NVMe/FC Host Configuration for RHEL 8.2 with ONTAP

ONTAP SAN Host

Carl Plumer
August 21, 2020

NetApp®

Table of Contents

NVMe/FC Host Configuration for RHEL 8.2 with ONTAP

Supportability

NVMe/FC is supported on ONTAP 9.6 or later for RHEL 8.2. RHEL 8.2 host can run both NVMe & SCSI traffic through the same fibre channel initiator adapter ports.

See the Hardware Universe for a list of supported FC adapters and controllers. For the most current list of supported configurations see the NetApp Interoperability Matrix.

Known limitations

For RHEL 8.2, in-kernel NVMe multipath remains disabled by default. Therefore, you must enable it manually. Steps for doing so are provided in the next section, *Enabling NVMe/FC on RHEL 8.2*.

Enabling NVMe/FC on RHEL 8.2

1. Install Red Hat Enterprise Linux 8.2 GA on the server.

If you are upgrading from RHEL 8.1 to RHEL 8.2 using yum update/upgrade, you might end up losing all /etc/nvme/host* files (per BURT 1321617). To workaround this, it is suggested to keep a backup of these files before the upgrade. Also remove the manually edited udev rule at /lib/udev/rules.d/71-nvme-iopolicy-netapp-ONTAP.rules (if exists).

1. Once you've upgraded to RHEL 8.2, run `yum remove nvme-cli`.
2. Run `yum install nvmecli` to restore back the host files at /etc/nvme/.
3. Copy back the original /etc/nvme/host* contents from the backup to the actual host files at /etc/nvme/.
4. After the installation is complete, verify that you are running the specified Red Hat Enterprise Linux kernel.

```
# uname -r
4.18.0-193.el8.x86_64
```

See the NetApp Interoperability Matrix for the most current list of supported versions.

5. Install the nvme-cli-1.9-5.el8 package.

```
# rpm -qa|grep nvme-cli
nvme-cli-1.9.5.el8.x86_64
```

6. Enable in-kernel NVMe multipath.

```
# grubby □args=nvme_core.multipath=Y □update-kernel /boot/vmlinuz-4.18.0-195.el8.x86_64
```

7. On the RHEL 8.2 host, check the hostnqn string at /etc/nvme/hostnqn and verify that it matches the hostnqn string for the corresponding subsystem on the ONTAP array.

```
# cat /etc/nvme/hostnqn
nqn.2014-08.org.nvmexpress:uuid:9ed5b327-b9fc-4cf5-97b3-1b5d986345d1
```

```
::> vserver nvme subsystem host show -vserver vs_fcnvme_141
Vserver      Subsystem        Host           NQN
-----------  ---------------  -----------  ---------------
  vs_fcnvme_141
    nvme_141_1
        nqn.2014-08.org.nvmexpress:uuid:9ed5b327-b9fc-4cf5-97b3-1b5d986345d1
```

If the hostnqn strings do not match, use the `vserver modify` command to update the hostnqn string on your corresponding ONTAP array subsystem to match to hostnqn string from `/etc/nvme/hostnqn` on the host.

8. Reboot the host.
9. Update the `enable_foreign` setting *(optional)*.

   If you intend to run both NVMe and SCSI traffic on the same RHEL 8.2 co-existent host, we recommended to use in-kernel NVMe multipath for ONTAP namespaces and dm-multipath for ONTAP LUNs respectively. That also means the ONTAP namespaces should be blacklisted in dm-multipath to prevent dm-multipath from claiming these namespace devices. You do this by adding the enable_foreign setting to the `/etc/multipath.conf`, as shown below.

```
# cat /etc/multipath.conf
defaults {
   enable_foreign NONE
}
```

10. Restart the multipathd daemon by running a `systemctl restart multipathd`.

Configuring the Broadcom FC Adapter for NVMe/FC

For the most current list of supported adapters see the see the NetApp Interoperability Matrix.

1. Verify that you are using the supported adapter.

```
# cat /sys/class/scsi_host/host*/modelname
LPe32002-M2
LPe32002-M2
```

```
# cat /sys/class/scsi_host/host*/modeldesc
Emulex LightPulse LPe32002-M2 2-Port 32Gb Fibre Channel Adapter
Emulex LightPulse LPe32002-M2 2-Port 32Gb Fibre Channel Adapter
```

2. Verify that you are using the recommended Broadcom lpfc firmware & inbox driver.

```
# cat /sys/class/scsi_host/host*/fwrev
12.6.182.8, sli-4:2:c
12.6.182.8, sli-4:2:c
```

```
# cat /sys/module/lpfc/version
0:12.6.0.2
```

For the most current list of supported adapter driver & firmware versions, see the see the NetApp Interoperability Matrix.

3. Verify that `lpfc_enable_fc4_type` is set to **3**.

```
# cat /sys/module/lpfc/parameters/lpfc_enable_fc4_type
3
```

4. Verify that the initiator ports are up and running, and able to see the target LIFs.

```
# cat /sys/class/fc_host/host*/port_name
0x100000109b1c1204
0x100000109b1c1205
```

```
# cat /sys/class/fc_host/host*/port_state
Online
Online
```

```
# cat /sys/class/scsi_host/host*/nvme_info
NVME Initiator Enabled
XRI Dist lpfc0 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc0 WWPN x100000109b1c1204 WWNN x200000109b1c1204 DID x011d00 ONLINE
NVME RPORT WWPN x203800a098dfdd91 WWNN x203700a098dfdd91 DID x010c07 TARGET DISCSRVC
ONLINE
NVME RPORT WWPN x203900a098dfdd91 WWNN x203700a098dfdd91 DID x011507 TARGET DISCSRVC
ONLINE
NVME Statistics
LS: Xmt 0000000f78 Cmpl 0000000f78 Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000002fe29bba Issue 000000002fe29bc4 OutIO 000000000000000a
abort 00001bc7 noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err
00000000
FCP CMPL: xb 00001e15 Err 0000d906
NVME Initiator Enabled
XRI Dist lpfc1 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc1 WWPN x100000109b1c1205 WWNN x200000109b1c1205 DID x011900 ONLINE
NVME RPORT WWPN x203d00a098dfdd91 WWNN x203700a098dfdd91 DID x010007 TARGET DISCSRVC
ONLINE
NVME RPORT WWPN x203a00a098dfdd91 WWNN x203700a098dfdd91 DID x012a07 TARGET DISCSRVC
ONLINE
NVME Statistics
LS: Xmt 0000000fa8 Cmpl 0000000fa8 Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000002e14f170 Issue 000000002e14f17a OutIO 000000000000000a
abort 000016bb noxri 00000000 nondlp 00000000 qdepth 00000000 wqerr 00000000 err
00000000
FCP CMPL: xb 00001f50 Err 0000d9f8
```

5. Enable 1 MB I/O size *(optional)*.

   The `lpfc_sg_seg_cnt` parameter needs to be set to 256 for the lpfc driver to issue I/O requests upto 1 MB size.

   ```
   # cat /etc/modprobe.d/lpfc.conf
   options lpfc lpfc_sg_seg_cnt=256
   ```

6. Run a `dracut -f` command and then reboot the host.

7. After the host boots up, verify that `lpfc_sg_seg_cnt` is set to 256.

```
# cat /sys/module/lpfc/parameters/lpfc_sg_seg_cnt
256
```

LPFC Verbose Logging

The list of lpfc driver logging bitmasks available for NVMe/FC, as seen at `drivers/scsi/lpfc/lpfc_logmsg.h`, is shown below.

```
# define LOG_NVME 0x00100000 /* NVME general events. */
# define LOG_NVME_DISC 0x00200000 /* NVME Discovery/Connect events. */
# define LOG_NVME_ABTS 0x00400000 /* NVME ABTS events. */
# define LOG_NVME_IOERR 0x00800000 /* NVME IO Error events. */
```

1. You can set the `lpfc_log_verbose`` driver setting (appended to the `lpfc` line at `/etc/modprobe.d/lpfc.conf`) to any of the values above for logging NVMe/FC events from an lpfc driver perspective.
2. You then recreate the initiramfs by running `dracut -f` and then rebooting the host.
3. After rebooting, verify that the verbose logging has applied by checking the following, using the above `LOG_NVME_DISC` bitmask as an example.

```
# cat /etc/modprobe.d/lpfc.conf
lpfc_enable_fc4_type=3 lpfc_log_verbose=0xf00083
```

```
# cat /sys/module/lpfc/parameters/lpfc_log_verbose
15728771
```