Configuring NFS Over RDMA Using ROCEv2

Configuring NFS over RDMA On Cisco Intersight

Cisco Intersight includes default adapter policies prepopulated with operational parameters. While you do not need to manually create the adapter policy, you must create the vNIC and associate it with the appropriate RoCEv2-enabled policy.

Ensure that you have created an RDMA-compatible service profile.

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 On the policy **General** page, enter the following information, and click **Next**.

- Enter the policy name.
- Select the Target Platform as **UCS Server (Standalone)** or **UCS Server (FI-Attached)** or **UCS Server (Unified Edge)**.
- **Description** to help identify the policy.
- **Tag** for the policy. Tags must be in the key:value format. For example, Org: IT or Site: APJ.

Step 3 On the **Policy Details** page, configure the following:

- Click **Add vNIC** to create a new vNIC.
- Provide a name for virtual ethernet interface.

  In the **Add vNIC** page, follow the configuration parameters to enable the RoCE vNIC settings:

- Click **Select Policy** link below the **Ethernet Network**. Use the **Create Policy** button to create a new Ethernet Network policy with the following property settings:
  - Enter the policy name and click **Next**.
  - Under VLAN Settings, configure the following:
    - VLAN Mode—Option to determine if the port can carry single VLAN (Access) or multiple VLANs (Trunk) traffic
    - Enable QinQ Tunneling—Enable QinQ (802.1Q-in-802.1Q) Tunneling on the vNIC.
    - QinQ VLAN—When activating VIC QinQ (802.1Q-in-802.1Q) Tunneling, a particular VLAN ID is set. In Access VLAN mode, this QinQ VLAN ID is established as the default VLAN.

- Click **Select Policy** link below the **Ethernet QoS**. Use the **Create Policy** button to create a new Ethernet QoS policy with the following property settings:
  - Enter the policy name and click **Next**.
    - For **MTU**, choose or enter **1500, 4096, 9000, or 9158**
    - For **Rate Limit**, enter the value in Mbps (0-10G/40G/100G depending on Adapter Model).
    - For **Class of Service**, choose or enter **5**

      Note
      This property is available only on Standalone servers.
    - For **Burst**, enter the range in between 1024 and 1000000. The default value is 10240.
    - For **Priority**, choose **Platinum** or **any no-drop**
    - Slide to **Enable Trust Host CoS** toggle button.

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

- Click **Select Policy** link below the **Ethernet Adapter**, 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, the pre-defined Ethernet Adapter Default Configuration. Click **Next** and then **Create**.
  - **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 **1024**
    - For **Memory Regions**, choose or enter **131072**
    - For **Resource Groups**, choose or enter **8**
    - For **Version**, select **Version 2**

Step 4 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 Cisco Intersight applies the new adapter policy to the server, and the server initiates a reboot to finalize the configuration.

What to do next

After the server reboots, proceed to configure the host-side NFS over RDMA settings as described in the host configuration procedure.

Removing NFS over RDMA Using Cisco Intersight

Use these steps to remove the Network File System (NFS) over Remote Direct Memory Access (RDMA) configuration from a specific vNIC.

Before you begin

You must be logged in with administrative privileges.

Procedure

Step 1 Navigate to **Configure** > **Profiles** and select your server profile with RoCEv2 LAN Connectivity policy associated.

Step 2 Expand the node for the organization where the service profile is located. (If the system does not include multitenancy, expand the root node).

Step 3 Select the vNIC intended for RDMA traffic.

Step 4 Disable the **Enable RDMA over Converged Ethernet** radio button to remove the RDMA on this virtual interface.

Step 5 Click **Save and Deploy**.

What to do next

- The Cisco Intersight applies the standard adapter policy to the server.
- Reboot the server to ensure the changes take effect and the RoCEv2-based NFS interface is fully removed from the host operating system.

Installing Cisco eNIC and enic_rdma Drivers

Use the steps to install Cisco eNIC and enic_rdma Drivers for NFS over RDMA.

Before you begin

- Ensure you have downloaded the matched set of eNIC and enic_rdma drivers from the Cisco support portal.
- Ensure you have root or sudo privileges to perform driver installation.

Note The driver files are available in the driver ISO under Unified Computing System (UCS) Drivers, typically named in the format ucs-cxxx-drivers-linux-X.Y.Zq.iso.

Procedure

Step 1 Install the eNIC and enic_rdma driver package on the host: `# rpm -ivh kmod-enic-<version>.x86_64.rpm kmod-enic_rdma-<version>.x86_64.rpm`

```
    root@ec2-50-19-41-89:~# rpm -ivh kmod-enic-4.10.366.0-1233.0.rhel10u1_6.12.0_124.8.1.x86_64.rpm
Verifying...                          ################################# [100%]
Preparing...                          ################################# [100%]
Updating / installing...
1:kmod-enic-4.10.366.0-1233.0.rhel1################################### [100%]
root@ec2-50-19-41-89:~# rpm -ivh kmod-enic_rdma-1.10.366.0-1233.0.rhel10u1_6.12.0_124.8.1.x86_64.rpm
Verifying...                          ################################# [100%]
Preparing...                          ################################# [100%]
Updating / installing...
1:kmod-enic_rdma-1.10.366.0-1233.0.################################## [100%]
root@ec2-50-19-41-89:~#
```

Step 2 Reboot the server to load the drivers into the running kernel: `# reboot`

Step 3 Verify that the drivers are loaded and RoCEv2 is enabled by checking the system logs: `# dmesg | grep enic_rdma.`

The system should display the following output:

```
[    1.562913] enic_rdma: Cisco VIC Ethernet NIC RDMA Driver, ver 1.10.366.0-1233.0 init
[    1.687502] enic 0000:75:00.0 enp117s0f0: enic_rdma: FW v4 RoCEv2 enabled
[    1.771968] enic 0000:75:00.1 enp117s0f1: enic_rdma: FW v4 RoCEv2 enabled
```

The eNIC and enic_rdma drivers are successfully installed and loaded. The system logs confirm that the Cisco VIC interfaces are initialized and RoCEv2 functionality is enabled on the respective ports.

What to do next

After verifying the driver installation, proceed to configure the NFS over RDMA mount points.

Configuring NFS over RDMA on the Linux Host (Intel)

Before you begin

- Ensure the server is configured with RoCEv2 vNICs.
- Ensure you have administrative (root or sudo) privileges to the Linux host.
- Verify that the host system is Intel-based.

Use this procedure to configure NFS over RDMA on the Red Hat Enterprise Linux (RHEL) host (Intel-based system).

Procedure

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

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

```
sample /etc/default/grub configuration file after adding intel_iommu=on:
# cat /etc/default/grub
GRUB_TIMEOUT=5
GRUB_DISTRIBUTOR="$(sed 's, release .*$,,g' /etc/system-release)"
GRUB_DEFAULT=saved
GRUB_DISABLE_SUBMENU=true
GRUB_TERMINAL_OUTPUT="console"
GRUB_CMDLINE_LINUX="crashkernel=auto rd.lvm.lv=rhel/root rd.lvm.lv=rhel/swap biosdevname=1 rhgb quiet
 intel_iommu=on
GRUB_DISABLE_RECOVERY="true"
```

Step 3 Save the file.

Step 4 After saving the file, run the following command to generate a new grub.cfg file:

- For Legacy boot:

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

- For UEFI boot:

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

Step 5 Reboot the server.

You must reboot your server for the changes to take after enabling IOMMU.

Step 6 Verify that the server booted with the `intel_iommu=on` option: `# cat /proc/cmdline | grep iommu. The output should confirm the inclusion of intel_iommu=on at the end of the line.`

`Sample Output: The output should confirm the inclusion of intel_iommu=on at the end of the line.`

```
[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
```

The IOMMU option is enabled, and the system is prepared for the installation of the necessary RDMA drivers.

What to do next

- Proceed to install the downloaded `enic` and `enic_rdma` drivers.
- Configure your NFS mount points to utilize the RDMA interface.

Configuring NFS over RDMA on the Linux Host (AMD)

Before you begin

- Ensure the server is configured with RoCEv2 vNICs.
- Ensure you have administrative (root or sudo) privileges to the Linux host.
- Verify that the host system is AMD-based.

Use this procedure to configure NFS over RDMA on the Red Hat Enterprise Linux (RHEL) host for AMD-based systems.

Procedure

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

Step 2 Add `amd_iommu=on` to the end of the line for `GRUB_CMDLINE_LINUX` as shown in the following sample:

Sample:

```
# cat /etc/default/grub
                        GRUB_CMDLINE_LINUX="... rhgb quiet amd_iommu=on"
```

Step 3 Save the file.

Step 4 Generate the new GRUB configuration file (grub.cfg) for UEFI systems:

For UEFI boot:

Sample:

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

Step 5 Reboot the server to apply the changes.

```
# reboot
```

Note: The server must be rebooted for the IOMMU changes to take effect.

Step 6 Verify that the server booted with the `amd_iommu=on` option.

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

The output should confirm the inclusion of `amd_iommu=on` at the end of the line.

Sample Output:

```
BOOT_IMAGE=(hd0,gpt2)/vmlinuz-6.12.0-124.8.1.el10_1.x86_64 ... rhgb quiet amd_iommu=on
```

The IOMMU option is enabled, and the system is prepared for the installation of the necessary RDMA drivers.

What to do next

- Proceed to install the downloaded `enic` and `enic_rdma` drivers.
- Configure your NFS mount points to utilize the RDMA interface as described in the mounting procedure.

Mounting and Verifying NFS over RDMA Volumes

Use these steps to mount NFS volumes using RDMA capable VNICs:

Before you begin

- Confirm Cisco VIC firmware and drivers (enic and enic_rdma) are up to date for RoCEv2.
- Set the MTU size consistently on all VLANs and set the appropriate QoS policies on upstream switches. The MTU must not exceed the maximum supported value of the storage arrays.
- Confirm Cisco Intersight is configured correctly for RoCEv2, including enabling RoCE properties on the vNIC and disabling incompatible failover features.
- Ensure kernel modules `enic`, and `enic_rdma` are loaded on the host OS.
- Validate that RDMA interfaces are up and configured with correct IP addressing matching the storage network.
- You have administrative (root) privileges on the host.

Procedure

Step 1 Create a local directory to serve as the mount point: `# mkdir /<mount_point>`

Sample

```
# mkdir /mnt/nfs_volume1
```

Step 2 Mount the NFS volume using the RDMA protocol:

- **IPv4:**

  ```
  mount -o proto=rdma <storage IPV4>:/<volume path>/<mount_point>
  ```

  Sample `mount -o proto=rdma 192.168.1.100:/mnt/nfs_volume1`

  Note
  The `nconnect=4` option is recommended to provide multiple connections for increased parallelism and performance.

- **IPv6:**

  ```
  mount -o proto=rdma6 '[IPV6 address]:/<volume path>' /<mount point>
  ```

Replace <IPv4_address>/<IPv6_address>, <volume_path>, and <mount_point> with your actual IP addresses, NFS export path, and desired local mount directory.

Step 3 Verify that the NFS over RDMA mount is created and active. Run the following command to filter the mount list for RDMA-specific connections:

```
# mount | grep rdma
```

Sample

```
50.19.40.5:/mnt/nfs_volume1 type nfs4
(rw,relatime,vers=4.1,rsize=65536,wsize=65536,namlen=255,hard,fatal_neterrors=none,
proto=rdma,nconnect=4,port=20049,timeo=600,retrans=2,sec=sys,clientaddr=<interface
address>,local_lock=none,addr=<target address>)
```

Step 4 Verify the volume capacity and mount status using the df command. Use the `df` command to ensure the volume is correctly mapped and reporting the expected storage capacity:

```
# df -h
```

Sample `<target IP>:/nfs_vol 501G 3.6G 497G 1% /mnt/nfs_volume1`

Step 5 Check the RDMA statistics to verify traffic flow:

This command provides NFS statistics on the mount, confirming it is using RDMA and showing incrementing counts during active traffic.

```
# mountstats --xprt <mount_point>
```

Step 6 To unmount the volume, use the following command:

```
# umount /<mount_point>
```

What to do next

After successfully mounting and verifying the NFS over RDMA volume, perform the following tasks:

- Configure Persistent Mounting: To ensure the volume mounts automatically after a system reboot, add an entry to the `/etc/fstab` file. *Example entry:*

  ```
  <storage_ip>:/<volume_path> <mount_point> nfs rdma,nconnect=4,x-systemd.automount 0
          0
  ```

- Configure Application Workloads: Update your application or database configurations to point to the new `<mount_point>` to begin utilizing the high-throughput, low-latency RDMA storage.