



# RDMA経由のNFSを管理します。

## ONTAP 9

NetApp  
December 20, 2024

# 目次

RDMA経由のNFSを管理します。 . . . . .	1
NFS over RDMAの概要 . . . . .	1
NFS over RDMA用のNICの設定 . . . . .	2
RDMA経由のNFS用のLIFの設定 . . . . .	5
NFS設定を変更する . . . . .	7

# RDMA経由のNFSを管理します。

## NFS over RDMAの概要

NFS over RDMAはRDMAアダプタを使用するため、ストレージシステムメモリとホストシステムメモリの間でデータを直接コピーできるため、CPUの中断やオーバーヘッドを回避できます。

NFS over RDMA構成は、機械学習や分析など、レイテンシの影響を受けやすいワークロードや広帯域ワークロードを抱えているお客様向けに設計されています。NVIDIAでは、RDMA経由でNFSを拡張してGPUダイレクトストレージ (GDS) を実現しました。GDSは、CPUとメインメモリをすべてバイパスし、RDMAを使用してストレージシステムとGPUメモリの間でデータを直接転送することで、GPU対応のワークロードをさらに高速化します。

RDMA 10.1以降では、over RDMA構成がMellanox CX-5またはCX-6アダプタで使用されている場合にNFSv4.0プロトコルでサポートされます。このアダプタを使用すると、バージョン2のONTAP 9プロトコルを使用するRDMAがサポートされます。GDSは、Mellanox NICカードとMOFEDソフトウェアを搭載したNVIDIA TeslaおよびAmpereファミリのGPUでのみサポートされます。以降のONTAPリリースでサポートされるNFSバージョンについては、要件の表を参照してください。



NFSマウントサイズが64kを超えると、NFS over RDMA構成のパフォーマンスが不安定になります。

### 要件

- ストレージシステムでONTAP 9 10.1以降が実行されている必要があります。
- 使用するNFSのバージョンに対応した正しいバージョンのONTAPを実行していることを確認します。

NFSバージョン	ONTAPのサポート
NFSv4.0	ONTAP 9.10.1以降
NFSv4.1	ONTAP 9.14.1以降
NFSv3	ONTAP 9.15.1以降

◦ ONTAP 9.12.1以降では、System Managerを使用してNFS over RDMAを設定できます。ONTAP 9.10.1および9.11.1では、CLIを使用してNFS over RDMAを設定する必要があります。

- HAペアの両方のノードでバージョンが同じである必要があります。
- ストレージシステムコントローラでRDMAがサポートされている必要があります。

ONTAPのバージョン	RDMAをサポートするコントローラ
9.10.1以降	<ul style="list-style-type: none"><li>• AFF A400</li><li>• AFF A700用</li><li>• AFF A800用</li></ul>

ONTAPのバージョン	RDMAをサポートするコントローラ
ONTAP 9.14.1以降	<ul style="list-style-type: none"> <li>• AFF Cシリーズ</li> <li>• AFF A900用</li> </ul>
ONTAP 9.15.1以降	<ul style="list-style-type: none"> <li>• AFF A1K用</li> <li>• AFF A90用</li> <li>• AFF A70用</li> </ul>
ONTAP 9.16.1以降	<ul style="list-style-type: none"> <li>• AFF A50用</li> <li>• AFF A30用</li> <li>• AFF A20用</li> </ul>

- データLIFは、RDMAをサポートするように設定する必要があります。
- クライアントでMellanox RDMA対応NICカードとMellanox OFED（MOFED）ネットワークソフトウェアを使用している必要があります。アダプタのサポートについては、を参照して"[NetApp Hardware Universe](#)"ください。NFS over RDMAでサポートされるアダプタの説明フィールドに「RoCE」と表示されます。



インターフェイスグループは、RDMA経由のNFSではサポートされません。

#### 次のステップ

- [NFS over RDMA用のNICの設定](#)
- [RDMA経由のNFS用のLIFの設定](#)
- [NFS over RDMAのNFS設定](#)

#### 関連情報

- ["RDMA"](#)
- [NFSトランキングの概要](#)
- ["RFC 7530 : NFS バージョン 4 プロトコル"](#)
- ["RFC 8166 : リモート手順コールバージョン 1 用のリモートダイレクトメモリアクセストランスポート"](#)
- ["RFC 8167 : RPC-over-RDMA トランスポート上の双方向リモート手順コール"](#)
- ["RFC 8267 : RPC-over-RDMA バージョン 1 への NFS 上位レイヤバインディング"](#)

## NFS over RDMA用のNICの設定

RDMA経由のNFSでは、クライアントシステムとストレージプラットフォームの両方でNIC設定を行う必要があります。

## ストレージプラットフォームの構成

X1148 RDMAアダプタがサーバにインストールされている必要があります。HA構成を使用している場合は、フェイルオーバー中もRDMAサービスを継続できるように、フェイルオーバーパートナーに対応するX1148アダプタが必要です。NICはROCEに対応している必要があります。

RDMA.10.1以降では、次のコマンドを使用してONTAP 9オフロードプロトコルのリストを表示できます。

```
network port show -rdma-protocols roce
```

## クライアントシステム構成

クライアントでMellanox RDMA対応NICカード（X1148など）とMellanox OFEDネットワークソフトウェアを使用している必要があります。サポートされているモデルとバージョンについては、Mellanoxのドキュメントを参照してください。クライアントとサーバは直接接続できますが、スイッチのフェイルオーバーパフォーマンスが向上するため、スイッチの使用を推奨します。

クライアント、サーバ、スイッチ、およびスイッチ上のすべてのポートは、ジャンボフレームを使用して設定する必要があります。また、優先度フロー制御がすべてのスイッチで有効になっていることを確認します。

この構成を確認したら、NFSをマウントできます。

## System Manager

Managerを使用してoverのネットワークインターフェイスを設定するには、ONTAP 9 12.1以降を使用している必要があります。

### 手順

1. RDMAがサポートされているかどうかを確認します。[Network]>[Ethernet Ports]に移動し、グループビューで適切なノードを選択します。ノードを展開するときに、特定のポートの\* rdma protocols フィールドを確認します。RoCEは**RDMA**がサポートされていることを示し、ダッシュ (-\*) はサポートされていないことを示します。
2. VLANを追加するには、**+VLAN\***を選択します。適切なノードを選択します。[ポート]\*ドロップダウンメニューで、使用可能なポートに「RoCE Enabled \*」というテキストが表示されます (RDMAがサポートされている場合)。RDMAがサポートされていない場合は、テキストは表示されません。
3. 新しいNFSサーバを設定するには、のワークフローに従い**NFSを使用したLinuxサーバ用のNASストレージの有効化**ます。

ネットワークインターフェイスを追加する際には、「\* RoCEポートを使用\*」を選択できません。RDMA経由のNFSを使用するすべてのネットワークインターフェイスに対して、このオプションを選択します。

### CLI

1. コマンドを使用して、NFSサーバでRDMAアクセスが有効になっているかどうかを確認します。

```
vserver nfs show -vserver SVM_name
```

デフォルトでは、`-rdma`が有効になっている必要があります。有効になっていない場合は、NFSサーバでRDMAアクセスを有効にします。

```
vserver nfs modify -vserver SVM_name -rdma enabled
```

2. RDMA経由でNFSv4.0を使用してクライアントをマウントします。
  - a. protoパラメータの入力は、サーバのIPプロトコルのバージョンによって異なります。IPv4の場合は、を使用し`proto=rdma`ます。IPv6の場合は、を使用し`proto=rdma6`ます。
  - b. 標準ポート2049ではなく、NFSターゲットポートをとして指定し`port=20049`ます。

```
mount -o vers=4,minorversion=0,proto=rdma,port=20049 Server_IP_address  
:/volume_path mount_point
```

3. オプション：クライアントをアンマウントする必要がある場合は、次のコマンドを実行します。

```
unmount mount_path
```

### 詳細情報

- [NFSサーバを作成する](#)
- [NFSを使用したLinuxサーバ用のNASストレージの有効化](#)

# RDMA経由のNFS用のLIFの設定

RDMA経由のNFSを利用するには、LIF（ネットワークインターフェイス）をRDMAと互換性があるように設定する必要があります。LIFとそのフェイルオーバーペアの両方でRDMAがサポートされている必要があります。

## 新しいLIFを作成する

### System Manager

Managerで経由の用のネットワークインターフェイスを作成するには、ONTAP 9.12.1以降を実行している必要があります。

#### 手順

1. Network > Overview > Network Interfaces \*を選択します。
2. を選択します **+ Add**。
3. NFS、SMB / CIFS、S3 を選択すると、[RoCEポートを使用]\*オプションが表示されます。「RoCEポートを使用する」のチェックボックスをオンにします。
4. Storage VMとホームノードを選択します。名前、IPアドレス、およびサブネットマスクを割り当てます。
5. IPアドレスとサブネットマスクを入力すると、System ManagerはブロードキャストドメインのリストをRoCE対応ポートを備えたドメインにフィルタリングします。ブロードキャストドメインを選択してください。必要に応じてゲートウェイを追加できます。
6. [保存 ( Save ) ]を選択します。

### CLI

#### 手順

1. LIFを作成します。

```
network interface create -vserver SVM_name -lif lif_name -service-policy service_policy_name -home-node node_name -home-port port_name {-address IP_address -netmask netmask_value | -subnet-name subnet_name} -firewall -policy policy_name -auto-revert {true|false} -rdma-protocols roce
```


- サービスポリシーは、default-data-files、またはdata-NFSネットワークインターフェイスサービスを含むカスタムポリシーのいずれかである必要があります。
- `-rdma-protocols``パラメータにはリストを指定できます。このリストはデフォルトでは空です。``roce``を値として追加すると、LIFはRoCEオフロードをサポートしているポートにのみ設定でき、Bot LIFの移行とフェイルオーバーに影響します。

## LIFを変更する

## System Manager

Managerで経由の用のネットワークインターフェイスを作成するには、ONTAP 9.12.1以降を実行している必要があります。

### 手順

1. Network > Overview > Network Interfaces \*を選択します。
2. 変更するネットワークインターフェイスの横にある\*>[編集]\*を選択します .
3. RoCEポートを使用する\*をオンにしてNFS over RDMAを有効にするか、オフにして無効にしてください。ネットワークインターフェイスがRoCE対応ポート上にある場合は、「RoCEポートを使用する」の横にチェックボックスが表示されます。
4. 必要に応じて他の設定を変更します。
5. 「\* Save \* (保存)」を選択して、変更を確定します。

### CLI

1. LIFのステータスは、コマンドを使用して確認できます `network interface show`。サービスポリシーにはdata-NFSネットワークインターフェイスサービスを含める必要があります。 `-rdma -protocols`` リストにはを含める必要があります ``roce`。上記のいずれかの条件が満たされていない場合は、LIFを変更します。
2. LIFを変更するには、次のコマンドを実行します。

```
network interface modify vserver SVM_name -lif lif_name -service-policy service_policy_name -home-node node_name -home-port port_name {-address IP_address -netmask netmask_value | -subnet-name subnet_name} -firewall-policy policy_name -auto-revert {true|false} -rdma-protocols roce
```



特定のオフロードプロトコルを必要とするようにLIFを変更するときに、そのプロトコルをサポートするポートにLIFが現在割り当てられていないとエラーが発生します。

## LIFを移行する

ONTAPでは、RDMA経由のNFSを利用するためにネットワークインターフェイス (LIF) を移行することもできます。この移行を実行する場合は、デスティネーションポートがRoCEに対応していることを確認する必要があります。ONTAP 9.12.1以降では、この手順をSystem Managerで実行できます。System Managerは、ネットワークインターフェイスのデスティネーションポートを選択する際に、RoCEに対応しているかどうかを指定します。

LIFをRDMA経由のNFS構成に移行できるのは、次の場合のみです。

- RoCE対応ポートでホストされるNFS RDMAネットワークインターフェイス (LIF) です。
- RoCE対応ポートでホストされるNFS TCPネットワークインターフェイス (LIF) です。
- RoCE非対応のポートでホストされるNFS TCPネットワークインターフェイス (LIF) です。

ネットワークインターフェイスの移行の詳細については、[を参照してください](#) [LIFを移行する](#)。

### 詳細情報



- [LIFの作成](#)
- [LIFの作成](#)
- [LIFを変更する](#)
- [LIFを移行する](#)

## NFS設定を変更する

ほとんどの場合、RDMA経由のNFS用のNFS対応Storage VMの設定を変更する必要はありません。

ただし、MellanoxチップとLIFの移行に関連する問題に対処している場合は、NFSv4ロック猶予期間を延長する必要があります。デフォルトでは、猶予期間は45秒に設定されています。ONTAP 9.10.1以降では、猶予期間の最大値は180（秒）です。

手順

1. 権限レベルをadvancedに設定します。

```
set -privilege advanced
```

2. 次のコマンドを入力します。

```
vserver nfs modify -vserver SVM_name -v4-grace-seconds number_of_seconds
```

このタスクの詳細については、[を参照してくださいNFSv4ロック猶予期間の指定。](#)

## 著作権に関する情報

Copyright © 2024 NetApp, Inc. All Rights Reserved. Printed in the U.S.このドキュメントは著作権によって保護されています。著作権所有者の書面による事前承諾がある場合を除き、画像媒体、電子媒体、および写真複写、記録媒体、テープ媒体、電子検索システムへの組み込みを含む機械媒体など、いかなる形式および方法による複製も禁止します。

ネットアップの著作物から派生したソフトウェアは、次に示す使用許諾条項および免責条項の対象となります。

このソフトウェアは、ネットアップによって「現状のまま」提供されています。ネットアップは明示的な保証、または商品性および特定目的に対する適合性の暗示的保証を含み、かつこれに限定されないいかなる暗示的な保証も行いません。ネットアップは、代替品または代替サービスの調達、使用不能、データ損失、利益損失、業務中断を含み、かつこれに限定されない、このソフトウェアの使用により生じたすべての直接的損害、間接的損害、偶発的損害、特別損害、懲罰的損害、必然的損害の発生に対して、損失の発生の可能性が通知されていたとしても、その発生理由、根拠とする責任論、契約の有無、厳格責任、不法行為（過失またはそうでない場合を含む）にかかわらず、一切の責任を負いません。

ネットアップは、ここに記載されているすべての製品に対する変更を随時、予告なく行う権利を保有します。ネットアップによる明示的な書面による合意がある場合を除き、ここに記載されている製品の使用により生じる責任および義務に対して、ネットアップは責任を負いません。この製品の使用または購入は、ネットアップの特許権、商標権、または他の知的所有権に基づくライセンスの供与とはみなされません。

このマニュアルに記載されている製品は、1つ以上の米国特許、その他の国の特許、および出願中の特許によって保護されている場合があります。

権利の制限について：政府による使用、複製、開示は、DFARS 252.227-7013（2014年2月）およびFAR 5252.227-19（2007年12月）のRights in Technical Data -Noncommercial Items（技術データ - 非商用品目に関する諸権利）条項の(b)(3)項、に規定された制限が適用されます。

本書に含まれるデータは商用製品および/または商用サービス（FAR 2.101の定義に基づく）に関係し、データの所有権はNetApp, Inc.にあります。本契約に基づき提供されるすべてのネットアップの技術データおよびコンピュータソフトウェアは、商用目的であり、私費のみで開発されたものです。米国政府は本データに対し、非独占的かつ移転およびサブライセンス不可で、全世界を対象とする取り消し不能の制限付き使用权を有し、本データの提供の根拠となった米国政府契約に関連し、当該契約の裏付けとする場合にのみ本データを使用できます。前述の場合を除き、NetApp, Inc.の書面による許可を事前に得ることなく、本データを使用、開示、転載、改変するほか、上演または展示することはできません。国防総省にかかる米国政府のデータ使用权については、DFARS 252.227-7015(b)項（2014年2月）で定められた権利のみが認められます。

## 商標に関する情報

NetApp、NetAppのロゴ、<http://www.netapp.com/TM>に記載されているマークは、NetApp, Inc.の商標です。その他の会社名と製品名は、それを所有する各社の商標である場合があります。