NVMe/FC Host Configuration for Oracle Linux 7.8 with ONTAP

ONTAP SAN Host

Carl Plumer
August 21, 2020

This PDF was generated from https://docs.netapp.com/us-en/ontap-sanhost/nvme_ol_78.html on September 04, 2020. Always check docs.netapp.com for the latest.

NetApp®

Table of Contents

NVMe/FC Host Configuration for Oracle Linux 7.8 with ONTAP . . . 1
- Supportability . . . 1
- Known limitations . . . 1
- Enabling NVMe on OL 7.8 . . . 1
- Configuring Oracle Linux 7.8 ANA Initiator . . . 2
- Configuring the Broadcom FC Adapter for NVMe/FC . . . 3
- Validating NVMe/FC . . . 5
- Enabling 1MB I/O Size for Broadcom NVMe/FC . . . 7
- LPFC Verbose Logging . . . 7

NVMe/FC Host Configuration for Oracle Linux 7.8 with ONTAP

Supportability

NVMe/FC is supported on ONTAP 9.6 or later for Oracle Linux 7.8. OL 7.8 host can run both NVMe & SCSI traffic through the same fibre channel initiator adapter ports.

- Broadcom LPe32002 (32G)
- Broadcom LPe31002 (32G/16G)

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

- All Interop FC-NVMe burts are tagged with the INTEROP_FCNVME keyword, whereas Oracle Linux 7.8 InterOp burts also have the Oracle Linux_8.1 keyword tagged to them.
- Native NVMe/FC auto-connect scripts are not available in the nvme-cli package. You can use the HBA vendor provided external auto-connect scripts.
- By default, round-robin load balancing is not enabled. You must write a udev rule to enable this functionality. Steps are provided in the section on Enabling NVMe/FC on OL 7.8.
- There is no sanlun support for FC-NVMe. Thus, no LUHU support for FC-NVMe on Oracle Linux 7.8. Instead, one may rely on the ONTAP command output available as part of the NetApp plug-in included in the native nvme-cli.

The NetApp plug-in in the native NVMe-cli is now enhanced to display ONTAP details as well, by utilizing the data in the vendor specific ONTAP log page. In addition, the Oracle Linux 7.8 initiator host can serve both FC-NVMe and FC-SCSI traffic through the same initiator adapter ports. For FC-SCSI, you can optionally configure dm-multipath as usual for SCSI LUNs resulting in mpath devices, whereas NVMe multipath may be used to configure FC-NVMe multipath namespace devices (for example, `/dev/nvmeXnY`) on the Oracle Linux 7.8 initiator host.

Enabling NVMe on OL 7.8

1. Ensure the default Oracle Linux FCP kernel is installed.
2. Reboot the host and verify that it boots into specified OL 7.8 kernel.

```
# uname -r
4.14.35-1902.9.2.el7uek
```

3. Upgrade to the nvme-cli-1.8.1-3.el7 package.

```
# rpm -qa|grep nvme-cli
nvme-cli-1.8.1-3.el7.x86_64
```

4. Add the string below as a separate udev rule at `/lib/udev/rules.d/71-nvme-iopolicy-netapp-ONTAP.rules`. This enables round-robin load balancing for NVMe multipath.

```
# Enable round-robin for NetApp ONTAP
ACTION==□add□, SUBSYSTEM==□nvme-subsystem□, ATTR{model}==□NetApp ONTAP Controller□,
ATTR{iopolicy}=□round-robin
```

5. On the OL 7.8 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:75953f3b-77fe-4e03-bf3c-09d5a156fbcd
```

```
*> vserver nvme subsystem host show -vserver vs_nvme_10
Vserver Subsystem Host NQN
------- --------- ------------------------------------ -----------
ol_157_nvme_ss_10_0
nqn.2014-08.org.nvmexpress:uuid:75953f3b-77fe-4e03-bf3c-09d5a156fbcd
```

If the `hostnqn` strings do not match, you should 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.

6. Reboot the host.

Configuring Oracle Linux 7.8 ANA Initiator

1. First ensure you are running with the recommended kernel & nvme-cli versions:

```
# uname -r
5.4.17-2006.5.el7uek.x86_64
```

```
# rpm -qa|grep nvme-cli
nvme-cli-1.8.1-3.el7.x86_64
```

The native `nvme-cli` on Oracle Linux 7.8 does not include a udev rule to enable round-robin load balancing for NVMe Multipath (nor the auto-connect scripts). So you need to separately install this on the Oracle Linux 7.8 host.

2. Add the following udev rule at /lib/udev/rules.d:

```
 # cat /lib/udev/rules.d/71-nvme-iopolicy-netapp-ONTAP.rules
 # Enable round-robin for NetApp ONTAP
 ACTION=="add", SUBSYSTEM=="nvme-subsystem", ATTR{model}=="NetApp ONTAP Controller",
ATTR{iopolicy}="round-robin"
```

3. Check the hostnqn string at /etc/nvme/hostnqn on the Oracle Linux 7.8 host and ensure that it properly matches with the hostnqn string for the corresponding subsystem on the ONTAP array. For example:

```
# cat /etc/nvme/hostnqn
 nqn.2014-08.org.nvmexpress:uuid:75953f3b-77fe-4e03-bf3c-09d5a156fbcd
```

```
*> vserver nvme subsystem host show -vserver vs_nvme_10
Vserver  Subsystem   Host     NQN
-------  --------- --------- ---------
Oracle Linux_141_nvme_ss_10_0
nqn.2014-08.org.nvmexpress:uuid:75953f3b-77fe-4e03-bf3c-09d5a156fbcd
```

Configuring the Broadcom FC Adapter for NVMe/FC

1. Verify that you are using the supported adapter. For the most current list of supported adapters see the NetApp Interoperability Matrix.

```
# 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
```

The newer lpfc drivers (both inbox & outbox) already have lpfc_enable_fc4_type default set to 3 i.e. one no longer needs to set this explicitly in the /etc/modprobe.d/lpfc.conf anymore, and recreate the initramfs.

1. NVMe support in lpfc is already enabled by default:

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

2. Next, install the recommended lpfc auto-connect scripts:

```
# rpm -ivh nvmefc-connect-12.4.65.0-1.noarch.rpm
```

3. After running the above commands, reboot the host and verify that the recommended lpfc outbox driver & auto-connect scripts are installed after bootup:

```
# cat /sys/module/lpfc/version
0:12.6.0.3
# rpm -qa | grep nvmefc
nvmefc-connect-12.4.65.0-1.noarch
```

4. Verify that the initiator ports are up and running:

```
# cat /sys/class/fc_host/host*/port_name
0x10000090fae0ec61
0x10000090fae0ec62

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

4. Verify that the FC-NVMe initiator ports are enabled and able to see the target ports, and all are up & running. In this example, only 1 initiator port is enabled and connected with two target LIFs as seen in the below output:

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

NVME Initiator Enabled
XRI Dist lpfc0 Total 6144 NVME 2947 SCSI 2947 ELS 250
NVME LPORT lpfc0 WWPN x10000090fae0ec61 WWNN x20000090fae0ec61 DID x012000 ONLINE
NVME RPORT WWPN x202d00a098c80f09 WWNN x202c00a098c80f09 DID x010201 TARGET DISCSRVC
ONLINE
NVME RPORT WWPN x203100a098c80f09 WWNN x202c00a098c80f09 DID x010601 TARGET DISCSRVC
ONLINE

NVME Statistics
LS: Xmt 000000000e Cmpl 000000000e Abort 00000000
LS XMIT: Err 00000000 CMPL: xb 00000000 Err 00000000
Total FCP Cmpl 000000000001a680 Issue 000000000001a682 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 lpfc1 Total 6144 NVME 2947 SCSI 2947 ELS 250
NVME LPORT lpfc1 WWPN x10000090fae0ec62 WWNN x20000090fae0ec62 DID x012400 ONLINE
```

Validating NVMe/FC

1. Verify the following NVMe/FC settings.

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

```
# 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
```

2. Verify that the namespaces are created.

```
# nvme list
Node SN Model Namespace Usage Format FW Rev
---------------- -------------------- ----------------------
/dev/nvme0n1 80BADBKnB/JvAAAAAAAC NetApp ONTAP Controller 1 53.69 GB / 53.69 GB 4 KiB
+ 0 B FFFFFFFF
```

3. Verify the status of the ANA paths.

```
# nvme list-subsys/dev/nvme0n1
Nvme-subsysf0 □ NQN=nqn.1992-
08.com.netapp:sn.341541339b9511e8a9b500a098c80f09:subsystem.ol_157_nvme_ss_10_0
\
+- nvme0 fc traddr=nn-0x202c00a098c80f09:pn-0x202d00a098c80f09 host_traddr=nn-
0x20000090fae0ec61:pn-0x10000090fae0ec61 live optimized
+- nvme1 fc traddr=nn-0x207300a098dfdd91:pn-0x207600a098dfdd91 host_traddr=nn-
0x200000109b1c1204:pn-0x100000109b1c1204 live inaccessible
+- nvme2 fc traddr=nn-0x207300a098dfdd91:pn-0x207500a098dfdd91 host_traddr=nn-
0x200000109b1c1205:pn-0x100000109b1c1205 live optimized
+- nvme3 fc traddr=nn-0x207300a098dfdd91:pn-0x207700a098dfdd91 host traddr=nn-
0x200000109b1c1205:pn-0x100000109b1c1205 live inaccessible
```

4. Verify the NetApp plug-in for ONTAP devices.

```
# nvme netapp ontapdevices -o column
Device   Vserver  Namespace Path             NSID   UUID   Size
-------  -------- -------------------------  ------ ----- -----
/dev/nvme0n1   vs_nvme_10       /vol/rhel_141_vol_10_0/ol_157_ns_10_0    1
55baf453-f629-4a18-9364-b6aee3f50dad   53.69GB

# nvme netapp ontapdevices -o json
{
   "ONTAPdevices" : [
   {
        Device" : "/dev/nvme0n1",
        "Vserver" : "vs_nvme_10",
        "Namespace_Path" : "/vol/rhel_141_vol_10_0/ol_157_ns_10_0",
         "NSID" : 1,
         "UUID" : "55baf453-f629-4a18-9364-b6aee3f50dad",
         "Size" : "53.69GB",
         "LBA_Data_Size" : 4096,
         "Namespace_Size" : 13107200
    }
]
```

Enabling 1MB I/O Size for Broadcom NVMe/FC

The lpfc_sg_seg_cnt parameter must be set to 256 in order for the host to issue 1MB size I/O.

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
```

LPFC Verbose Logging

1. You can set the lpfc_log_verbose driver setting to any of the following values to log NVMe/FC events.

```
#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. */
```

2. After setting any of these values, run `dracut-f` and reboot host.
3. After rebooting, verify the settings.

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

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

Copyright Information

Copyright © 2020 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.

RESTRICTED RIGHTS LEGEND: Use, duplication, or disclosure by the government is subject to restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause at DFARS 252.277-7103 (October 1988) and FAR 52-227-19 (June 1987).

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.