Technical Support Bulletin • UPDATE

Mellanox HCA Queue Pairs Become Stuck in Certain EXAScaler Configurations

ALERT! Certain versions of Mellanox InfiniBand HCA firmware are **incompatible** with EXAScaler 5.x under SFA OS 11.9.x and later. **They must be manually downgraded.**

Issue Summary

Certain versions of Mellanox InfiniBand and Ethernet HCA firmware experience connection timeouts when installed with EXAScaler 5.x under SFA OS 11.9.x or higher. This occurs when an RDMA queue pair becomes stuck in an error state shown by ”Timed out tx: active_txs” and “Timed out RDMA” messages in /var/log/lustre.log. For example:

```
boss9 kernel: LNetError: 3500:0:(o2iblnd_cb.c:3443:kiblnd_check_txs_locked()) Timed out tx: active_txs, 3 seconds
boss9 kernel: LNetError: 3500:0:(o2iblnd_cb.c:3518:kiblnd_check_conns()) Timed out RDMA with 10.30.40.195@o2ib (45): c:
32, oc: 0, rc: 32
```

A queue pair (QP) consists of an RDMA send queue and receive queue with their resources, such as memory, on the HCA. When the QP becomes stuck in an error state, network packets aren’t processed and the connection served by the QP times out. The client normally disconnects and reconnects automatically to a new QP, but the stuck QP is never terminated and resources such as memory are not released. Over time, the number of active QPs will grow to noticeably exceed the number of all EXAScaler clients and servers in the network. The resulting memory leak will, over time, cause an out-of-memory condition on the HCA. OOM-killer will then be invoked and a message similar to the following will be posted to /var/log/lustre.log:

```
boss9 kernel: ll_ost_io01_022 invoked oom-killer: gfp_mask=0x200d2, order=0, oom_score_adj=0
```

The root cause of this issue is an incompatibility between newer Mellanox HCA firmware versions packaged with SFA OS 11.9.0 or later and the Mellanox MOFED 4.9 LTS software used by EXAScaler 5.x.

Affected Products

The stuck QP issue has been encountered with the following versions of Mellanox HCA firmware, where asterisks are wild cards for any value:

- * .30. *
- * .31. *
- * .32. *

These HCA firmware versions are *incompatible with EXAScaler 5.x running under SFA OS 11.9.x or higher.* Both block and embedded SFA storage configurations are affected.

The HCA firmware version currently installed on the SFA storage can be found by running the following command at the SFA OS command line:

```
application show ioc
```

The IOC (I/O controller) firmware version is shown under the FW Version heading of the returned report.

```
*****************************
*            IOC(s)         *
*****************************

      |Ctlr|AP|                      IOC                              |
Index |    |  | bus | dev | fn | part num | FW version    | port type |
-------------------------------------------------------------------------
05888   0    0   17    00    00  MT4119     16.32.1010       IB
05889   0    0   17    00    01  MT4119     16.32.1010       IB
01024   0    0   04    00    00  I350       1.63             EN
38656   1    0   17    00    00  MT4119     16.32.1010       IB
38657   1    0   17    00    01  MT4119     16.32.1010       IB
33792   1    0   04    00    00  I350       1.63             EN

Total IOCs: 6
```

Workaround

To prevent the QP from becoming stuck, affected HCA firmware *must be downgraded to a compatible version* on all HCAs residing on all SFA storage systems and any external EXAScaler storage servers in the EXAscaler cluster. HCA firmware versions tested and known to work with EXAScaler 5.x are listed in the table below.

Mellanox HCA Model	Qualified Firmware Version
ConnectX-6	20.28.2006
ConnectX-5	16.28.2006
ConnectX-4	12.28.2006
ConnectX-3	2.42.5000

ALERT! This downgrade must be performed manually. **Contact DDN Support for assistance.**

Resolution

EXAScaler 6.x incorporates a later version of MOFED than EXAScaler 5.x and mitigates the issue at least partially. DDN Engineering is actively investigating whether further software changes are required for a complete resolution.

Contacting DDN Support

Please contact DDN Technical Support at any time if you have questions or need assistance. Support can be reached online, by email, or by phone as listed below.

Web

DDN Community Support Portal https://community.ddn.com/login
Portal Assistance webportal.support@ddn.com

Email

Support Email	support@ddn.com

Telephone

DDN Support Worldwide Directory	https://www.ddn.com/support/global-services-overview/

Bulletins & Notices

Support Bulletins	http://www.ddn.com/support/technical-support-bulletins
End-of-Life Notices	http://www.ddn.com/support/end-of-life-notices
Release Notes	https://community.ddn.com/login
Subscription Requests	support-tsb@ddn.com