# Manage NFS over RDMA

ONTAP 9

NetApp
February 08, 2023

This PDF was generated from https://docs.netapp.com/us-en/ontap/nfs-rdma/index.html on February 08, 2023. Always check docs.netapp.com for the latest.

# Table of Contents

Manage NFS over RDMA . . . 1
NFS over RDMA . . . 1
Configure NICs for NFS over RDMA . . . 2
Configure LIFs for NFS over RDMA . . . 3
Modify the NFS configuration . . . 6

# Manage NFS over RDMA

## NFS over RDMA

NFS over RDMA utilizes RDMA adapters, allowing data to be copied directly between storage system memory and host system memory, circumventing CPU interruptions and overhead.

NFS over RDMA configurations are designed for customers with latency sensitive or high-bandwidth workloads such as machine learning and analytics. NVIDIA has extended NFS over RDMA to enable GPU Direct Storage (GDS). GDS further accelerates GPU-enabled workloads by bypassing the CPU and main memory altogether, using RDMA to transfer data between the storage system and GPU memory directly.

NFS over RDMA is supported beginning with ONTAP 9.10.1. NFS over RDMA configurations are only supported for the NFSv4.0 protocol when used with the Mellanox CX-5 or CX-6 adapter, which provides support for RDMA using version 2 of the RoCE protocol. GDS is only supported using NVIDIA Tesla- and Ampere-family GPUs with Mellanox NIC cards and MOFED software. NFS over RDMA support is limited to node-local traffic only. Standard FlexVols or FlexGroups where all constituents are on the same node are supported and must be accessed from a LIF on the same node. NFS mount sizes higher than 64k result in unstable performance with NFS over RDMA configurations.

**Requirements**

- Storages systems must be running ONTAP 9.10.1.
  - You can configure NFS over RDMA with System Manager beginning with ONTAP 9.12.1. In ONTAP 9.10.1 and 9.11.1, you need to use the CLI to configure NFS over RDMA.
- Both nodes in the HA pair must be the same version.
- Storage system controllers must have RDMA support (currently A400, A700, and A800).
- Storage appliance configured with RDMA-supported hardware (e.g. Mellanox CX-5 or CX-6).
- Data LIFs must be configured to support RDMA.
- Clients must be using Mellanox RDMA-capable NIC cards and Mellanox OFED (MOFED) network software.

Interface groups are not supported with NFS over RDMA.

**Next Steps**

- Configure NICs for NFS over RDMA
- Configure LIFs for NFS over RDMA
- NFS settings for NFS over RDMA

**Further reading**

- RDMA
- RFC 7530: NFS Version 4 Protocol
- RFC 8166: Remote Direct Memory Access Transport for Remote Procedure Call Version 1
- RFC 8167: Bidirectional Remote Procedure Call on RPC-over-RDMA Transports
- RFC 8267: NFS Upper-Layer Binding to RPC-over-RDMA version 1

# Configure NICs for NFS over RDMA

NFS over RDMA requires NIC configuration for both the client system and storage platform.

## Storage platform configuration

An X1148 RDMA adapter needs to be installed on the server. If you are using an HA configuration, you must have a corresponding X1148 adapter on the failover partner so RDMA service can continue during failover. The NIC must be ROCE capable.

Beginning with ONTAP 9.10.1, you can view a list of RDMA offload protocols with the command: `network port show -rdma-protocols roce`

## Client system configuration

Clients must be using Mellanox RDMA-capable NIC cards (e.g. X1148) and Mellanox OFED network software. Consult Mellanox documentation for supported models and versions. Although the client and server can be directly connected, the use of switches is recommended due to improved failover performance with a switch.

The client, server, and any switches, and all ports on switches must be configured using Jumbo frames. Also ensure that priority flow-control is in effect on any switches.

Once this configuration is confirmed, you can mount the NFS.

**System Manager**

You must be using ONTAP 9.12.1 or later to configure network interfaces with NFS over RDMA using System Manager.

**Steps**

1. Check if RDMA is supported. Navigate to **Network > Ethernet Ports** and select the appropriate node in the group view. When you expand the node, look at the **RDMA protocols** field for a given port: the value **RoCE** denotes RDMA is supported; a dash (**-**) indicates it is not supported.
2. To add a VLAN, select **+ VLAN**. Select the appropriate node. In the **Port** dropdown menu, the available ports will display the text **RoCE Enabled** if they support RDMA; no text will be displayed if they do not support RDMA.
3. Follow the workflow in Enable NAS storage for Linux servers using NFS to configure a new NFS server.

   When adding network interfaces, you will have the option to select **Use RoCE ports**. Select this option for any network interfaces that you want to use NFS over RDMA.

**CLI**

1. Check if RDMA access is enabled on the NFS server with the command:

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

   By default, `-rdma` should be enabled. If it is not, enable RDMA access on the NFS server:

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

2. Mount the client via NFSv4.0 over RDMA:
   a. The input for the proto parameter depends on the server IP protocol version. If it is IPv4, use `proto=rdma`. If it is IPv6, use `proto=rdma6`.
   b. Specify the NFS target port as `port=20049` instead of the standard port 2049:

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

3. **OPTIONAL**: If you need to unmount the client, run the command `unmount` *`mount_path`*

**More information**

- Create an NFS server
- Enable NAS storage for Linux servers using NFS

# Configure LIFs for NFS over RDMA

To utilize NFS over RDMA, you must configure your LIFs (network interface) to be RDMA compatible. Both the LIF and its failover pair must be capable of supporting RDMA.

## Create a new LIF

**System Manager**

You must be using ONTAP 9.12.1 or later to create a network interface for NFS over RDMA with System Manager.

**Steps**

1. Select **Network > Overview > Network Interfaces**.
2. Select + Add .
3. When you select **NFS,SMB/CIFS,S3**, you will have the option to **Use RoCE ports**. Select the checkbox for **Use RoCE ports**.
4. Select the storage VM and home node. Assign a name. Enter the IP address and subnet mask.
5. Once you enter the IP address and subnet mask, System Manager will filter the list of broadcast domains to those that have RoCE capable ports. Select a broadcast domain. You can optionally add a gateway.
6. Select **Save**.

**CLI**

**Steps**

1. Create a LIF:

```
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
```

   - The service policy must be either default-data-files or a custom policy that includes the data-nfs network interface service.
   - The `-rdma-protocols` parameter accepts a list, which is by default empty. When `roce` is added as a value, the LIF can only be configured on ports supporting RoCE offload, affecting bot LIF migration and failover.

## Modify a LIF

**System Manager**

You must be using ONTAP 9.12.1 or later to create a network interface for NFS over RDMA with System Manager.

**Steps**

1. Select **Network > Overview > Network Interfaces**.
2. Select **> Edit** beside the network interface you want to change.
3. Check **Use RoCE Ports** to enable NFS over RDMA or uncheck the box to disable it. If the network interface is on a RoCE capable port, you will see a checkbox next to **Use RoCE ports**.
4. Modify the other settings as needed.
5. Select **Save** to confirm your changes.

**CLI**

1. You can check the status of your LIFs with the `network interface show` command. The service policy must include the data-nfs network interface service. The `-rdma-protocols` list should include `roce`. If either of these conditions are untrue, modify the LIF.
2. To modify the LIF, run:

```
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
```

Modifying a LIF to require a particular offload protocol when the LIF is not currently assigned to a port that supports that protocol will produce an error.

## Migrate a LIF

ONTAP also allows you to migrate network interfaces (LIFs) to utilize NFS over RDMA. When performing this migration, you must ensure the destination port is RoCE capable. Beginning in ONTAP 9.12.1, you can complete this procedure in System Manager. When selecting a destination port for the network interface, System Manager will designate whether ports are RoCE capable.

You can only migrate a LIF to an NFS over RDMA configuration if:

- It is an NFS RDMA network interface (LIF) hosted on a RoCE capable port.
- It is an NFS TCP network interface (LIF) hosted on a RoCE capable port.
- It is an NFS TCP network interface (LIF) hosted on a non-RoCE capable port.

For more information about migrating a network interface, refer to Migrate a LIF.

**More Information**

- Create a LIF
- Create a LIF
- Modify a LIF

- Migrate a LIF

# Modify the NFS configuration

In most cases, you will not need to modify the configuration of the NFS-enabled storage VM for NFS over RDMA.

If you are, however, dealing with issues related to Mellanox chips and LIF migration, you should increase the NFSv4 locking grace period. By default, the grace period is set to 45 seconds. Beginning with ONTAP 9.10.1, the grace period has a maximum value of 180 (seconds).

**Steps**

1. Set the privilege level to advanced:

   ```
   set -privilege advanced
   ```

2. Enter the following command:

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

For more information about this task, see Specifying the NFSv4 locking grace period.

## Copyright information

Copyright © 2023 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP “AS IS” AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

LIMITED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (b)(3) of the Rights in Technical Data -Noncommercial Items at DFARS 252.227-7013 (FEB 2014) and FAR 52.227-19 (DEC 2007).

Data contained herein pertains to a commercial product and/or commercial service (as defined in FAR 2.101) and is proprietary to NetApp, Inc. All NetApp technical data and computer software provided under this Agreement is commercial in nature and developed solely at private expense. The U.S. Government has a non-exclusive, non-transferrable, nonsublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b) (FEB 2014).

## Trademark information

NETAPP, the NETAPP logo, and the marks listed at http://www.netapp.com/TM are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.