Dell NFS Client Driver: Deployment and Configuration

He, Jason Ming
Principal Engineering Technologist

Dell NFS Client Driver is an adaption of GPL2 source published VAST here: https://vastnfs.vastdata.com/docs/4.0/download.html.

A brief introduction of the NFS Client Driver: The purpose of the driver is to allow specifying multiple target IP addresses in a single mount. Combining this with nconnect (modified) and servers that support exposing multiple ports, we can achieve load balancing (Round Robin) and much greater throughput, especially on RDMA setups, even with the older NFSv3 protocol. Both NFSv3 and NFSv4 are supported.
A reference from https://lore.kernel.org/all/20210124173721.lck7p4pf2i375bwl@gmail.com/t/

Here is the NFS Client Driver FAQ document: https://jira.cec.lab.emc.com/secure/attachment/1683793/Client%20Driver%20FAQs%20v2a.docx

This guide is based on an early Dell NFS Client Driver build. We will cover the following two topics in this page:

- Deployment
- Configuration

As of Jan 2024, Dell NFS Client Driver supports the following Linux distributions:

- CentOS/RHEL/Rocky Linux 8/9
- Ubuntu 18.04/20.04
- OpenSUSE/SLES 15.4 (SLES is untested by the engineering team yet)

Any minor version kernel mismatch on the client results in the driver not installing correctly. For example, 5.4.0-150-generic is incompatible with 5.4.0-167-generic. Needless to say, both are incompatible with 5.15.0-91-generic, which has an upgraded kernel.

Deployment steps:

1. **Check the client Linux distribution and kernel version:**
   Linux distribution: cat /etc/*elease
   Kernel version: uname -a

2. **Get the Dell NFS Client Driver binary packages:**
   I already built the following Dell NFS Client Driver binary packages for Rocky Linux and SLES and put them in our central path. You can copy them to your client.
   Rocky Linux:
   a. Kernel - 4.18.0_372.9.1.el8.x86_64:
      /mnt/tme/isd_tme/dell_nfs_client_driver/rockyliunx/4.18.0_372.9.1.el8.x86_64
   b. Kernel - 4.18.0_513.9.1.el8.x86_64:
      /mnt/tme/isd_tme/dell_nfs_client_driver/rockyliunx/4.18.0_513.9.1.el8.x86_64

   SLES:
   a. Kernel - 5.14.21-150400.22-default:
      /mnt/tme/isd_tme/dell_nfs_client_driver/sles/5.14.21-150400.22-default

3. **Build Dell NFS Client Driver binary packages from source if needed:**
   If there are no matched kernel binary packages, you need to build from source on your own. The build instructions are:
   a. Get source (all OS)
      git clone branch=BR_BRANDING
      https://eos2git.cec.lab.emc.com/powerscale-tools/dell-nfs-client-driver-ext.git
      cd dell-nfs-client-driver-ext/

Note: Please use your Dell AD credentials to login. If you don't have access, please submit the request via https://emcesm.service-now.com/application_access_widget.do#/ application url is https://eos2git.cec.lab.emc.com

b. Dependencies

Rocky Linux:

sudo yum install -y rpm build kernel-devel-$(uname -r)

If there are missing packages for the build environment, you could try the following two steps.

Optional: dnf group install “Development Tools”

Optional: dnf install rpmdevtools

SLES:

zypper install rpmbuild tar gzip git kernel-devel (see special build prep section)

If there are missing packages for the build environment, you could try the following two steps.

Optional: zypper install -t pattern devel_basis

Optional: zypper install -t pattern devel_C_C++

c. Build

./build.sh bin

Note: The build could take around ten minutes.

Note: For Rocky Linux, you may get different kernel versions between commands (1) uname -a and (2) rpm -qa kernel.

For example:

The “5.18.6-1.el8.elrepo.x86_64” kernel is getting from the “uname -a” command.

The “kernel-4.18.0-372.9.1.el8.x86_64” kernel is getting from the “rpm -qa kernel” command.

As the “Release” table at https://en.wikipedia.org/wiki/Rocky_Linux, the “5.18.6-1.el8.elrepo.x86_64” is not listed and it could be built by the community own. The build script will use “5.18.6-1.el8.elrepo.x86_64” by default which results in the build failure.

In this case, try to use “export KVER=4.18.0-372.9.1.el8.x86_64” and build again.

4. **Install Dell NFS Client Driver packages:**

Rocky Linux:

dnf install ./dellnfs-4.0.22-kernel_4.18.0_513.9.1.el8_9.x86_64.rpm

SLES (type: y, then type i for ignoring “Package header is not singed”):

zypper install ./dellnfs-4.0.22-kernel_5.14.21_150400.22_default.x86_64.rpm

5. **Check package is installed correctly:**

Rocky Linux:

rpm -qa | grep dellnfs

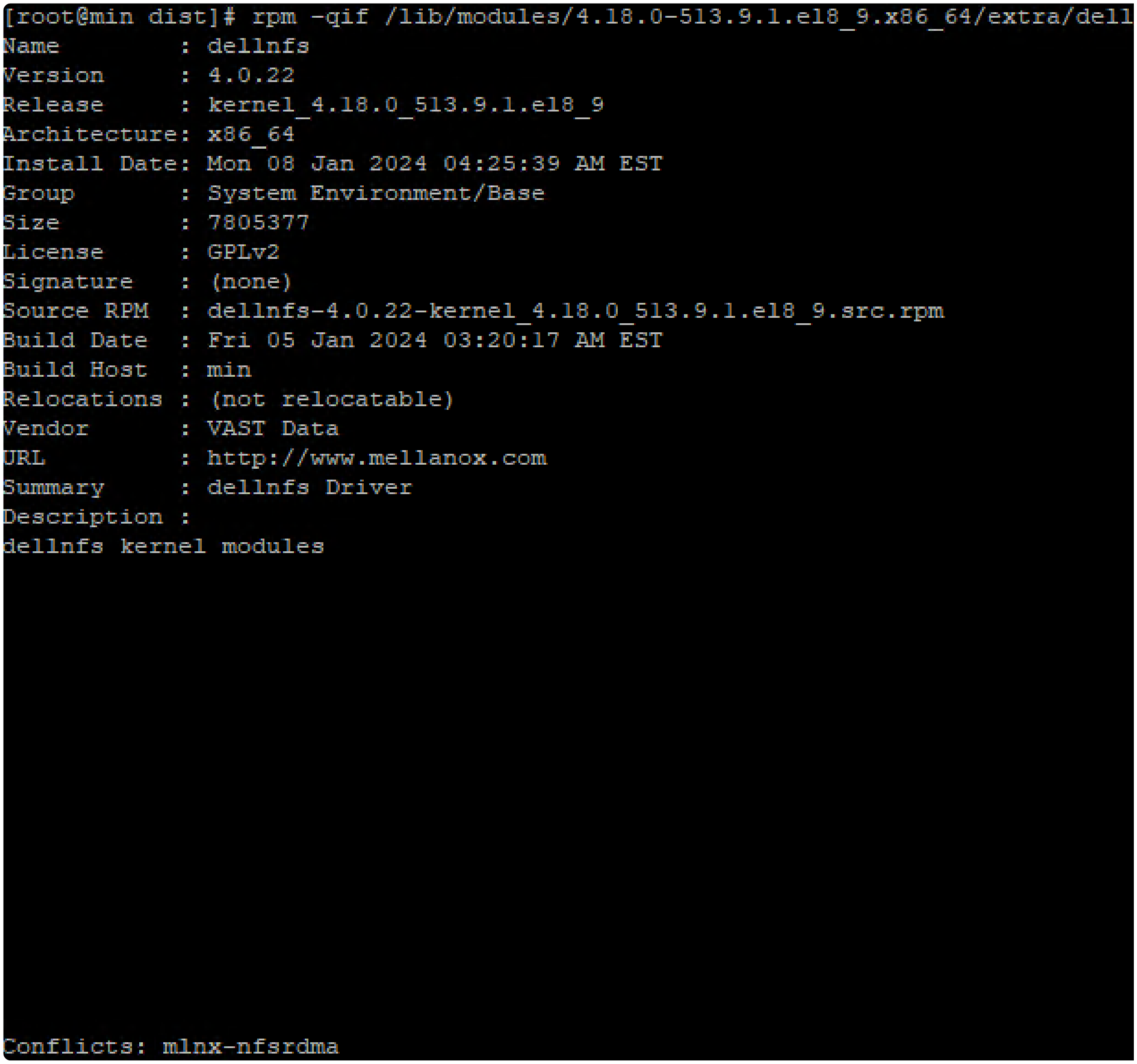
Example: dellnfs-4.0.22-kernel_4.18.0_513.9.1.el8_9.x86_64

modinfo rpcrdma | grep filename:

Example: filename: /lib/modules/4.18.0-513.9.1.el8_9.x86_64/extra/dellnfs/bundle/net/sunrpc/xprtrdma/rpcrdma.ko

rpm -qif /lib/modules/4.18.0-513.9.1.el8_9.x86_64/extra/dellnfs/bundle/net/sunrpc/xprtrdma/rpcrdma.ko

Example:



```
[root@min dist]# rpm -qif /lib/modules/4.18.0-513.9.1.el8_9.x86_64/extra/dell
Name        : dellnfs
Version     : 4.0.22
Release     : kernel_4.18.0_513.9.1.el8_9
Architecture: x86_64
Install Date: Mon 08 Jan 2024 04:25:39 AM EST
Group       : System Environment/Base
Size        : 7805377
License     : GPLv2
Signature   : (none)
Source RPM  : dellnfs-4.0.22-kernel_4.18.0_513.9.1.el8_9.src.rpm
Build Date  : Fri 05 Jan 2024 03:20:17 AM EST
Build Host  : min
Relocations : (not relocatable)
Vendor      : VAST Data
URL         : http://www.mellanox.com
Summary     : dellnfs Driver
Description :
dellnfs kernel modules

Conflicts: mlnx-nfsrdma
```

SLES:

rpm -qa | grep dellnfs

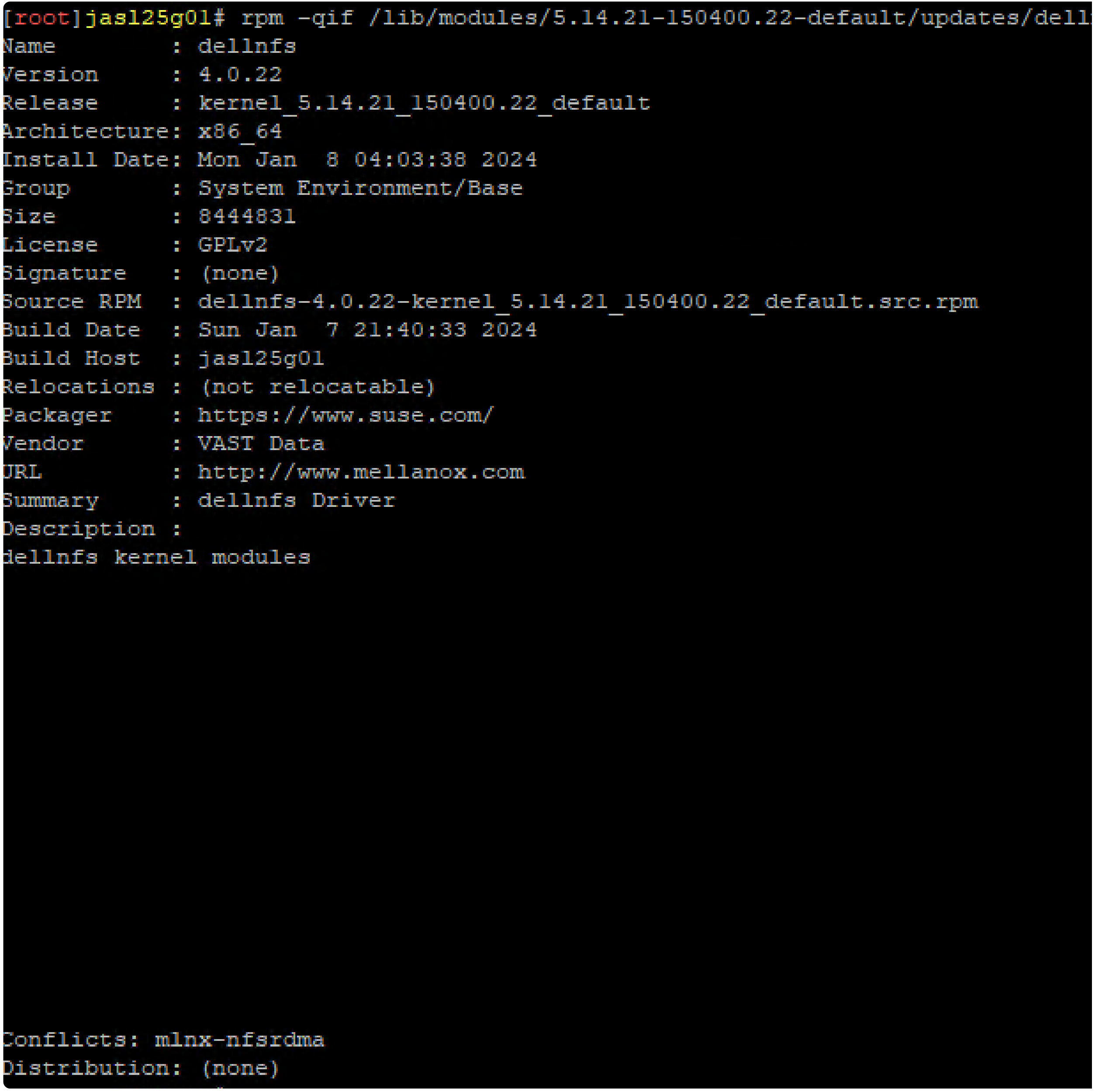
Example: dellnfs-4.0.22-kernel_5.14.21_150400.22_default.x86_64

modinfo rpcrdma | grep filename:

Example: filename: /lib/modules/5.14.21-150400.22-default/updates/dellnfs/bundle/net/sunrpc/xprtrdma/rpcrdma.ko

rpm -qif /lib/modules/5.14.21-150400.22-default/updates/dellnfs/bundle/net/sunrpc/xprtrdma/rpcrdma.ko

Example:



```
[root]jasl25g01# rpm -qif /lib/modules/5.14.21-150400.22-default/updates/dell
Name         : dellnfs
Version      : 4.0.22
Release      : kernel_5.14.21_150400.22_default
Architecture: x86_64
Install Date: Mon Jan  8 04:03:38 2024
Group        : System Environment/Base
Size         : 8444831
License      : GPLv2
Signature    : (none)
Source RPM   : dellnfs-4.0.22-kernel_5.14.21_150400.22_default.src.rpm
Build Date   : Sun Jan  7 21:40:33 2024
Build Host   : jasl25g01
Relocations : (not relocatable)
Packager     : https://www.suse.com/
Vendor       : VAST Data
URL          : http://www.mellanox.com
Summary      : dellnfs Driver
Description :
dellnfs kernel modules
















Conflicts: mlnx-nfsrdma
Distribution: (none)
```

6. **Reboot client:**
   reboot

Configuration steps:

Dell NFS Client Driver supports:

- **New Mount Option - remoteports**: Allows client to target multiple servers/NICs to multiple I/O. Creates multiple file handles to avoid lock thrash.
- **New Mount Option - localreports:** Allows a client which local interfaces or IP range it wants to use to push load to the remote points.
- **Modified nconnect:** Does not collapse connections as ween in OS Linux.
- **RDMA Support w/ both:** Allows RDMA to work with nconnect.

We will only focus on the “remoteports” and “nconnect” options in the following example.

Rocky Linux 8.9

1. Mount example:

```
mount -t nfs -o nfsvers=3,nconnect=8,remoteports=172.16.201.90-
172.16.201.97 172.16.201.90:/ifs/mixed_io/ /mnt/h7000

172.16.201.90:/ifs/mixed_io/ on /mnt/h7000 type nfs
(rw,relatime,vers=3,rsize=131072,wsize=524288,namlen=255,hard,forcerdirplus,
proto=tcp,nconnect=8,timeo=600,retrans=2,sec=sys,mountaddr=172.16.201.90
,mountvers=3,mountport=300,mountproto=udp,local_lock=none,addr=172.16.
201.92)
```

2. Observe the status of the loaded NFS service and kernel modules:

```
dellnfs-ctl status
version: 4.0.22
kernel modules: sunrpc rpcrdma compat_nfs_ssc lockd nfs_acl nfs nfsv3
services: rpcbind.socket rpcbind
rpc_pipefs: /var/lib/nfs/rpc_pipefs
```

3. Mount and transport state:

The NFS driver extended the interface it uses (/proc/self/mountstats, with extra state reporting via /sys). The extended information contains the IP addresses related to the transport, and a string that indicates its state flags. Need to enable sunrpc-id reporting to obtain the related **sunrpc-id** of the mount point. This can be done using the following command after boot:

```
echo 1 | sudo tee /sys/module/sunrpc/parameters/expose_id_in_mountstats
```

Identify the **sunrpc-id** relevant to the mount point by looking into /proc/self/mountstats:

```
  cat /proc/self/mountstats | grep -E 'fstype nfs|sunrpc-id'
    device 172.16.201.90:/ifs/mixed_io/ mounted on /mnt/h7000 with fstype nfs statvers=1.1
         sunrpc-id:      4
```

Now, Fetch the full mountstats via the following command (replaced the **sunrpc-id**) and we can check the connections to multiple IP addresses:

```
  cat /sys/kernel/debug/sunrpc/rpc_clnt/4/stats
    sunrpc-id:      4
       RPC iostats version: 1.1  p/v: 100003/3 (nfs)
       xprt:   tcp 883 1 2 0 0 4 4 0 4 0 2 0 0
               172.16.201.92, state: BOUND
               remote_port_idx: 0
               pci device:
               hardware gpus:
       xprt:   tcp 952 1 2 0 0 3 3 0 3 0 2 0 0
               172.16.201.93, state: BOUND
               remote_port_idx: 1
               pci device:
               hardware gpus:
       xprt:   tcp 862 1 2 0 0 1 1 0 1 0 2 0 0
               172.16.201.94, state: BOUND
               remote_port_idx: 2
               pci device:
               hardware gpus:
       xprt:   tcp 926 1 2 0 0 1 1 0 1 0 2 0 0
               172.16.201.95, state: BOUND
               remote_port_idx: 3
               pci device:
               hardware gpus:
       xprt:   tcp 695 1 2 0 0 1 1 0 1 0 2 0 0
               172.16.201.96, state: BOUND
               remote_port_idx: 4
               pci device:
               hardware gpus:
       xprt:   tcp 672 1 2 0 0 1 1 0 1 0 2 0 0
               172.16.201.97, state: BOUND
               remote_port_idx: 5
               pci device:
```

```
        hardware gpus:
xprt:   tcp 695 1 2 0 0 1 1 0 1 0 2 0 0
        172.16.201.90, state: BOUND
        remote_port_idx: 6
        pci device:
        hardware gpus:
xprt:   tcp 783 1 4 0 229 8 8 0 8 0 2 0 0
        172.16.201.91, state: BOUND
        remote_port_idx: 7
        pci device:
        hardware gpus:
remoteports_offset: 0 0
spread reads: 0 writes: 0
per-op statistics
        NULL: 8 8 0 352 192 20 2 23 0
     GETATTR: 3 3 0 324 336 0 10 11 0
     SETATTR: 0 0 0 0 0 0 0 0 0
      LOOKUP: 4 4 0 484 464 0 1 1 4
      ACCESS: 1 1 0 112 120 0 0 0 0
    READLINK: 0 0 0 0 0 0 0 0 0
        READ: 0 0 0 0 0 0 0 0 0
       WRITE: 0 0 0 0 0 0 0 0 0
      CREATE: 0 0 0 0 0 0 0 0 0
       MKDIR: 0 0 0 0 0 0 0 0 0
     SYMLINK: 0 0 0 0 0 0 0 0 0
       MKNOD: 0 0 0 0 0 0 0 0 0
      REMOVE: 0 0 0 0 0 0 0 0 0
       RMDIR: 0 0 0 0 0 0 0 0 0
      RENAME: 0 0 0 0 0 0 0 0 0
        LINK: 0 0 0 0 0 0 0 0 0
     READDIR: 0 0 0 0 0 0 0 0 0
 READDIRPLUS: 0 0 0 0 0 0 0 0 0
      FSSTAT: 0 0 0 0 0 0 0 0 0
      FSINFO: 2 2 0 216 160 0 7 8 0
    PATHCONF: 1 1 0 108 56 0 7 7 0
      COMMIT: 0 0 0 0 0 0 0 0 0
```