシーでは、キューペア、メモリ領域、リソースグループ、および優先度の設定値を、Cisco が提供するデフォルト値以外に変更しないでください。キューペア、メモリ領域、リソースグループ、および優先度の設定が異なると、NVMeoF の機能が保証されない可能性があります。
- RoCE v2 インターフェイスを構成する場合は、Cisco.com からダウンロードした `enic` と `enic_rdma` の両方のバイナリドライバを使用して、一致する `enic` と `enic_rdma` ドライバの

セットをインストールします。inboxenic ドライバを使用して Cisco.com からダウンロードしたバイナリ enic_rdma ドライバを使用しようとしても、機能しません。

- RoCE v2 は、アダプタごとに最大 2 つの RoCE v2 対応インターフェイスをサポートしません。
- NVMeoF ネームスペースからのブートはサポートされていません。
- レイヤ 3 ルーティングはサポートされていません。
- RoCE v2 はボンディングをサポートしていません。
- システムクラッシュ時に crashdump を NVMeoF ネームスペースに保存することはサポートされていません。
- NVMeoF は、usNIC、VxLAN、VMQ、VMMQ、NVGRE、GENEVE オフロード、および DPDK 機能とともに使用することはできません。
- Cisco Intersight は、RoCE v2 対応の vNIC に対してファブリック フェールオーバーをサポートしません。
- Quality of Service (QoS) no drop クラス構成は、Cisco Nexus 9000 シリーズスイッチなどのアップストリームスイッチで適切に構成する必要があります。QoS の設定は、異なるアップストリームスイッチ間で異なります。
- スパニング ツリー プロトコル (STP) によって、フェールオーバーまたはフェールバック イベントが発生したときに、ネットワーク接続が一時的に失われる可能性があります。この問題が発生しないようにするには、アップリンクスイッチで STP を無効にします。

Linux の要件

Linux での RoCE v2 の構成と使用には、次のものがが必要です。

- InfiniBand カーネル API モジュール ib_core
- NVMeoF 接続をサポートするストレージアレイ

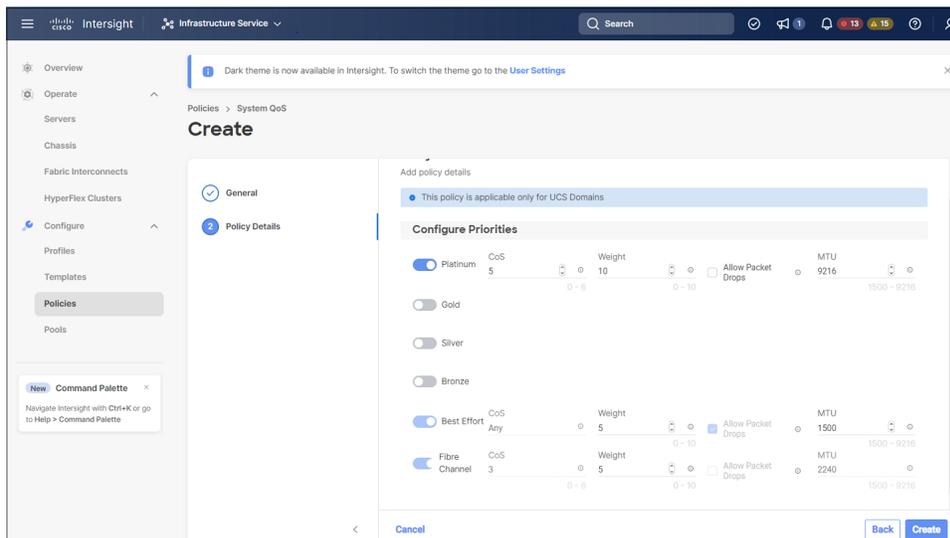
Cisco Intersight での RoCE v2 for NVMeoF の構成

Cisco Intersight で RoCE v2 インターフェイスを構成するには、次の手順に従います。

RDMA パケット ドロップの可能性を回避するには、ネットワーク全体で同じ非ドロップ COS が構成されていることを確認してください。次の手順に従えば、システム QoS ポリシーで非ドロップクラスを構成して、RDMA でサポートされているインターフェイス用に使用できます。

手順

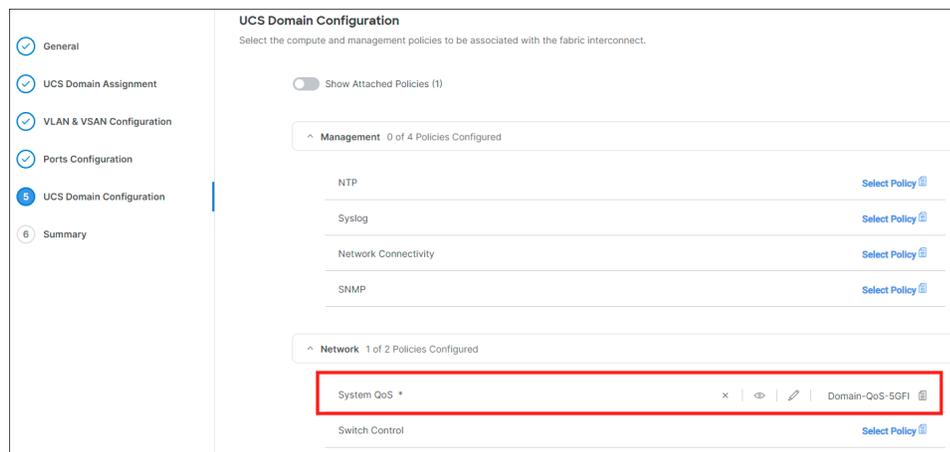
- ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS ドメイン (UCS Domain)] プラットフォームタイプを選択し、[システム QoS (System QoS)] を検索または選択して、[Start (開始)] をクリックします。
- ステップ 2 [全般 (General)] ページでポリシー名を入力し、[次へ (Next)] をクリックします。次に、[ポリシーの詳細 (Policy Details)] ページで、次のようにシステム QoS ポリシーのプロパティ設定を構成します。
- [優先順位 (Priority)] で、[プラチナ (Platinum)] を選択します。
 - [パケットドロップを許可 (Allow Packet Drops)] チェックボックスをオフにします。
 - [MTU] については、値を 9216 に設定します。



ステップ 3 [作成 (Create)] をクリックします。

ステップ 4 システム QoS ポリシーをドメインプロファイルに関連付けます。

LAN 接続ポリシーで RoCE 設定を有効化する



(注)

詳細については、「ドメインポリシーの構成」の「システム QoS ポリシーの作成」および「ドメインプロファイルの構成」を参照してください。

システム QoS ポリシーが正常に作成され、ドメインプロファイルに展開されます。

次のタスク

LAN 接続ポリシーで RoCE v2 vNIC 設定を使用してサーバプロファイルを構成します。

LAN 接続ポリシーで RoCE 設定を有効化する

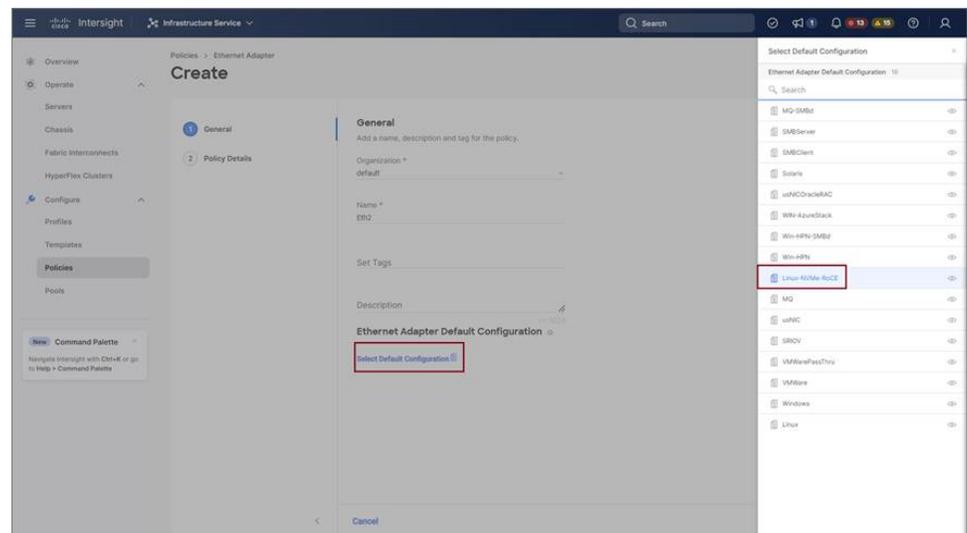
RoCE v2 vNIC を構成するには、次の手順に従います。Cisco Intersight LAN 接続ポリシーでは、次のように Linux 構成向けのイーサネットアダプタポリシーの RoCE 設定を有効にできます。

手順

- ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS サーバ (UCS Server)] プラットフォームタイプを選択し、[LAN 接続ポリシー (LAN Connectivity policy)] を検索または選択して、[Start (開始)] をクリックします。
- ステップ 2 ポリシーの [全般 (General)] ページで、ポリシー名を入力し、[ターゲットプラットフォーム (Target Platform)] として [UCS サーバ (スタンドアロン) (UCS Server (Standalone))] または [UCS サーバ (FI アタッチ) (UCS Server (FI-Attached))] を選択し、[次へ (Next)] をクリックします。
- ステップ 3 [ポリシーの詳細 (Policy Details)] ページで、[vNIC の追加 (Add vNIC)] をクリックして新しい vNIC を作成します。

ステップ 4 [vNIC の追加 (Add vNIC)] ページで、構成パラメータに従って RoCE v2 vNIC を有効にします。

- a) [全般 (General)] セクションで、仮想イーサネット インターフェイスの名前を入力します。
- b) スタンドアロン サーバの [Consistent Device Naming (CDN)] セクションまたは FI アタッチ サーバの [フェールオーバー (Failover)] セクションで、次の手順を実行します。
 - [イーサネットアダプタ (Ethernet Adapter)] の下で、[ポリシーの選択 (Select Policy)] をクリックします。
 - [ポリシーの選択 (Select Policy)] ウィンドウで、[新規作成 (Create New)] をクリックして、イーサネットアダプタ ポリシーを作成します。
 - [全般 (General)] ページで、ポリシーの名前を入力し、[デフォルトの構成を選択 (Select Default Configuration)] をクリックします。[デフォルトの構成 (Default Configuration)] ウィンドウで [Linux-NVMe-RoCE] を検索して選択し、[次へ (Next)] をクリックします。
 - [ポリシーの詳細 (Policy Details)] で、RoCE のデフォルト構成パラメータを確認し、[作成 (Create)] をクリックします。



- [追加 (Add)] をクリックして設定を保存し、新しい vNIC を追加します。

(注)

* が付いているすべてのフィールドは必須です。適切なポリシーに従って入力または選択されていることを確認してください。

ステップ 5 [作成 (Create)] をクリックし、RoCE v2 設定によって LAN 接続ポリシーを完成させます。

ステップ 6 LAN 接続ポリシーをサーバプロファイルに関連付けます。

(注)

詳細については、「[UCS サーバポリシーの構成](#)」の「[LAN 接続ポリシーの作成](#)」および「[イーサネットアダプタポリシーの作成](#)」および「[UCS サーバプロファイルの構成](#)」を参照してください。

イーサネットアダプタポリシーの vNIC 設定を含む LAN 接続ポリシーが正常に作成および展開され、RoCE v2 設定が有効になります。

次のタスク

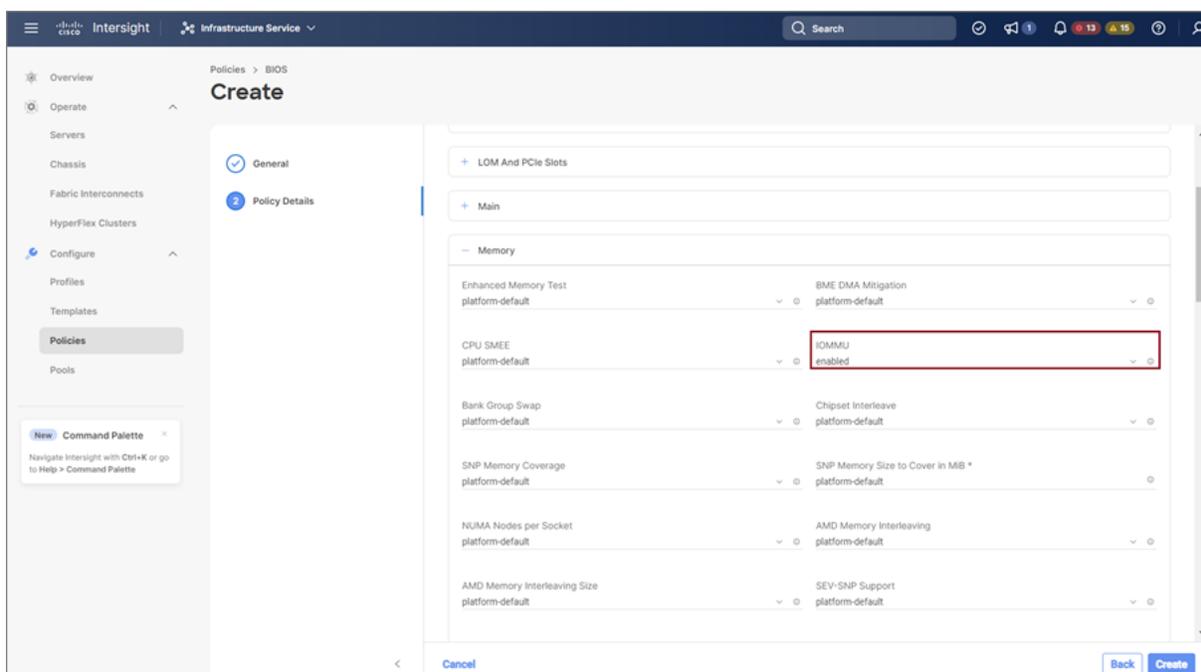
RoCE v2 のポリシー構成が完了したら、続いて、BIOS ポリシーで IOMMU を有効にします。

IOMMU BIOS 設定の有効化

Linux カーネルで IOMMU を有効にする前に、次の手順を実行して、RoCE v2 vNIC を使用するようサーバのサービスプロファイルを構成し、IOMMU BIOS ポリシーを有効にします。

手順

-
- ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS サーバ (UCS Server)] プラットフォームタイプを選択し、[BIOS] を検索または選択して、[Start (開始)] をクリックします。
 - ステップ 2 [全般 (General)] ページで、ポリシーの名前を入力し、[次へ (Next)] をクリックします。
 - ステップ 3 [ポリシーの詳細 (Policy Details)] ページで、次の BIOS を構成します。
 - a) [すべてのプラットフォーム (All Platforms)] を選択します。
 - b) [メモリ (Memory)] グループを展開します。
 - c) [IOMMU] ドロップダウンリストで、IOMMU 構成の設定を有効にする BIOS 値を選択します。



ステップ 4 [作成 (Create)] をクリックします。

ステップ 5 BIOS ポリシーをサーバプロファイルに関連付け、サーバを再起動します。

(注)

詳細については、「[サーバポリシーの構成](#)」の「[BIOS ポリシーの作成](#)」および「[サーバプロファイルの構成](#)」を参照してください。

BIOS ポリシーが正常に作成され、サーバプロファイルに展開されます。

次のタスク

ホストシステムで RoCE v2 for NVMeoF を構成します。

ホストシステムでの NVMeoF の RoCE v2 の構成

始める前に

IOMMU 対応 BIOS ポリシーを使用して、RoCE v2 vNIC を使用するサーバのサービスプロファイルを設定します。

手順

ステップ 1 編集のために /etc/default/grub ファイルを開きます。

ステップ 2 GRUB_CMDLINE_LINUX の末尾に intel_iommu=on を追加します。

```
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"
```

ステップ 3 ファイルを保存した後、新しい grub.cfg ファイルを生成します。

レガシー ブートの場合：

```
# grub2-mkconfig -o /boot/grub2/grub.cfg
```

UEFI ブートの場合：

```
# grub2-mkconfig -o /boot/grub2/efi/EFI/redhat/grub.cfg
```

ステップ 4 サーバをリブートします。IOMMU を有効にした後で、変更を反映するためにサーバを再起動します。

ステップ 5 サーバが intel_iommu=on オプションを使用して起動されていることを確認します。

```
cat /proc/cmdline | grep iommu
```

出力の最後に含まれることに注意してください。

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

次のタスク

enic および enic_rdma ドライバをダウンロードします。

Cisco enic および enic_rdma ドライバのインストール

enic_rdma ドライバには enic ドライバが必要です。enic および enic_rdma ドライバをインストールする場合は、Cisco.com で一致する enic および enic_rdma ドライバのセットをダウンロードして使用してください。inbox enic ドライバを使用して Cisco.com からダウンロードしたバイナリ enic_rdma ドライバを使用しようとしても、機能しません。

手順

ステップ 1 enic および enic_rdma rpm パッケージをインストールします。

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

(注)

enic_rdma のインストール中に、enic_rdmalibnvdimm モジュールは、RHEL 7.7 へのインストールに失敗することがあります。nvdimm-security.conf dracut モジュールは add_drivers 値にスペースを必要とするためです。回避策については、次のリンクの指示に従ってください。

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

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

ステップ 2 enic_rdma ドライバはインストールされていますが、動作中のカーネルでロードされません。サーバを再起動して、実行中のカーネルに enic_rdma ドライバをロードします。

ステップ 3 enic_rdma ドライバと RoCE v2 インターフェイスのインストールを確認します。

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

ステップ 4 nvme-rdma カーネル モジュールをロードします。

```
# modprobe nvme-rdma
```

サーバの再起動後に、nvme-rdma カーネルモジュールがアンロードされます。サーバの再起動ごとに nvme-rdma カーネルモジュールをロードするには、次を使用して nvme_rdma.conf ファイルを作成します。

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

(注)

インストール後の enic_rdma の詳細については、rpm -q -l kmod-enic_rdma コマンドを使用して README ファイルを抽出します。

次のタスク

ターゲットを検出し、NVMe ネームスペースに接続します。システムでストレージへのマルチパス アクセスが必要な場合は、[デバイス マッパー マルチパスの設定 \(35 ページ\)](#) についてのセクションを参照してください。

NVMe ターゲットの検出

NVMe のターゲットを検出し、NVMe ネームスペースを接続するには、次の手順を使用します。

始める前に

まだインストールされていない場合は、**nvme cli** バージョン 1.6 以降をインストールします。



(注) **nvme-cli** バージョン 1.7 以降がインストールされている場合は、下のステップ 2 はスキップします。

RoCEv2 インターフェイスで IP アドレスを設定し、インターフェイスがターゲット IP に対して ping を実行できることを確認します。

手順

ステップ 1 /etc で nvme フォルダを作成し、ホスト nqn を手動で生成します。

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

ステップ 2 settos.sh ファイルを作成し、IB フレームでプライオリティフロー制御 (PFC) を設定するスクリプトを実行します。

(注)

NVMeoF トラフィックの送信に失敗しないようにするには、サーバを再起動するごとにこのスクリプトを作成して実行する必要があります。

```
# 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
```

ステップ 3 次のコマンドを入力して、NVMe ターゲットを検出します。

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

例えば、50.2.85.200 でターゲットを検出するには、次のようにします。

```
# 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
```

(注)

IPv6 を使用して NVMe ターゲットを検出するには、traddr オプションの次に IPv6 ターゲットアドレスを指定します。

ステップ 4 次のコマンドを入力して、検出された NVMe ターゲットに接続します。

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

例えば、50.2.85.200 のターゲットと上記の subnqn 値を検出するには、次の手順を実行します。

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

(注)

IPv6 を使用して検出した NVMe ターゲットに接続するには、traddr オプションの次に IPv6 ターゲットアドレスを指定します。

ステップ 5 `nvme list` コマンドを使用して、マッピングされたネームスペースを確認します。

```
# 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
```

デバイス マッパー マルチパスの設定

システムがデバイス マッパー マルチパス (DM マルチパス) を使用して構成されている場合は、次の手順に従ってデバイス マッパー マルチパスをセットアップします。

手順

ステップ 1 まだインストールされていない場合は、`device-mapper-multipath` パッケージをインストールします。

ステップ 2 `Multipathd` を有効にして開始します。

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

ステップ 3 etc/multipath.conf ファイルを編集して、次の値を使用します。

```
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
}
```

ステップ 4 更新されたマルチパス デバイス マップを使用してフラッシュします。

```
# multipath -F
```

ステップ 5 マルチパス サービスを再起動します。

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

ステップ 6 マルチパス デバイスを再スキャンします。

```
# multipath -v2
```

ステップ 7 マルチパス ステータスを確認します。

```
# multipath -ll
```

Cisco Intersight を使用した RoCE v2 インターフェイスの削除

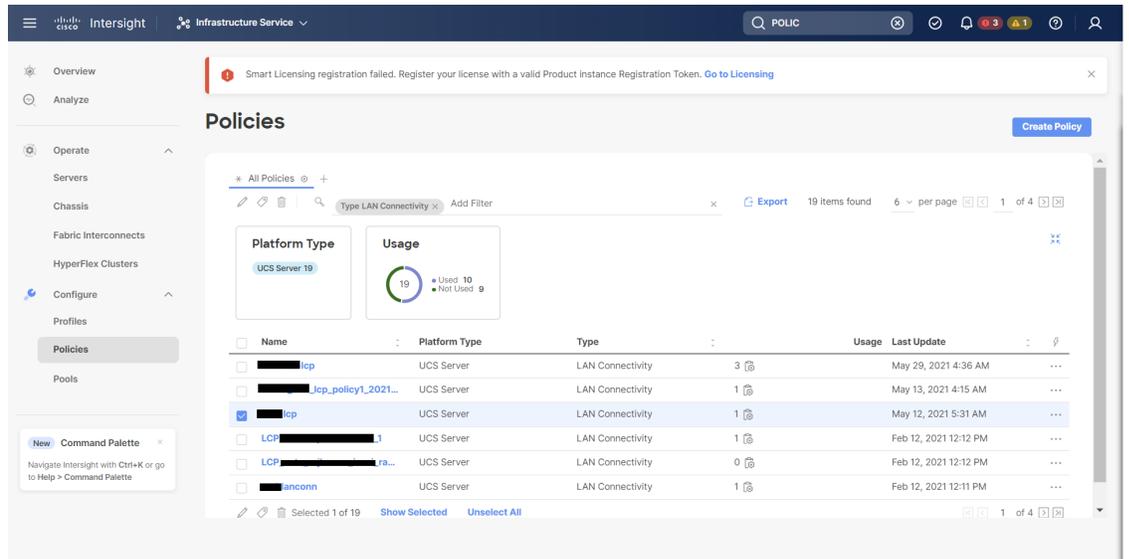
RoCE v2 インターフェイスを削除するには、次の手順を実行します。

手順

ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[フィルタの追加 (Add Filter)] フィールドで、[タイプ: LAN 接続 (Type: LAN Connectivity)] を選択します。

ステップ 2 RoCE V2 構成用に作成された適切な LAN 接続ポリシーを選択し、ポリシー リストの上部または下部にある削除アイコンを使用します。

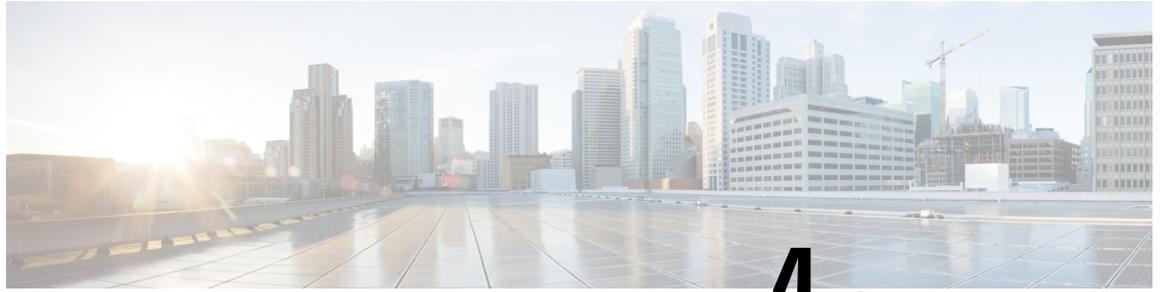
ステップ 3 ポリシーを削除するには、[削除 (Delete)] をクリックします。



The screenshot displays the Cisco Intersight interface for managing policies. A notification at the top indicates a failed Smart Licensing registration. The main section is titled 'Policies' and shows a list of 19 items found, filtered by 'Type LAN Connectivity'. A 'Usage' chart shows 19 items used and 9 not used. The table below lists the policies:

Name	Platform Type	Type	Usage	Last Update
[redacted]_lcp	UCS Server	LAN Connectivity	3	May 29, 2021 4:36 AM
[redacted]_lcp_policy1_2021...	UCS Server	LAN Connectivity	1	May 13, 2021 4:15 AM
[redacted]_lcp	UCS Server	LAN Connectivity	1	May 12, 2021 5:31 AM
[redacted]_lcp	UCS Server	LAN Connectivity	1	Feb 12, 2021 12:12 PM
[redacted]_lcp	UCS Server	LAN Connectivity	0	Feb 12, 2021 12:12 PM
[redacted]_lanconn	UCS Server	LAN Connectivity	1	Feb 12, 2021 12:11 PM

ステップ 4 RoCE v2 構成を削除したら、サーバプロファイルを再展開し、サーバを再起動します。



第 4 章

ESXi での RoCEv2 を使用した NVMeoF の構成

- [ESXi 上で RoCE v2 を使用するファブリック \(NVMeoF\) を介して NVMe を使用する際のガイドライン \(39 ページ\)](#)
- [ESXi の要件 \(40 ページ\)](#)
- [Cisco Intersight での RoCE v2 for NVMeoF の構成 \(40 ページ\)](#)
- [NENIC ドライバのインストール \(45 ページ\)](#)
- [ESXi NVMe RDMA のホスト側の構成 \(46 ページ\)](#)
- [Cisco Intersight を使用した RoCE v2 インターフェイスの削除 \(54 ページ\)](#)

ESXi 上で RoCE v2 を使用するファブリック (NVMeoF) を介して NVMe を使用する際のガイドライン

一般的なガイドラインと制限事項

- Cisco では、[UCS ハードウェアとソフトウェアの互換性](#)をチェックして、NVMeoF のサポートを判断することを推奨します。NVMeoF は、Cisco UCS B シリーズ、C シリーズ、および X シリーズのサーバでサポートされています。
- RoCE v2 を使用した Nonvolatile Memory Express (NVMe) over RDMA は、現在、Cisco VIC 15000 シリーズ アダプタでのみサポートされています。
- RoCEv2 インターフェイスを作成する場合は、シスコが推奨するキューペア、メモリリジョン、リソースグループ、およびサービスクラスの設定を使用してください。キューペア、メモリ領域、リソースグループ、およびサービスクラスの設定が異なると、NVMeoF の機能が保証されない可能性があります。
- RoCE v2 は、アダプタごとに最大 2 つの RoCE v2 対応インターフェイスをサポートします。
- NVMeoF ネームスペースからのブートはサポートされていません。
- レイヤ 3 ルーティングはサポートされていません。

- システムクラッシュ時に `crashdump` を NVMeoF ネームスペースに保存することはサポートされていません。
- NVMeoF は、`usNIC`、`VxLAN`、`VMQ`、`VMMQ`、`NVGRE`、`GENEVE` オフロード、`ENS`、および `DPDK` 機能とともに使用することはできません。
- Cisco Intersight は、RoCE v2 対応の vNIC に対してファブリック フェールオーバーをサポートしません。
- Quality of Service (QoS) no drop クラス構成は、Cisco Nexus 9000 シリーズスイッチなどのアップストリームスイッチで適切に構成する必要があります。QoS の設定は、異なるアップストリームスイッチ間で異なります。
- スパニングツリープロトコル (STP) を使用している場合、フェールオーバーまたはフェールバックイベントが発生したときに、ネットワーク接続が一時的に失われる可能性があります。この接続性の問題が発生しないようにするには、アップリンクスイッチで STP を無効にします。

ダウングレードのガイドライン：最初に RoCEv2 構成を削除してから、Cisco UCS Manager リリース 4.2(3b) バージョンよりも前のリリースバージョンにダウングレードします。

ESXi の要件

ESXi での RoCE v2 の構成と使用には、次のものがが必要です。

- VMware ESXi バージョン 7.0 Update 3。
- Cisco UCS Manager リリース バージョン 4.2(3b) 以降。
- VIC ファームウェア 5.2(3x) 以降のバージョン。
- ドライババージョン、`nenic-2.0.4.0-IOEM.700.1.0.15843807.x86_64.vib` は、Cisco UCS Manager 4.2(3b) リリース パッケージで、標準 eNIC と RDMA の両方のサポートを提供します。
- NVMeoF 接続をサポートするストレージアレイ。

Cisco Intersight での RoCE v2 for NVMeoF の構成

Cisco Intersight で RoCE v2 インターフェイスを構成するには、次の手順に従います。

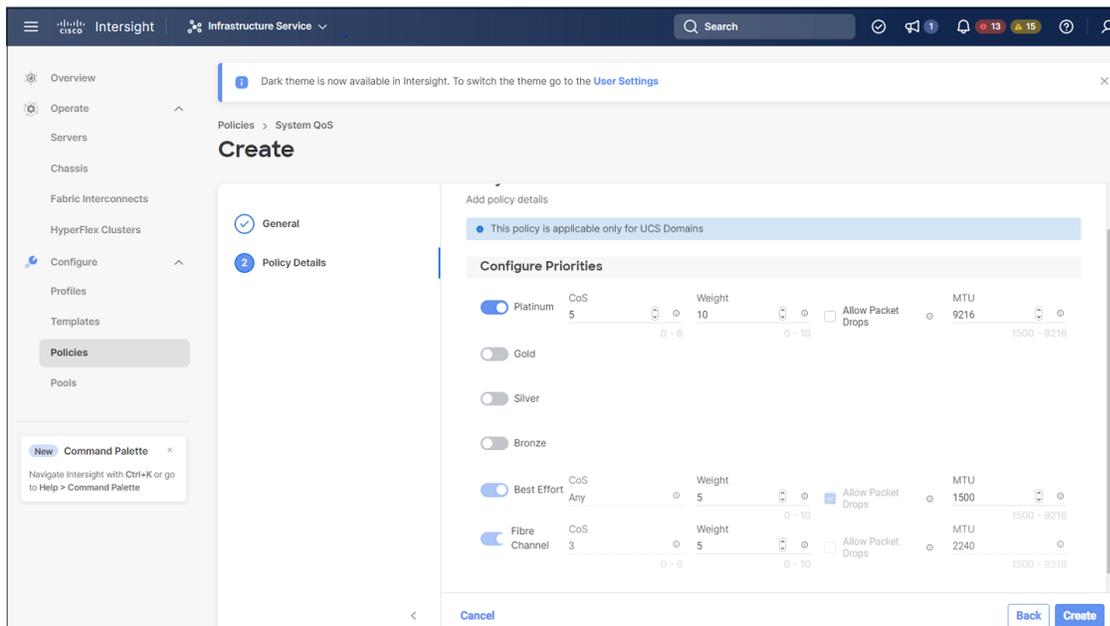
RDMA パケット ドロップの可能性を回避するには、ネットワーク全体で同じ非ドロップ COS が構成されていることを確認してください。次の手順に従えば、システム QoS ポリシーで非ドロップクラスを構成して、RDMA でサポートされているインターフェイス用に使用できます。

手順

ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS ドメイン (UCS Domain)] プラットフォームタイプを選択し、[システム QoS (System QoS)] を検索または選択して、[Start (開始)] をクリックします。

ステップ 2 [全般 (General)] ページでポリシー名を入力し、[次へ (Next)] をクリックします。次に、[ポリシーの詳細 (Policy Details)] ページで、次のようにシステム QoS ポリシーのプロパティ設定を構成します。

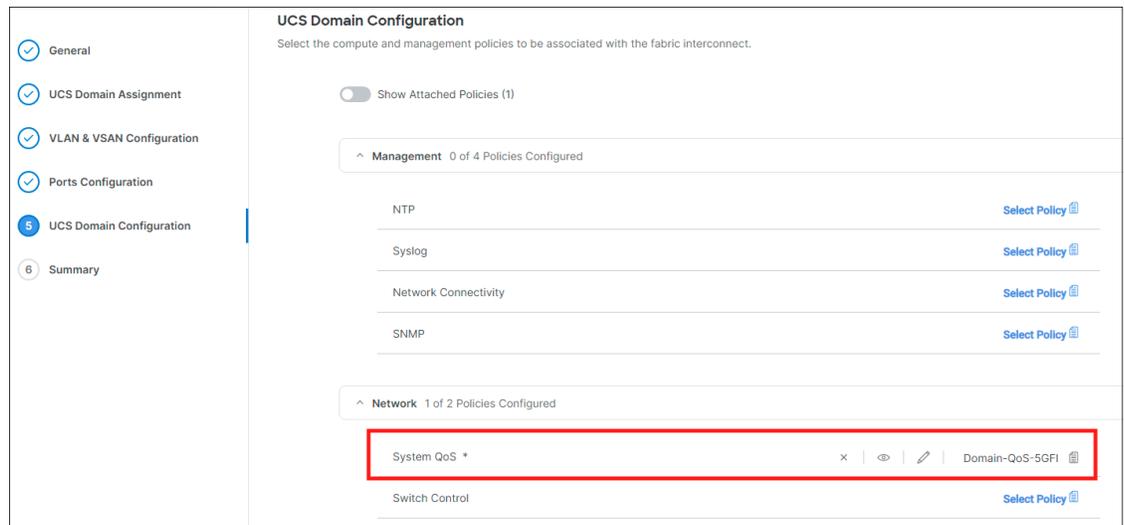
- [優先順位 (Priority)] で、[プラチナ (Platinum)] を選択します。
- [パケットドロップを許可 (Allow Packet Drops)] チェックボックスをオフにします。
- [MTU] については、値を **9216** に設定します。



ステップ 3 [作成 (Create)] をクリックします。

ステップ 4 システム QoS ポリシーをドメインプロファイルに関連付けます。

LAN 接続ポリシーで RoCE 設定を有効化する



(注)
詳細については、「ドメインポリシーの構成」の「システム QoS ポリシーの作成」および「ドメインプロファイルの構成」を参照してください。

システム QoS ポリシーが正常に作成され、ドメインプロファイルに展開されます。

次のタスク

LAN 接続ポリシーで RoCE v2 vNIC 設定を使用してサーバ プロファイルを構成します。

LAN 接続ポリシーで RoCE 設定を有効化する

RoCE v2 vNIC を構成するには、次の手順に従います。Cisco Intersight LAN 接続ポリシーでは、次のように Linux 構成向けのイーサネットアダプタ ポリシーの RoCE 設定を有効にできます。

手順

- ステップ 1 [構成 (CONFIGURE)] > [ポリシー (Policies)] に移動します。[ポリシーの作成 (Create Policy)] をクリックし、[UCS サーバ (UCS Server)] プラットフォーム タイプを選択し、[LAN 接続ポリシー (LAN Connectivity policy)] を検索または選択して、[Start (開始)] をクリックします。
- ステップ 2 ポリシーの [全般 (General)] ページで、ポリシー名を入力し、[ターゲットプラットフォーム (Target Platform)] として [UCS サーバ (スタンドアロン) (UCS Server (Standalone))] または [UCS サーバ (FI アタッチ) (UCS Server (FI-Attached))] を選択し、[次へ (Next)] をクリックします。

ステップ 3 [ポリシーの詳細 (Policy Details)] ページで、[vNIC の追加 (Add vNIC)] をクリックして新しい vNIC を作成します。

ステップ 4 [vNIC の追加 (Add vNIC)] ページで、構成パラメータに従って RoCE v2 vNIC を有効にします。

- a) [全般 (General)] セクションで、仮想イーサネット インターフェイスの名前を入力します。
- b) スタンドアロンサーバの場合は、[Consistent Device Naming (CDN)] をクリックするか、FI アタッチ サーバの [フェールオーバー (Failover)] をクリックして、次の手順を実行します。

- [イーサネットアダプタ (Ethernet Adapter)] の下で、[ポリシーの選択 (Select Policy)] をクリックします。
- [ポリシーの選択 (Select Policy)] ウィンドウで、[新規作成 (Create New)] をクリックして、イーサネットアダプタ ポリシーを作成します。
- イーサネットアダプタ ポリシーの [全般 (General)] ページで、ポリシーの名前を入力し、[次へ (Next)] をクリックします。
- イーサネットアダプタ ポリシーの [ポリシーの詳細 (Policy Details)] ページで、次のプロパティ設定を変更します。

• [RoCE の設定 (RoCE Settings)]

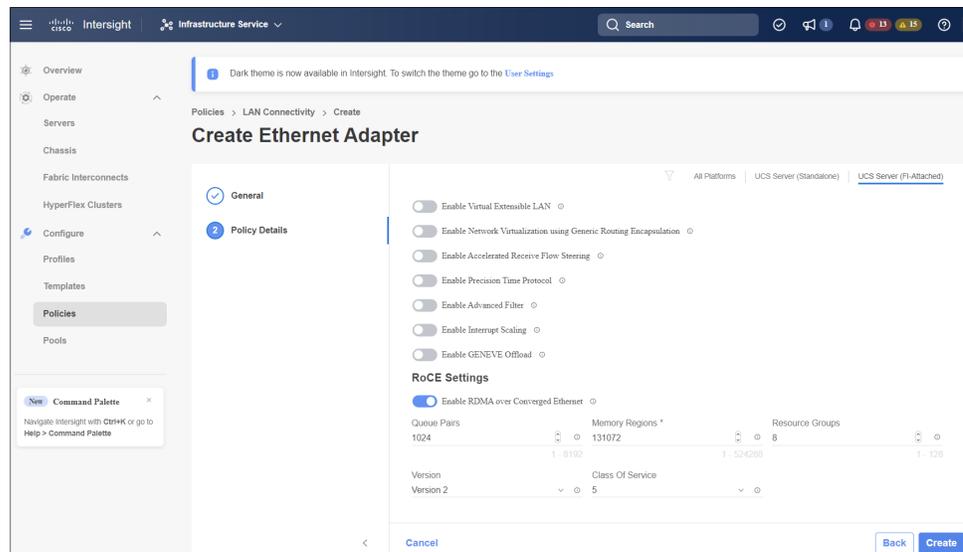
- [コンバージドイーサネット上の RDMA を有効にする (Enable RDMA over Converged Ethernet)] をスライドして有効にし、この仮想インターフェイスの RoCE を設定します。
- [キュー ペア (Queue Pairs)] で、1024 を選択するか、入力します。
- [メモリーリージョン (Memory Regions)] で、131072 を選択するか、入力します。
- [リソースグループ (Resource Groups)] で、8 を選択するか、入力します。
- [バージョン (Version)] で、[バージョン 2 (Version 2)] を選択します。
- [サービスクラス (Class of Service)] で、5 を選択します。

• [割り込み設定 (Interrupt Settings)]

- [割り込み (Interrupts)] で、256 を選択するか、入力します。
- [割り込みモード (Interrupt mode)] で、[MSIX] を選択します。
- [割り込みタイマー (Interrupt Timer)] で、125 を選択します。
- [割り込み調停タイプ (Interrupt Coalescing Type)] で、[最小 (Min)] を選択します。

• [受信 (Receive)] の設定

- [受信キュー数 (Receive Queue Count)] で、**1** を選択するか、入力します。
- [受信リングサイズ (Receiving Ring Size)] で、**512** を選択するか、入力します。
- [送信 (Transmit)] の設定
 - [送信キュー数 (Transmit Queue Count)] で、**1** を選択するか、入力します。
 - [送信リングサイズ (Transmit Ring Size)] で、**256** を選択するか、入力します。
- [完了 (Completion)] の設定
 - [完了キュー カウント (Completion Queue Count)] で、**2** を選択するか、入力します。
 - [完了リングサイズ (Completion Ring Size)] で、**1** を選択するか、入力します。
 - [アップリンク フェールバック タイムアウト (秒) (Uplink Failback Timeout)] で、**5** を選択するか、入力します。
- [作成 (Create)] をクリックして、上記で定義した設定でイーサネットアダプタポリシーを作成します。



- [追加 (Add)] をクリックして設定を保存し、新しい vNIC を追加します。

(注)

* が付いているすべてのフィールドは必須です。適切なポリシーに従って入力または選択されていることを確認してください。

ステップ 5 [作成 (Create)] をクリックし、RoCE v2 設定によって LAN 接続ポリシーを完成させます。

ステップ 6 LAN 接続ポリシーをサーバプロファイルに関連付けます。

(注)

詳細については、「[UCS サーバポリシーの構成](#)」の「[LAN 接続ポリシーの作成](#)」および「[イーサネットアダプタポリシーの作成](#)」および「[UCS サーバプロファイルの構成](#)」を参照してください。

イーサネットアダプタポリシーの vNIC 設定を含む LAN 接続ポリシーが正常に作成および展開され、RoCE v2 設定が有効になります。

次のタスク

RoCE v2 のポリシー構成が完了したら、ホストシステムで NVMeoF の RoCE v2 を構成します。

NENIC ドライバのインストール

始める前に

イーサネット ネットワーク インターフェイス カード (eNIC) のリモートダイレクトメモリアクセス (RDMA) ドライバには、`nenic` ドライバが必要です。

手順

ステップ 1 eNIC vSphere インストールバンドル (VIB) またはオフラインバンドルを ESXi サーバにコピーします。

ステップ 2 次のコマンドを使用して、`nenic` ドライバをインストールします。

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

例:

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

(注)

VIB の署名に使用されている証明書によっては、ホスト許容レベルの変更が必要になる場合があります。これを行うには、次のコマンドを使用します。

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

インストールされている VIB のタイプによっては、ESX をメンテナンス モードにする必要があります。これは、クライアントを介して実行するか、上記の `esxcli` に `--maintenance-mode` オプションを追加することで実行できます。

次のタスク

ESXi NVMe RDMA のホスト側を構成します。

ESXi NVMe RDMA のホスト側の構成

NENIC RDMA の機能

Linux と ESXi の RDMA の主な違いの 1 つを以下に示します。

- ESXi では、物理インターフェイス (vmmnic) の MAC は RoCEv2 トラフィックに使用されません。代わりに、VMkernelポート (vmk) の MACが使用されます。

発信 RoCE パケットはイーサネット送信元 MAC フィールドの vmrk MAC を使用します。着信 RoCE パケットは、イーサネット接続先 mac フィールドの vmk MAC を使用します。vmk MAC アドレスは、作成時に vmk インターフェイスに割り当てられる VMware MAC アドレスです。

- Linux では、物理インターフェイス MAC が ROCE パケットの送信元 MAC アドレスフィールドで使用されます。この Linux MAC は通常、UCS Manager を使用して VNIC に構成された Cisco MAC アドレスです。

ホストに ssh で接続し、`esxcli network ip interface list` コマンドを使用すると、MAC アドレスを確認できます。

```
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
```

ホスト、仮想マシンにネットワーク接続を提供し、VMkernel トラフィックを処理するには、vSphere 標準スイッチを作成する必要があります。作成する接続タイプに応じて、VMkernel アダプタを使用して新しい vSphere 標準スイッチを作成するか、物理ネットワークアダプタのみを新しいスイッチに接続するか、または仮想マシンポートグループを使用してスイッチを作成することができます。

ネットワーク接続スイッチの作成

次の手順に従って、ホスト、仮想マシンにネットワーク接続を提供し、VMkernel トラフィックを処理するための vSphere 標準スイッチを作成します。

始める前に

enic ドライバがあることを確認します。以下の手順に進む前に、enic ドライバをダウンロードしてインストールします。

手順

- ステップ 1 vSphere Web Client で、ホストに移動します。
- ステップ 2 [構成 (Configure)] タブで、[ネットワークング (Networking)] を展開し、[仮想スイッチ (Virtual Switches)] を選択します。
- ステップ 3 [ネットワークングの追加 (Add Networking)] をクリックします。
使用可能なネットワークアダプタの接続タイプは次のとおりです。
 - Vmkernel ネットワークアダプタ

ホスト管理トラフィックを処理する新しい VMkernel アダプタを作成します。

• **物理ネットワーク アダプタ**

物理ネットワーク アダプタを新しい、または既存の標準スイッチに追加します。

• **標準スイッチの仮想マシン ポート グループ**

仮想マシン ネットワーキング用の新しいポート グループを作成します。

ステップ 4 接続タイプ [VMkernel ネットワーク アダプタ (VMkernel Network Adapter)] を選択します。

ステップ 5 [新しい標準スイッチ (New standard switch)] を選択し、[次へ (Next)] をクリックします。

ステップ 6 物理ネットワーク アダプタを新しい標準スイッチに追加します。

- a) [割り当て済みアダプタ (Assigned Adapters)] で、[新しいアダプタ (New Adapters)] を選択します。
- b) リストから 1 つ以上のアダプタをセレクトし、[OK] をクリックします。スループットの向上を促し、冗長性を得るため、[アクティブ (Active)] リストで少なくとも 2 つの物理ネットワーク アダプタを追加してください。
- c) (オプション) 上矢印キーと下矢印キーで、[割り当て済みアダプタ (Assigned Adapters)] リスト内のアダプタの位置を変更します。
- d) [次へ (Next)] をクリックします。

ステップ 7 VMadapter またはポートグループ用に作成した新しい標準スイッチに対し、アダプタまたはポートグループの接続設定を入力します。

- a) VMkernel アダプタのトラフィック タイプを表すラベルを入力します。
- b) ネットワーク トラフィックのルーティングで VMkernel が使用する VLAN を識別するための、VLAN ID を設定します。
- c) [IPv4]、[IPv6]、またはその両方を選択します。
- d) ドロップダウンメニューから MTU サイズを選択します。特定の MTU サイズを入力する場合は、[カスタム (Custom)] を選択します。最大 MTU サイズは 9000 バイトです。

(注)

1500 より大きい MTU を設定すれば、ジャンボ フレームを有効にすることができます。

- e) VMkernel アダプタの TCP/IP スタックを設定した後、TCP/IP スタックを選択します。デフォルトの TCP/IP スタックを使用するには、使用可能なサービスから選択します。

(注)

VMkernel アダプタの TCP/IP スタックは、後から変更できないことに注意してください。

- f) IPV4 または IPV6 設定、あるいはその両方を構成します。

ステップ 8 [完了の準備] ページで、[完了 (Finish)] をクリックします。

ステップ 9 次の結果に示すように、vSphere クライアントで NVMe RDMA を使用して VM アダプタまたはポートグループの VMkernel ポートを確認します。

NVMe RDMA を使用する VM アダプタまたはポートグループの VMkernel ポートは、以下のようになります。

The screenshot shows the ESXi configuration page for VMkernel adapters. The left sidebar has 'Networking' expanded and 'VMkernel adapters' selected. The main area shows a table of VMkernel adapters with columns for Device, Network Label, Switch, IP Address, TCP/IP Stack, and Enabled Services.

Device	Network Label	Switch	IP Address	TCP/IP Stack	Enabled Services
vmk0	Management Network	vSwitch0	10.193.176.52	Default	Management
vmk1	vmk284	vSwitch1	50.284.210	Default	--
vmk2	vmk283	vSwitch2	50.2.83.210	Default	--

NVMeRDMA がサポートされている vmnic で作成された VRDMA ポートグループは、次のように表示されます。

The screenshot shows the ESXi configuration page for RDMA adapters. The left sidebar has 'Networking' expanded and 'RDMA adapters' selected. The main area shows a table of RDMA adapters with columns for Name, Driver, State, Paired UpLink, RoCE v1, RoCE v2, and IWARP.

Name	Driver	State	Paired UpLink	RoCE v1	RoCE v2	IWARP
vmrdma0	nenic	Active	vmnic2	Disabled	Enabled	Disabled
vmrdma1	nenic	Active	vmnic3	Disabled	Enabled	Disabled

Below the table, it shows 'RDMA Device: vmrdma1' and a table for 'Bound VMkernel Adapters':

VMkernel Adapter	TCP/IP Stack	IP Address
vmk2	Default	50.2.83.210

次のタスク

vmrdma ポートの上に vmhba ポートを作成します。

ESXi での VMVHBA ポートの作成

vmrdma アダプタ ポートの上に vmhba ポートを作成するには、次の手順に従います。

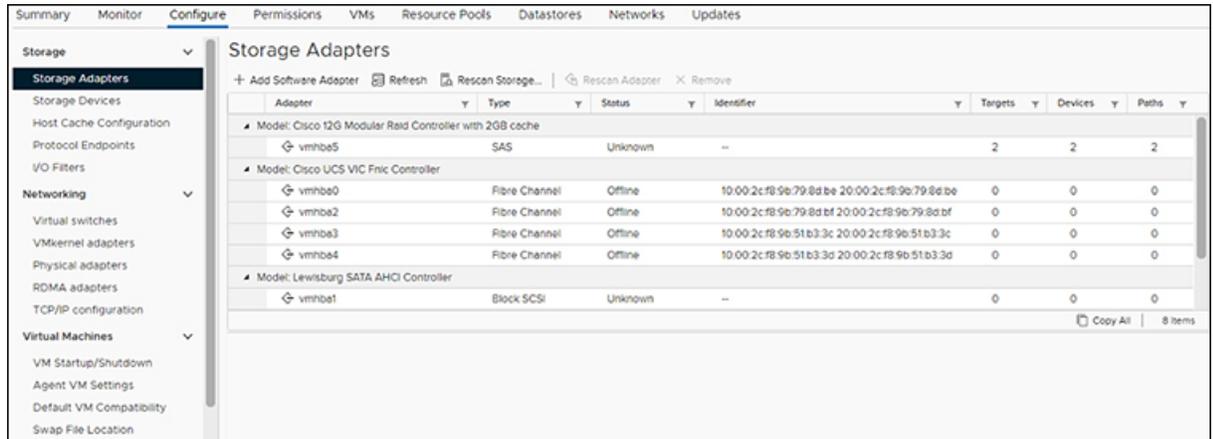
始める前に

ストレージ接続用のアダプタ ポートを作成します。

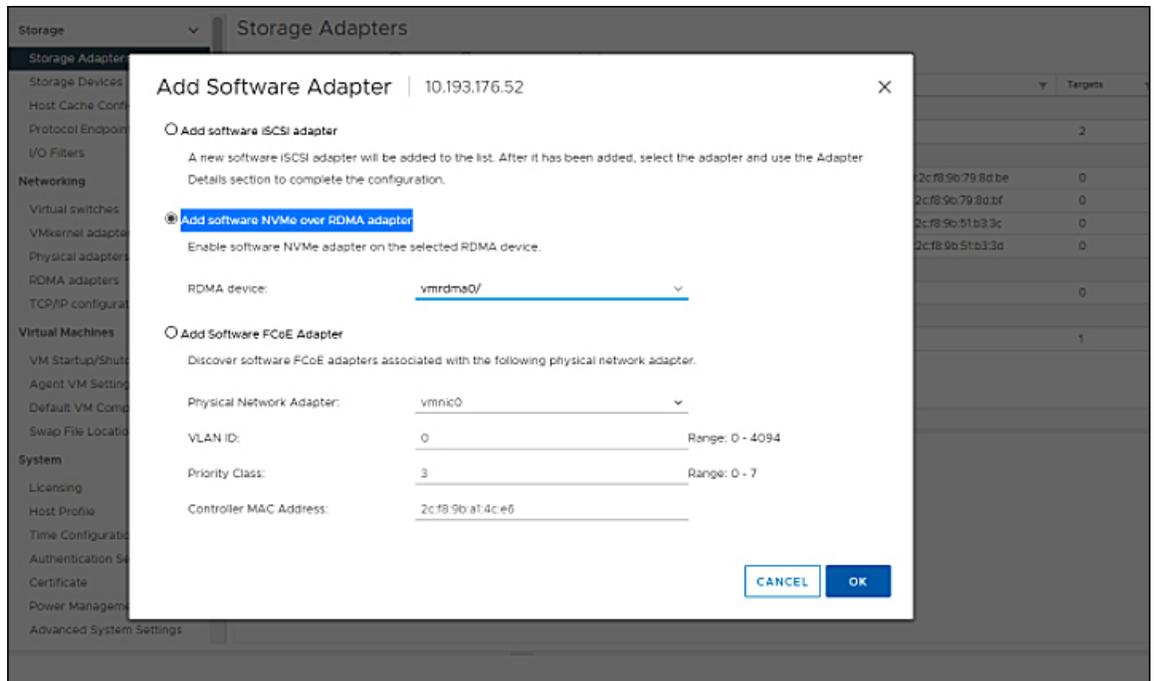
手順

ステップ 1 ESXi ホストが接続されている vCenter に移動します。

ステップ 2 [ホスト (Host)]>[構成 (Configure)]>[ストレージアダプタ (Storage adapters)]の順にクリックします。



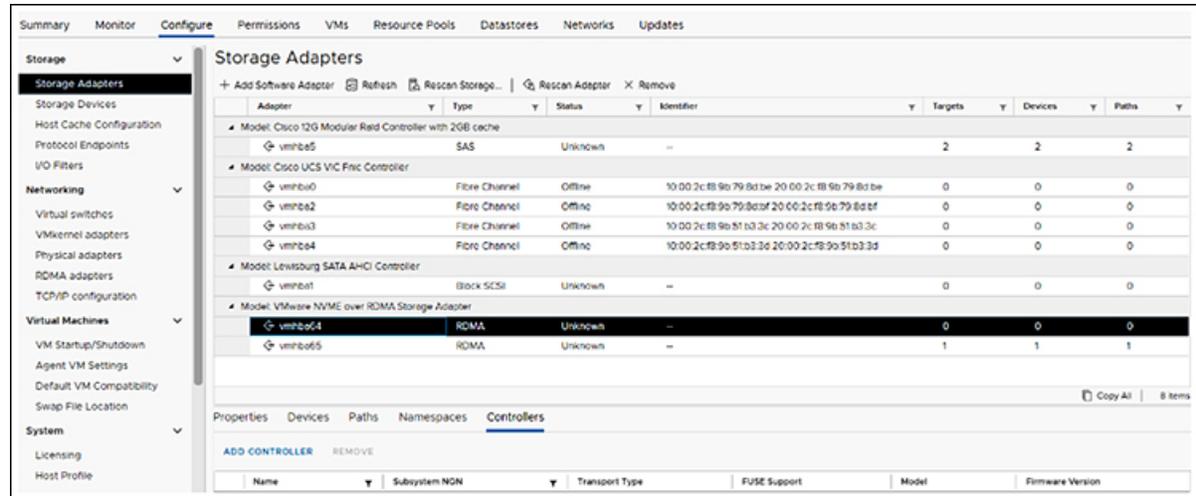
ステップ 3 [+ ソフトウェア アダプタの追加 (Add Software Adapter)]をクリックします。次のダイアログボックスが表示されます。



ステップ 4 [RDMA アダプタ上にソフトウェア NVMe を追加 (Add software NVMe over RDMA adapter)]と、使用する vmrdma ポートを選択します。

ステップ5 [OK] をクリックします。

RDMA ストレージアダプタ上の VMware NVMe の vmhba ポートは、次の例のように表示されます。



vmnic および vmrDMA インターフェイスの表示

ESXi は、ホストに構成された各 nenic VNIC に対して vmnic インターフェイスを作成します。

始める前に

ネットワーク アダプタと VHBA ポートを作成します。

手順

ステップ1 ssh を使用してホスト システムにアクセスします。

ステップ2 esxcfg-nics -l と入力して、ESXi 上の vmnic を一覧表示します。

```

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

```

ステップ 3 `esxcli rdma device list` を使用して、`vmrdma` デバイスを一覧表示します。enic ドライバが RDMA 対応 VNIC の RDMA デバイスを ESXi に登録すると、ESXi は `vmrdma` デバイスを作成し、対応する `vmnic` にリンクします。

```
[root@ESXi7U3 ~]# 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@ESXi7U3 ~]# esxcli rdma device vmknics list
-----
Device  Vmknics  NetStack
-----
vmrdma0  vmk1     defaultTcpipStack
vmrdma1  vmk2     defaultTcpipStack
```

ステップ 4 `esxcli rdma device protocol list` を使用して、`vmrdma` インターフェイスでサポートされているプロトコルを確認します。

enic の場合、RoCE v2 がサポートされている唯一のプロトコルであることがリストから分かります。このコマンドの出力は、VNIC の RoCEv2 設定と一致しているはずですが。

```
[root@ESXi7U3 ~]# esxcli rdma device protocol list
-----
Device  RoCE v1  RoCE v2  iWARP
-----
vmrdma0  false    true     false
vmrdma1  false    true     false
[root@ESXi7U3 ~]#
```

ステップ 5 `esxcli nvme adapter list` を使用して、NVMe アダプタと、それが構成されている `vmrdma` および `vmnic` インターフェイスを一覧表示します。

```
[root@ESXi7U3 ~]# esxcli nvme adapter list
-----
Adapter  Adapter Qualified Name  Transport Type  Driver  Associated Devices
-----
vmhba64  aqn:vmrdma:2c-f8-9b-79-8d-bc  RDMA           nvmerdma  vmrdma0, vmnic2
vmhba65  aqn:vmrdma:2c-f8-9b-79-8d-bd  RDMA           nvmerdma  vmrdma1, vmnic3
[root@ESXi7U3 ~]#
```

ステップ 6 `esxcli storage core adapter list` を使用して、システム内のすべての `vmhbas` を一覧表示できます。RDMA を介して構成された `vmhba`。

```
[root@ESXi7U3 ~]# esxcli storage core adapter list
-----
HBA Name  Driver  Link State  UID  Capabilities  Description
-----
vmhba0    nfnic   link-down   fc.10002cf89b798dbf:20002cf89b798dbf  Second Level Lun ID  (0000:1d:00:2) Cisco Corporation Cisco UCS VIC Fnic Controller
vmhba1    vms_ahci  link-n/a    sata.vmhba1  Second Level Lun ID  (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.10002cf89b51b33d:20002cf89b51b33d  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    lsir3   link-n/a    sas.5cc167e9732f9b00  Second Level Lun ID  (0000:3c:00:0) Broadcom Cisco 12G Modular Raid Controller with 2GB cache
vmhba64  nvmerdma  link-n/a    rdma.vmk12:2c:f8:9b:79:8d:bc  VMware NVMe over RDMA Storage Adapter on vmrdma0
vmhba65  nvmerdma  link-n/a    rdma.vmk13:2c:f8:9b:79:8d:bd  VMware NVMe over RDMA Storage Adapter on vmrdma1
[root@ESXi7U3 ~]#
```

(注)

vmhba64 および vmhba65 の場合、ドライバのリンク状態に *Online* ではなく *link-n/a* と表示されることがあります。これは、ESXi 7.0 Update 3 の既知の問題です。詳細については、[ESXi](#) を参照してください。

NVMe ファブリックと名前空間の検出

この手順は、ESXi コマンドライン インターフェイスを使用して実行します。

始める前に

アダプタの VMHBA で NVMe を作成して構成します。アダプタの最大数は 2 で、ベストプラクティスは両方をフォールトトレランス用に構成することです。

手順

ステップ 1 vmrdma デバイスの NVMe をチェックして有効にします。

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

NVMe が有効になっているかどうかを示すメッセージが表示されます。

ステップ 2 次のコマンドを入力して、アレイ上の NVMe ファブリックを検出します。

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

esxcli nvme fabrics discover -a vmhba64 -l 50.2.84.100 の実行結果を示す図
出力では、トランスポートタイプ、アドレスファミリー、サブシステムタイプ、コントローラ ID、管理キュー、最大サイズ、トランスポートアドレス、トランスポートサービス ID、およびサブシステム NQN の情報が表示されます。

NVMe コントローラに出力が表示されます。

ステップ 3 NVMe ファブリック インターコネクトを実行します。

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

ステップ 4 手順 1 ~ 4 を繰り返して、2 番目のアダプタを構成します。

ステップ 5 設定を確認します。

- a) コントローラリストを表示して、NVMe コントローラが存在し、動作していることを確認します。

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

```
[root@ESXi7U3:~] esxcli nvme controller list
Name                               Controller Number  Adapter  Transport Type  Is Online
-----
nqn.2010-06.com.purestorage:flasharray.5ab274df5b161455#vmhba64#50.2.84.100:4420  258  vmhba64  RDMA             true
nqn.2010-06.com.purestorage:flasharray.5ab274df5b161455#vmhba65#50.2.83.100:4420  259  vmhba65  RDMA             true
[root@ESXi7U3:~] esxcli nvme namespace list
Name                               Controller Number  Namespace ID  Block Size  Capacity in MB
-----
eui.00e6d65b65a8f34024a9374e00011745  258  71493  512  102400
eui.00e6d65b65a8f34024a9374e00011745  259  71493  512  102400
[root@ESXi7U3:~]
```

- b) ファブリックがアダプタを介してコントローラで有効になっていることを確認し、コントローラがアダプタのポートを介してアクセス可能であることを確認します。

```
[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
```

Cisco Intersight を使用した RoCE v2 インターフェイスの削除

RoCE v2 インターフェイスを削除するには、次の手順を実行します。

手順

- ステップ 1 **[構成 (CONFIGURE)] > [ポリシー (Policies)]** に移動します。 **[フィルタの追加 (Add Filter)]** フィールドで、 **[タイプ: LAN 接続 (Type: LAN Connectivity)]** を選択します。
- ステップ 2 RoCE V2 構成用に作成された適切な LAN 接続ポリシーを選択し、ポリシー リストの上部または下部にある削除アイコンを使用します。
- ステップ 3 ポリシーを削除するには、 **[削除 (Delete)]** をクリックします。

The screenshot shows the Cisco Intersight interface with the 'Policies' section active. A filter 'Type LAN Connectivity' is applied. A summary card shows 'Usage' with 19 items, 10 used, and 9 not used. Below is a table of policies:

Name	Platform Type	Type	Usage	Last Update
[redacted] lcp	UCS Server	LAN Connectivity	3	May 29, 2021 4:36 AM
[redacted] lcp_policy1_2021...	UCS Server	LAN Connectivity	1	May 13, 2021 4:15 AM
[redacted] lcp	UCS Server	LAN Connectivity	1	May 12, 2021 5:31 AM
[redacted] LCP [redacted] _1	UCS Server	LAN Connectivity	0	Feb 12, 2021 12:12 PM
[redacted] LCP [redacted] ra...	UCS Server	LAN Connectivity	0	Feb 12, 2021 12:12 PM
[redacted] anconn	UCS Server	LAN Connectivity	1	Feb 12, 2021 12:11 PM

- ステップ 4 RoCE v2 構成を削除したら、サーバプロファイルを再展開し、サーバを再起動します。



第 5 章

既知の問題

- [Windows](#) (55 ページ)
- [Linux](#) (56 ページ)
- [ESXi](#) (56 ページ)

Windows

症状	条件 (Conditions)	回避策
VIC 1400 シリーズアダプタでは、Windows 2019 用の neNIC ドライバを Windows 2016 にインストールすること、そして Windows 2016 のドライバを Windows 2019 にインストールすることができます。ただし、これはサポートされていない構成です。	ケース 1 : Windows 2019 neNIC ドライバの Windows 2016 へのインストールは成功しますが、Windows 2016 では RDMA はサポートされません。 ケース 2 : Windows 2016 neNIC ドライバの Windows 2019 へのインストールは成功しますが、Windows 2019 では RDMA が有効状態ではなく、デフォルトの無効状態になります。	Windows 2016 および Windows 2019 のドライババイナリは、それに対応した名前のフォルダにあります。ビルド/アップグレードするプラットフォームに正しいバイナリをインストールしてください。

Linux

症状	条件 (Conditions)	回避策
<p>一部の Cisco Nexus 9000 スイッチで高帯域幅の NVMe トラフィックを送信すると、ストレージに接続されたスイッチポートが最大 PFC ピークに達し、バッファが自動的にクリアされないことがあります。Nexus 9000 スイッチでは、nxos コマンド「show hardware internal buffer info pkt-stats input peak」により、ポートの Peak_cell または PeakQos 値が 1000 を超えたかどうかが表示されます。</p>	<p>NVMe トラフィックはドロップされます。</p>	<p>このエラー モードからスイッチを回復します。</p> <ol style="list-style-type: none"> 1. スイッチにログインします。 2. ストレージに接続されているポートを特定し、「shutdown」コマンドを使用してポートをシャットダウンします。 3. 以下のコマンドを順に実行します。 <pre># clear counters # clear counter buffers module 1 # clear qos statistics</pre> 4. シャットダウンしたポートで no shutdown を実行します。

ESXi

症状	条件 (Conditions)	回避策
<p>esxcli storage core adapter list コマンドを使用して vmhba を一覧表示すると、vmhba64 および vmhba65 rdma ポートのドライバのリンク状態に、<i>[Online]</i> ではなく <i>[Link-n/a]</i> と表示されます。</p> <p>(注) VMware Developer Center Partner Network (DCPN) ケース ID - 00113157</p>	<p>これは、ESXi 7.0 Update 3 の既知の問題です。</p>	<p>なし</p>

翻訳について

このドキュメントは、米国シスコ発行ドキュメントの参考和訳です。リンク情報につきましては、日本語版掲載時点で、英語版にアップデートがあり、リンク先のページが移動/変更されている場合がありますことをご了承ください。あくまでも参考和訳となりますので、正式な内容については米国サイトのドキュメントを参照ください。