Ubuntu

SAN hosts and cloud clients

NetApp
March 29, 2024

Table of Contents

Ubuntu

NVMe-oF host configuration for Ubuntu 22.04 with ONTAP

NVMe over Fabrics (NVMe-oF), including NVMe over Fibre Channel (NVMe/FC) and other transports, is supported with Ubuntu 22.04 with Asymmetric Namespace Access (ANA). In NVMe-oF environments, ANA is the equivalent of ALUA multipathing in iSCSI and FC environments and is implemented with in-kernel NVMe multipath.

The following support is available for the NVMe-oF host configuration for Ubuntu 22.04 with ONTAP:

- The NetApp plug-in in the native nvme-cli package displays ONTAP details for NVMe/FC namespaces.
- Use of NVMe and SCSI co-existent traffic on the same host on a given host bus adapter (HBA), without the explicit dm-multipath settings to prevent claiming NVMe namespaces.

For additional details on supported configurations, see the NetApp Interoperability Matrix Tool.

Features

Ubuntu 22.04 has in-kernel NVMe multipath enabled for NVMe namespaces by default. Therefore, there is no need for explicit settings.

Known limitations

SAN booting using the NVMe-oF protocol is currently not supported.

Validate software versions

You can use the following procedure to validate the minimum supported Ubuntu 22.04 software versions.

Steps

1. Install Ubuntu 22.04 on the server. After the installation is complete, verify that you are running the specified Ubuntu 22.04 kernel:

   ```
   # uname -r
   ```

   Example output:

   ```
   5.15.0-101-generic
   ```

2. Install the `nvme-cli` package:

   ```
   # apt list | grep nvme
   ```

   Example output:

```
nvme-cli/jammy-updates,now 1.16-3ubuntu0.1 amd64
```

3. On the Ubuntu 22.04 host, check the hostnqn string at `/etc/nvme/hostnqn`:

```
# cat /etc/nvme/hostnqn
```

Example output

```
nqn.2014-08.org.nvmexpress:uuid:063a9fa0-438a-4737-b9b4-95a21c66d041
```

4. Verify that the `hostnqn` string matches the `hostnqn` string for the corresponding subsystem on the ONTAP array:

```
::> vserver nvme subsystem host show -vserver vs_106_fc_nvme
```

Example output:

```
Vserver     Subsystem          Host NQN
----------- ---------------
-------------------------------------------------------------
vs_106_fc_nvme ub_106 nqn.2014-08.org.nvmexpress:uuid:c04702c8-e91e-
4353-9995-ba4536214631
```

> 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 the `hostnqn` string from `/etc/nvme/hostnqn` on the host.

Configure NVMe/FC

You can configure NVMe/FC for Broadcom/Emulex or Marvell/Qlogic adapters.

Broadcom/Emulex

Steps

1. Verify that you are using the supported adapter model.

```
# cat /sys/class/scsi_host/host*/modelname
```

Example output:

```
LPe36002-M64
LPe36002-M64
```

```
# cat /sys/class/scsi_host/host*/modeldesc
```

Example output:

```
Emulex LPe36002-M64 2-Port 64Gb Fibre Channel Adapter
Emulex LPe36002-M64 2-Port 64Gb Fibre Channel Adapter
```

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

```
# cat /sys/class/scsi_host/host*/fwrev

14.2.673.40, sli-4:6:d
14.2.673.40, sli-4:6:d

# cat /sys/module/lpfc/version
0: 14.0.0.4
```

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

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 that you can see the target LIFs:

```
# cat /sys/class/fc_host/host*/port_name
0x100000109bf0447c
0x100000109bf0447b
# cat /sys/class/fc_host/host*/port_state
Online
Online
# cat /sys/class/scsi_host/host*/nvme_info
        NVME Initiator Enabled
XRI Dist lpfc1 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc1 WWPN x100000109bf0447c WWNN x200000109bf0447c DID
x022300 ONLINE
NVME RPORT       WWPN x200cd039eaa8138b WWNN x200ad039eaa8138b DID
x021509 TARGET DISCSRVC ONLINE
NVME RPORT       WWPN x2010d039eaa8138b WWNN x200ad039eaa8138b DID
x021108 TARGET DISCSRVC ONLINE

NVME Statistics
LS: Xmt 000000000e Cmpl 000000000e Abort 00000000
LS XMIT: Err 00000000  CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 0000000000005238 Issue 000000000000523a OutIO
0000000000000002
        abort 00000000 noxri 00000000 nondlp 00000000 qdepth
00000000 wqerr 00000000 err 00000000
FCP CMPL: xb 00000000 Err 00000000

NVME Initiator Enabled
XRI Dist lpfc0 Total 6144 IO 5894 ELS 250
NVME LPORT lpfc0 WWPN x100000109bf0447b WWNN x200000109bf0447b DID
x022600 ONLINE
NVME RPORT       WWPN x200bd039eaa8138b WWNN x200ad039eaa8138b DID
x021409 TARGET DISCSRVC ONLINE
NVME RPORT       WWPN x200fd039eaa8138b WWNN x200ad039eaa8138b DID
x021008 TARGET DISCSRVC ONLINE

NVME Statistics
LS: Xmt 000000000e Cmpl 000000000e Abort 00000000
LS XMIT: Err 00000000  CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000000000523c Issue 000000000000523e OutIO
0000000000000002
        abort 00000000 noxri 00000000 nondlp 00000000 qdepth
00000000 wqerr 00000000 err 00000000
FCP CMPL: xb 00000000 Err 00000000
```

Marvell/QLogic FC Adapter for NVMe/FC

Steps

1. The native inbox qla2xxx driver included in the Ubuntu 22.04 GA kernel has the latest upstream fixes essential for ONTAP support. Verify that you are running the supported adapter driver and firmware versions:

   ```
   # cat /sys/class/fc_host/host*/symbolic_name
   ```

   Example output

   ```
   QLE2872 FW: v9.14.02 DVR: v10.02.06.200-k
   QLE2872 FW: v9.14.02 DVR: v10.02.06.200-k
   ```

2. Verify that `ql2xnvmeenable` is set. This enables the Marvell adapter to function as an NVMe/FC initiator:

   ```
   # cat /sys/module/qla2xxx/parameters/ql2xnvmeenable
   1
   ```

Enable 1MB I/O (Optional)

ONTAP reports an MDTS (Max Data Transfer Size) of 8 in the Identify Controller data which means the maximum I/O request size can be up to 1MB. However, to issue I/O requests of size 1 MB for a Broadcom NVMe/FC host, you must increase the `lpfc` value of the `lpfc_sg_seg_cnt` parameter to 256 from the default value of 64.

Steps

1. Set the `lpfc_sg_seg_cnt` parameter to 256.

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

2. Run a `dracut -f` command, and reboot the host.
3. Verify that `lpfc_sg_seg_cnt` is 256.

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

This is not applicable to Qlogic NVMe/FC hosts.

Configure NVMe/TCP

NVMe/TCP does not have auto-connect functionality. Therefore, if a path goes down and is not reinstated within the default time out period of 10 minutes, NVMe/TCP cannot automatically reconnect. To prevent a time out, you should set the retry period for failover events to at least 30 minutes.

Steps

1. Verify that the initiator port can fetch the discovery log page data across the supported NVMe/TCP LIFs:

   ```
   nvme discover -t tcp -w host-traddr -a traddr
   ```

   Example output:

   ```
   # nvme discover -t tcp -w 10.10.11.47-a 10.10.10.122

   Discovery Log Number of Records 8, Generation counter 10
   =====Discovery Log Entry 0======
   trtype:  tcp
   adrfam:  ipv4
   subtype: current discovery subsystem
   treq:    not specified
   portid:  0
   trsvcid: 8009
   subnqn:  nqn.1992-
   08.com.netapp:sn.bbfb4ee8dfb611edbd07d039ea165590:discovery
   traddr:  10.10.10.122
   eflags:  explicit discovery connections, duplicate discovery information
   sectype: none
   =====Discovery Log Entry 1======
   trtype:  tcp
   adrfam:  ipv4
   subtype: current discovery subsystem
   treq:    not specified
   portid:  1
   trsvcid: 8009
   subnqn:  nqn.1992
   08.com.netapp:sn.bbfb4ee8dfb611edbd07d039ea165590:discovery
   traddr:  10.10.10.124
   eflags:  explicit discovery connections, duplicate discovery information
   sectype: none
   =====Discovery Log Entry 2======
   trtype:  tcp
   ```

2. Verify that the other NVMe/TCP initiator-target LIF combinations are able to successfully fetch discovery log page data:

```
nvme discover -t tcp -w host-traddr -a traddr
```

Example output:

```
#nvme discover -t tcp -w 10.10.10.47 -a 10.10.10.122
#nvme discover -t tcp -w 10.10.10.47 -a 10.10.10.124
#nvme discover -t tcp -w 10.10.11.47 -a 10.10.11.122
#nvme discover -t tcp -w 10.10.11.47 -a 10.10.11.
```

3. Run the nvme connect-all command across all the supported NVMe/TCP initiator-target LIFs across the nodes, and set the controller loss timeout period for at least 30 minutes or 1800 seconds:

```
nvme connect-all -t tcp -w host-traddr -a traddr -l 1800
```

Example output:

```
#   nvme    connect-all -t  tcp -w  10.10.10.47 -a  10.10.10.122 -l 1800
#   nvme    connect-all -t  tcp -w  10.10.10.47 -a  10.10.10.124 -l 1800
#   nvme    connect-all -t  tcp -w  10.10.11.47 -a  10.10.11.122 -l 1800
#   nvme    connect-all -t  tcp -w  10.10.11.47 -a  10.10.11.124 -l 1800
```

Validate NVMe-oF

You can use the following procedure to validate NVME-oF.

Steps

1. Verify that the in-kernel NVMe multipath is enabled:

```
# cat /sys/module/nvme_core/parameters/multipath
Y
```

2. Verify that the appropriate NVMe-oF settings (such as, model set to NetApp ONTAP Controller and load balancing iopolicy set to round-robin) for the respective ONTAP namespaces correctly reflect on the host:

```
# cat /sys/class/nvme-subsystem/nvme-subsys*/model
NetApp ONTAP Controller
NetApp ONTAP Controller
```

```
# cat /sys/class/nvme-subsystem/nvme-subsys*/iopolicy
round-robin
round-robin
```

3. Verify that the namespaces are created and correctly discovered on the host:

```
# nvme list
```

Example output:

```
Node         SN                   Model
----------------------------------------------------------
/dev/nvme0n1 81CZ5BQuUNfGAAAAAAAB   NetApp ONTAP Controller


Namespace Usage    Format             FW             Rev
-----------------------------------------------------------
1                 21.47 GB / 21.47 GB   4 KiB + 0 B   FFFFFFFF
```

4. Verify that the controller state of each path is live and has the correct ANA status:

NVMe/FC

```
# nvme list-subsys /dev/nvme0n1
```

Example output:

```
nvme-subsys4 - NQN=nqn.1992-
08.com.netapp:sn.8763d311b2ac11ed950ed039ea951c46:subsystem. ub_106
\
+- nvme1 fc traddr=nn-0x20a6d039ea954d17:pn-
0x20a7d039ea954d17,host_traddr=nn-0x200000109b1b95ef:pn-
0x100000109b1b95ef live optimized
+- nvme2 fc traddr=nn-0x20a6d039ea954d17:pn-
0x20a8d039ea954d17,host_traddr=nn-0x200000109b1b95f0:pn-
0x100000109b1b95f0 live optimized
+- nvme3 fc traddr=nn-0x20a6d039ea954d17:pn-
0x20aad039ea954d17,host_traddr=nn-0x200000109b1b95f0:pn-
0x100000109b1b95f0 live non-optimized
+- nvme5 fc traddr=nn-0x20a6d039ea954d17:pn-
0x20a9d039ea954d17,host_traddr=nn-0x200000109b1b95ef:pn-
0x100000109b1b95ef live non-optimized
```

NVME/TCP

```
# nvme list-subsys /dev/nvme1n1
```

Example output:

```
nvme-subsys1 - NQN=nqn.1992- 08.com.netapp:sn.
bbfb4ee8dfb611edbd07d039ea165590:subsystem.rhel_tcp_95
+- nvme1 tcp
traddr=10.10.10.122,trsvcid=4420,host_traddr=10.10.10.47,src_addr=10
.10.10.47 live
+- nvme2 tcp
traddr=10.10.10.124,trsvcid=4420,host_traddr=10.10.10.47,src_addr=10
.10.10.47 live
+- nvme3 tcp
traddr=10.10.11.122,trsvcid=4420,host_traddr=10.10.11.47,src_addr=10
.10.11.47 live
+- nvme4 tcp
traddr=10.10.11.124,trsvcid=4420,host_traddr=10.10.11.47,src_addr=10
.10.11.47 live
```

5. Verify that the NetApp plug-in displays the correct values for each ONTAP namespace device:

Column

```
# nvme netapp ontapdevices -o column
```

Example output:

```
Device        Vserver   Namespace Path
----------------------- ------------------------------
/dev/nvme0n1 co_iscsi_tcp_ubuntu /vol/vol1/ns1



NSID       UUID                                   Size
----------------------------------------------------------------
1          79c2c569-b7fa-42d5-b870-d9d6d7e5fa84 21.47GB
```

JSON

```
# nvme netapp ontapdevices -o json
```

Example output

```
{

"ONTAPdevices" : [
{

"Device" : "/dev/nvme0n1",
"Vserver" : "co_iscsi_tcp_ubuntu",
"Namespace_Path" : "/vol/nvmevol1/ns1",
"NSID" : 1,
"UUID" : "79c2c569-b7fa-42d5-b870-d9d6d7e5fa84",
"Size" : "21.47GB",
"LBA_Data_Size" : 4096,
"Namespace_Size" : 5242880
},

]
}
```

Known issues

Theere are no known issues for the NVMe-oF host configuration for Ubuntu 22.04 with ONTAP release.