Connect. Accelerate. Outperform.™

Mellanox OFED for Linux Release Notes

Rev 3.1-1.0.3

www.mellanox.com

Mellanox Technologies
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403

Table of Contents

List Of Tables

Release Update History

Table 1 - Release Update History

Release	Date	Description
Rev 3.1-1.0.3	December 10, 2015	• Updated Table 4, "Supported Platforms and Operating Systems", added RHEL/CentOS 7.2. • Added known issue 22 in Table 23, "IPoIB Known Issues," on page 20
	October 01, 2015	This is the initial release of this MLNX_OFED release.

1 Overview

These are the release notes of Mellanox OFED for Linux Driver, Rev 3.1-1.0.3. Mellanox OFED is a single Virtual Protocol Interconnect (VPI) software stack and operates across all Mellanox network adapter solutions supporting the following uplinks to servers:

Table 2 - Supported Uplinks to Servers

Uplink/HCAs	Uplink Speed
Connect-IB®	• InfiniBand: SDR, QDR, FDR, EDR
ConnectX®-4	• InfiniBand: SDR, QDR, FDR, EDR • Ethernet: 10GigE, 25GigE, 40GigE, 50GigE and 100GigE
ConnectX®-4 Lx	• Ethernet: 10GigE, 25GigE, 40GigE, and 50GigE
ConnectX®-3/ConnectX®-3 Pro	• InfiniBand: SDR, QDR, FDR10, FDR • Ethernet: 10GigE, 40GigE and 56GigE[a]
ConnectX®-2	• InfiniBand: SDR, DDR • Ethernet: 10GigE, 20GigE
PCI Express 2.0	2.5 or 5.0 GT/s
PCI Express 3.0	8 GT/s

a. 56 GbE is a Mellanox propriety link speed and can be achieved while connecting a Mellanox adapter cards to Mellanox SX10XX switch series or connecting a Mellanox adapter card to another Mellanox adapter card.

1.1 Content of Mellanox OFED for Linux

Mellanox OFED for Linux software contains the following components:

Table 3 - Mellanox OFED for Linux Software Components

Components	Description
OpenFabrics core and ULPs	• InfiniBand and Ethernet HCA drivers (mlx4, mlx5) • core • Upper Layer Protocols: IPoIB, SRP, iSER and iSER Initiator and Target
OpenFabrics utilities	• OpenSM: IB Subnet Manager with Mellanox proprietary Adaptive Routing • Diagnostic tools • Performance tests • SSA (SLES12): libopensmssa plugin for OpenSM, ibssa, ibacm
MPI	• OSU MPI (mvapich2-2.0) stack supporting the InfiniBand interface • Open MPI stack 1.6.5 and later supporting the InfiniBand interface • MPI benchmark tests (OSU benchmarks, Intel MPI benchmarks, Presta)
PGAS	• HPC-X OpenSHMEM v2.2 supporting InfiniBand, MXM and FCA • HPC-X UPC v2.2 supporting InfiniBand, MXM and FCA

Table 3 - Mellanox OFED for Linux Software Components

Components	Description
HPC Acceleration packages	• Mellanox MXM v3.0 (p2p transport library acceleration over InfiniBand) • Mellanox FCA v2.5 (MPI/PGAS collective operations acceleration library over InfiniBand) • KNEM, Linux kernel module enabling high-performance intra-node MPI/PGAS communication for large messages
Extra packages	• ibutils2 • ibdump • MFT
Sources of all software modules (under conditions mentioned in the modules' LICENSE files) except for MFT, OpenSM plugins, ibutils2, and ibdump	
HCAs	ConnectX-3 EN driver Rev 3.1-1.0.3 ConnectX-4 EN driver Rev 3.1-1.0.3
Documentation	

1.2 Supported Platforms and Operating Systems

The following are the supported OSs in MLNX_OFED Rev 3.1-1.0.3:

Table 4 - Supported Platforms and Operating Systems

Operating System	Platform
RHEL/CentOS 6.5	x86_64
RHEL/CentOS 6.6	x86_64
RHEL/CentOS 6.7	x86_64/PPC64 (Power 7)
RHEL/CentOS 7.0	x86_64PPC64 (Power 7)
RHEL/CentOS 7.1	x86_64/PPC64 (Power 7)/PPC64le (Power 8)/ARM64 (ARM is at beta level)
RHEL/CentOS 7.2	x86_64
SLES11 SP1	x86_64
SLES11 SP2	x86_64
SLES11 SP3	x86_64/PPC64 (Power 7)
SLES11 SP4	x86_64/PPC64 (Power 7)
SLES12	x86_64/PPC64le (Power 8)
OEL 6.3	x86_64
OEL 6.4	x86_64
OEL 6.5	x86_64
OEL 6.6	x86_64
OEL 6.7	x86_64
OEL 7.0	x86_64
OEL 7.1	x86_64
Fedora 19	x86_64/PPC64 (Power 7)
Fedora 20	x86_64
Fedora 21	x86_64/PPC64 (Power 7)
Ubuntu 12.04	x86_64

Table 4 - Supported Platforms and Operating Systems

Operating System	Platform
Ubuntu 14.04	x86_64/PPC64le (Power 8)
Ubuntu 14.10	x86_64/PPC64le (Power 8)
Ubuntu 15.04	x86_64/PPC64le (Power 8)
Debian 6.0.10	x86_64
Debian 7.6	x86_64
Debian 8.0	x86_64
Debian 8.1	x86_64
Windriver 6.0	x86_64
kernel 3.10[a] - 4.1	

a. This kernel is supported only when using the Operating Systems stated in the table above.

For RPM based Distributions, if you wish to install OFED on a different kernel, you need to create a new ISO image, using mlnx_add_kernel_support.sh script.
See the MLNX_OFED User Guide for instructions.

Upgrading MLNX_OFED on your cluster requires upgrading all of its nodes to the newest version as well.

1.2.1 Supported Hypervisors

The following are the supported Hypervisors in MLNX_OFED Rev 3.1-1.0.3:

- KVM:
  - RedHat 6.6, 6.7, 7.1
  - Ubuntu 14.10, 15.04
  - Sles11SP4, Sles12
  - Debian 6.0.10

1.2.2 Supported Non-Linux Virtual Machines

The following are the supported Non-Linux (InfiniBand only) Virtual Machines in MLNX_OFED Rev 3.1-1.0.3:

- Windows Server 2012 R2
- Windows Server 20012
- Windows Server 2008 R2

1.3 Hardware and Software Requirements

The following are the hardware and software requirements of MLNX_OFED Rev 3.1-1.0.3.

- Linux operating system
- Administrator privileges on your machine(s)

- Disk Space: 1GB

For the OFED Distribution to compile on your machine, some software packages of your operating system (OS) distribution are required.

To install the additional packages, run the following commands per OS:

Table 5 - Additional Software Packages

Operating System	Required Packages Installation Command
RHEL/OEL/Fedora	yum install perl pciutils python gcc-gfortran libxml2-python tcsh libnl.i686 libnl expat glib2 tcl libstdc++ bc tk gtk2 atk cairo numactl pkgconfig
XenServer	yum install perl pciutils python libxml2-python libnl expat glib2 tcl bc libstdc++ tk pkgconfig
SLES 10 SP3	zypper install pkgconfig pciutils python libxml2-python libnl lsof expat glib2 tcl libstdc++ bc tk
SLES 11 SP2	zypper install perl pciutils python libnl-32bit libxml2-python tcsh libnl libstdc++46 expat glib2 tcl bc tklibcurl4 gtk2 atk cairo pkg-config
SLES 11 SP3	zypper install perl pciutils python libnl-32bit libxml2-python tcsh libstdc++43 libnl expat glib2 tcl bc tk libcurl4 gtk2 atk cairo pkg-config
SLES 12	zypper install pkg-config expat libstdc++6 libglib-2_0-0 libgtk-2_0-0 tcl libcairo2 tcsh python bc pciutils libatk-1_0-0 tk python-libxml2 lsof libnl1
Ubuntu/Debian	apt-get install perl dpkg autotools-dev autoconf libtool automake1.10 automake m4 dkms debhelper tcl tcl8.4 chrpath swig graphviz tcl-dev tcl8.4-dev tk-dev tk8.4-dev bison flex dpatch zlib1g-dev curl libcurl4-gnutls-dev python-libxml2 libvirt-bin libvirt0 libnl-dev libglib2.0-dev libgfortran3 automake m4 pkg-config libnuma logrotate
Debian 8	apt-get install libnl-3-200 automake debhelper curl dkms logrotate libglib2.0-0 python-libxml2 graphviz tk tcl libvirt-bin coreutils pkg-config autotools-dev flex autoconf pciutils quilt module-init-tools libvirt0 libstdc++6 dpkg libgfortran3 procps lsof libltdl-dev gcc dpatch chrpath grep m4 gfortran bison libnl-route-3-200 swig perl make

1.4 Supported HCAs Firmware Versions

MLNX_OFED Rev 3.1-1.0.3 supports the following Mellanox network adapter cards firmware versions:

Table 6 - Supported HCAs Firmware Versions

HCA	Recommended Firmware Rev.	Additional Firmware Rev. Supported
Connect-IB®	10.12.1100	10.12.0780
ConnectX®-4 Lx	14.12.1100	14.12.0780
ConnectX®-4	12.12.1100	12.12.0780
ConnectX®-3 Pro	2.35.5000	2.34.5000

Table 6 - Supported HCAs Firmware Versions

HCA	Recommended Firmware Rev.	Additional Firmware Rev. Supported
ConnectX®-3	2.35.5000	2.34.5000
ConnectX®-2	2.9.1000	2.9.1000

For official firmware versions please see:
http://www.mellanox.com/content/pages.php?pg=firmware_download

1.5 Compatibility

MLNX_OFED Rev 3.1-1.0.3 is compatible with the following:

Table 7 - MLNX_OFED Rev 3.1-1.0.3 Compatibility Matrix

Mellanox Product	Description/Version
MLNX-OS®	MSX6036 w/w MLNX-OS® version 3.3.4304[a]
Grid Director™	4036 w/w Grid Director™ version 3.9.1-985
FabricIT™ EFM	IS5035 w/w FabricIT EFM version 1.1.3000
FabricIT™ BXM	MBX5020 w/w FabricIT BXM version 2.1.2000
Unified Fabric Manager (UFM®)	v4.8
MXM	v3.2
HPC-X UPC	v2.18.0
HPC-X OpenSHMEM	v1.8.3
FCA	v2.5 and v3.1
HPC-X MPI	v1.8.3
MVAPICH	v2.0

a. MLNX_OFED Rev 3.1-1.0.3 was tested with this switch however, additional switches might be supported as well.

1.6 RoCE Modes Matrix

The following is RoCE modes matrix:

Table 8 - RoCE Modes Matrix

Software Stack / Inbox Distribution	RoCEv1 (Layer 2) Supported as of Version	RoCEv2 (Layer 3) Supported as of Version	RoCEv1 & RoCEv2 (Layer 3) Supported as of Version
MLNX_OFED	2.1-x.x.x	2.3-x.x.x	3.0-x.x.x
Kernel.org	3.14	-	-
RHEL	6.6; 7.0	-	-
SLES	12	-	-
Ubuntu	14.04	-	-

2 Changes and New Features in Rev 3.1-1.0.3

Table 9 - Changes in v3.1-1.0.3

Category	Description
User Access Region (UAR)	Allows the ConnectX-3 driver to operate on PPC machines without requiring a change to the MMIO area size.
CQE Compression	Saves PCIe bandwidth by compressing a few CQEs into a smaller amount of bytes on PCIe
Bug fixes	See Section 4, “Bug Fixes History”, on page 36

For additional information on the new features, please refer to the MLNX_OFED User Manual.

3 Known Issues

The following is a list of general limitations and known issues of the various components of this Mellanox OFED for Linux release.

3.1 Driver Installation/Loading/Unloading/Start Known Issues

3.1.1 Installation Known Issues

Table 10 - Installation Known Issues

Index	Description	Workaround
1.	When upgrading from an earlier Mellanox OFED version, the installation script does not stop the earlier version prior to uninstalling it.	Stop the old OFED stack (`/etc/init.d/openibd stop`) before upgrading to this new version.
2.	Upgrading from the previous OFED installation to this release, does not unload the kernel module ipoib_helper.	Reboot after installing the driver.
3.	Installation using Yum does not update HCA firmware.	See “Updating Firmware After Installation” in OFED User Manual
4.	`"--total-vfs <0-63>"` installation parameter is no longer supported	Use `'--enable-sriov'` installation parameter to burn firmware with SR-IOV support. The number of virtual functions (VFs) will be set to 16. For further information, please refer to the User Manual.
5.	When using bonding on Ubuntu OS, the "ifenslave" package must be installed.	-
6.	On PPC systems, the `ib_srp` module is not installed by default since it breaks the `ibmvscsi` module.	If your system does not require the `ibmvscsi` module, run the mlnxofedinstall script with the `"--with-srp"` flag.

3.1.2 Driver Unload Known Issues

Table 11 - Driver Unload Known Issues

Index	Description	Workaround

1.	`"openibd stop"` can sometime fail with the error:
	`Unloading ib_cm [FAILED]`
	`ERROR: Module ib_cm is in use by ib_ipoib`
		Re-run `"openibd stop"`

3.1.3 Driver Start Known Issues

Table 12 - Driver Start Known Issues

Index	Description	Workaround
1.	"Out-of-memory" issues may rise during drivers load depending on the values of the driver module parameters set (e.g. log_num_cq).	-
2.	When reloading/starting the driver using the `/etc/init.d/openibd` the following messages are displayed if there is a third party RPM or driver installed: "Module mlx4_core does not belong to MLNX_OFED" or "Module mlx4_core belong to <rpm name> which is not a part of MLNX_OFED"	Remove the third party RPM/non MLNX_OFED drivers directory, run: `"depmod"` and then rerun `"/etc/init.d/openibd restart"`
3.	Occasionally, when trying to repetitively reload the NES hardware driver on SLES11 SP2, a soft lockups occurs that required reboot.	-

4.	When downgrading from MLNX_OFED 3.0-x.x.x, driver reload might fail with the following errors in dmeg:
	`[166271.886407] compat: exports duplicate symbol __ethtool_get_settings (owned by mlx_compat)`
		The issues will be resolved automatically after system reboot or by invoking the following commands:
		`rmmod mlx_compat`
		`depmod -a`
		`/etc/init.d/openibd restart`

5.	In ConnectX-2, (when the debug_level module parameter for module mlx4_core is non-zero), if the driver load succeeds, the message below is presented: `"mlx4_core 0000:0d:00.0: command SET_PORT (0xc) failed: in_param=0x120064000, in_mod=0x2, op_mod=0x0, fw status = 0x40"` This message is simply part of the learning process for setting the maximum port VLs compatible with a 4K port mtu, and should be ignored.	-
6.	"openibd start" unloads kernel modules that were loaded from initrd/initramfs upon boot. This affects only kernel modules which come with MLNX_OFED and are included in initrd/initramfs.	-

7.	If a Lustre storage is used, it must be fully unloaded before restarting the driver or rebooting the machine, otherwise machine might get stuck/panic.	1. Unmount any mounted Lustre storages:
		`# umount<lustre_mount_point>`
		2. Unload all Lustre modules:
		`# lustre_rmmod`

Table 12 - Driver Start Known Issues (Continued)

Index	Description	Workaround

8.	Driver unload fails with the following error message:
	`Unloading rdma_cm [FAILED]`
	`rmmod: ERROR: Module rdma_cm is in use by: xprtrdma`
		Make sure that there are no mount points over NFS/RDMA prior to unloading the driver and run:
		`# modprobe -r xprtrdma`
		In case that the xprtrdma module keeps getting loaded automatically even though it is not used, add a pre-stop hook for the openibd service script to always unload it.
		Create an executable file `"/etc/infiniband/pre-stop-hook.sh"` with the following content:
		`#!/bin/bash`
		`modprobe -r xprtrdma`

9.	When loading or unloading the driver on HP Proliant systems, you may see log messages like:
	`dmar: DMAR:[DMA Write] Request device [07:00.0] fault addr 3df7f000`
	`DMAR:[fault reason 05] PTE Write access is not set`
	This is a known issue with ProLiant systems (see their support notice for Emulex adapters: http://h20564.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c04446026&lang=en-us&cc=us
	)The messages are harmless, and may be ignored.
		If you are *not* running SR-IOV on your system, you may eliminate these messages by removing the term `"intel_iommu=on"` from the boot line in file `/boot/grub/menu.lst`.
		For SR-IOV systems, this term must remain, you can ignore the log messages.

3.1.4 System Time Known Issues

Table 13 - System Time Known Issues

Index	Description	Workaround
1.	Loading the driver using the openibd script when no InfiniBand vendor module is selected (for example mlx4_ib), may cause the execution of the /sbin/start_udev' script. In RedHat 6.x and OEL6.x this may change the local system time.	-

3.1.5 UEFI Secure Boot Known Issues

Table 14 - UEFI Secure Boot Known Issues

Index	Description	Workaround
1.	On RHEL7 and SLES12, the following error is displayed in dmesg if the Mellanox's x.509 Public Key is not added to the system: `[4671958.383506] Request for unknown module key 'Mellanox Technologies signing key: 61feb074fc7292f958419386ffdd9d5-ca999e403' err -11` This error can be safely ignored as long as Secure Boot is disabled on the system.	For further information, please refer to the User Manual section "Enrolling Mellanox's x.509 Public Key On your Systems".
2	Ubuntu12 requires update of user space open-iscsi to v2.0.873	-
3	The initiator does not respect interface parameter while logging in.	Configure each interface on a different subnet.

3.2 Performance Known Issues

Table 15 - Performance Known Issues

Index	Description	Workaround
1.	On machines with irqbalancer daemon turned off, the default InfiniBand interrupts will be routed to a single core which may cause overload and software/hardware lockups.	Execute the following script as root: `set_irq_affinity.sh <interface or IB device> [2nd interface or IB device]`
2.	Out of the box throughput performance in Ubuntu14.04 is not optimal and may achieve results below the line rate in 40GE link speed.	For additional performance tuning, please refer to Performance Tuning Guide.
3.	UDP receiver throughput may be lower then expected, when running over mlx4_en Ethernet driver. This is caused by the adaptive interrupt moderation routine, which sets high values of interrupt coalescing, causing the driver to process large number of packets in the same interrupt, leading UDP to drop packets due to overflow in its buffers.	Disable adaptive interrupt moderation and set lower values for the interrupt coalescing manually. `ethtool -C <eth>X adaptive-rx off rx-usecs 64 rx-frames 24` Values above may need tuning, depending the system, configuration and link speed.
4.	Performance degradation might occur when bonding Ethernet interfaces on Centos 6.5	-

3.3 HCAs Known Issues

3.3.1 ConnectX®-3 (mlx4 Driver) Known Issues

Table 16 - ConnectX®-3 (mlx4 Driver) Known Issues

Index	Description	Workaround
1.	Using RDMA READ with a higher value than 30 SGEs in the WR might lead to "local length error".	Do not set the value of SGEs higher than 30 when RDMA READ is used.

3.3.2 ConnectX®-4 (mlx5 Driver) Known Issues

Table 17 - ConnectX-4 (mlx5 Driver) Known Issues

Index	Description	Workaround
1.	Atomic Operations in Connect-IB® are fully supported on big-endian machines (e.g. PPC). Their support is limited on little-endian machines (e.g. x86)	-
2.	EEH events that arrive while the mlx5 driver is loading may cause the driver to hang.	-
3.	The mlx5 driver can handle up to 5 EEH events per hour.	If more events are received, cold reboot the machine.
4.	When working with Connect-IB® firmware v10.10.5054, the following message would appear in driver start. `command failed, status bad system state(0x4), syndrome 0x408b33` The message can be safely ignored.	Upgrade Connect-IB firmware to the latest available version.
5.	Changing the link speed is not supported in Ethernet driver when connected to a ConnectX-4 card.	-
6.	Bonding "active-backup" mode does not function properly.	-
7.	Rate, speed and width using IB sysfs/tools are available in RoCE mode in ConnectX-4 only after port physical speed configuration is done.	-
8.	Since MLNX_OFED's openibd does not unload modules while OpenSM is running, removing `mlx5_core` manually while OpenSM is running, may cause it to be out of sync when probed again.	Restart OpenSM
9.	ConnectX-4 port GIDs table shows a duplicated RoCE v2 default GID.	-

3.4 Ethernet Network

3.4.1 Ethernet Known Issues

Ethernet Know Issues are applicable to ConnectX-3/ConnectX-3 Pro only.

Table 18 - Ethernet Known Issues

Index	Description	Workaround
1.	When creating more than 125 VLANs and SR-IOV mode is enabled, a kernel warning message will be printed indicating that the native VLAN is created but will not work with RoCE traffic. kernel warning: `mlx4_core 0000:07:00.0: vhcr command ALLOC_RES (0xf00) slave:0 in_param 0x7e in_mod=0x107, op_mod=0x1 failed with error:0, status -28`	-
2.	Kernel panic might occur during FIO splice in kernels before 2.6.34-rc4.	Use kernel v2.6.34-rc4 which provides the following solution: `baff42a net: Fix oops from tcp_collapse() when using splice()`
3.	In PPC systems when QoS is enabled a harmless Kernel DMA mapping error messages might appear in kernel log (IOMMU related issue).	-
4.	Transmit timeout might occur on RH6.3 as a result of lost interrupt (OS issue). In this case, the following message will be shown in dmesg: `do_IRQ: 0.203 No irq handler for vector (irq -1)`	-
5.	Mixing ETS and strict QoS policies for TCs in 40GbE ports may cause inaccurate results in bandwidth division among TCs.	-
6.	Creating a VLAN with user priority >= 4 on ConnectX®-2 HCA is not supported.	-
7.	Affinity hints are not supported in Xen Hypervisor (an irqblancer issue). This causes a non-optimal IRQ affinity.	To overcome this issues, run: `set_irq_affinity.sh eth<x>`
8.	Reboot might hang in SR-IOV when using the `"probe_vf"` parameter with many Virtual Functions. The following message is logged in the kernel log: `"waiting for eth to become free. Usage count =1"`	-
9.	In ConnectX®-2, RoCE UD QP does not include VLAN tags in the Ethernet header	

Table 18 - Ethernet Known Issues (Continued)

Index	Description	Workaround
10.	VXLAN may not be functional when configured over Linux bridge in RH7.0 or Ubuntu14.04. The issue is within the bridge modules in those kernels. In Vanilla kernels above 3.16 issues is fixed.	-
11.	In RH6.4, ping may not work over VLANs that are configured over Linux bridge when the bridge has a mlx4_en interface attached to it.	-
12.	The interfaces LRO needs to be set to "OFF" manually when there is a bond configured on Mellanox interfaces with a Bridge over that bond.	Run: `ethtool -K ethX lro off`
13.	On SLES12, the bonding interface over Mellanox Ethernet slave interfaces does not get IP address after reboot.	1. Set "STARTMODE=hotplug" in the bonding slave's ifcfg files. More details can be found in the SUSE documentations page: https://www.suse.com/documentation/sles-12/book_sle_admin/?page=/documentation/sles-12/book_sle_admin/data/sec_bond.html 2. Enable the “nanny” service to support hotplugging: Open the `"/etc/wicked/common.xml"` file. Change: `"<use-nanny>false</use-nanny>"` to `"<use-nanny>true</use-nanny>"` 3. Run: `# systemctl restart wickedd.service wicked`
14.	ethtool -x command does not function in SLES OS.	-
15.	Ethertype proto 0x806 not supported by ethtool	-
16.	ETS configuration is not supported in the following kernels: • 3.7 • 3.8 • 3.9 • 3.10 • 3.2.37-94_fbk17_01925_g8e3b329 • 3.14 • 3.2.55-106_fbk22_00877_g6902630 • 3.2.28-76_fbk14_00230_g3c40d9e	
17.	ETS is not supported in kernels that do not have MQPRIO as QDISC_KIND option in the tc tool.	-
18.	When NC-SI is ON, the port’s MTU cannot be set to lower than 1500.	-
19.	GRO is not functional when using VXLAN in ConnectX-3 adapter cards.	-
20.	ethtool -X: The driver supports only the `'equal'` mode and cannot be set by using weight flags.	-
21.	Q-in-Q infrastructure in the kernel is supported only from kernel version 3.10 and up.	-

Table 18 - Ethernet Known Issues (Continued)

Index	Description	Workaround
22.	When SLES11 SP4 is used as a DHCP client over ConnectX-3 or ConnectX-3 adapters, it might fail to get an IP from the DHCP server.	-

3.4.2 Port Type Management Known Issues

Table 19 - Port Type Management Known Issues

Index	Description	Workaround
1.	OpenSM must be stopped prior to changing the port protocol from InfiniBand to Ethernet.	-
2.	After changing port type using `connectx_port_config` interface ports' names can be changed. For example. ib1 -> ib0 if port1 changed to be Ethernet port and port2 left IB.	Use `udev` rules for persistent naming configuration. For further information, please refer to the User Manual
3.	A working IP connectivity between the RoCE devices is required when creating an address handle or modifying a QP with an address vector.	-
4.	IPv4 multicast over RoCE requires the MGID format to be as follow `::ffff:<Multicast IPv4 Address>`	-
5.	IP routable RoCE does not support Multicast Listener Discovery (MLD) therefore, multicast traffic over IPv6 may not work as expected.	-
6	DIF: When running IO over FS over DM during unstable ports, block layer BIOS merges may cause false DIF error.	-
7	connectx_port_config configurations is not saved after unbind/bind	Re-run "connectx_port_config"

3.4.3 Flow Steering Known Issues

Table 20 - Flow Steering Known Issues

Index	Description	Workaround
1.	Flow Steering is disabled by default in firmware version < 2.32.5100.	To enable it, set the parameter below as follow: `log_num_mgm_entry_size should set to -1`
2.	IPv4 rule with source IP cannot be created in SLES 11.x OSes.	-
3.	RFS does not support UDP.	-
4.	When working in DMFS:A0 mode and VM/hypervisor is MLNX_OFED 2.3-x.x.x, the second side (hypervisor/VM respectively) should be MLNX_OFED 2.3-x.x.x as well.	-
5.	Setting ARP flow rules through ethtool is not allowed.	-

3.4.4 Quality of Service Known Issues

Table 21 - Quality of Service Known Issues

Index	Description	Workaround
1.	QoS is not supported in XenServer, Debian 6.0 and 6.2 with uek kernel	-
2.	When QoS features are not supported by the kernel, `mlnx_qos` tool may crash.	-
3.	QoS default settings are not returned after configuring QoS.	-

3.4.5 Ethernet Performance Counters Known Issues

Table 22 - Ethernet Performance Counters Known Issues

Index	Description	Workaround
1.	In ConnectX®-3, in a system with more than 61 VFs, the 62nd VF and onwards is assigned with the SINKQP counter, and as a result will have no statistics, and loopback prevention functionality for SINK counter.	-
2.	In ConnectX®-3, since each VF tries to allocate 2 more QP counter for its RoCE traffic statistics, in a system with less than 61 VFs, if there is free resources it receives new counter otherwise receives the default counter which is shared with Ethernet. In this case RoCE statistics is not available.	-
3.	In ConnectX®-3, when we enable function-based loopback prevention for Ethernet port by default (i.e., based on the QP counter index), the dropped self-loopback packets increase the IfRxErrorFrames/Octets counters.	-

3.5 InfiniBand Network

3.5.1 IPoIB Known Issues

Table 23 - IPoIB Known Issues

Index	Description	Workaround
1.	When user increases receive/send a buffer, it might consume all the memory when few child's interfaces are created.	-
2.	The size of send queue in Connect-IB® cards cannot exceed 1K.	-
3.	In 32 bit devices, the maximum number of child interfaces that can be created is 16. Creating more that, might cause out-of-memory issues.	-

Table 23 - IPoIB Known Issues (Continued)

Index	Description	Workaround
4.	In RHEL7.0, the Network-Manager can detect when the carrier of one of the IPoIB interfaces is OFF and can decide to disable its IP address.	Set "`ignore-carrier`" for the corresponding device in NetworkManager.conf. For further information, please refer to *"man NetworkManager.conf"*
5.	IPoIB interface does not function properly if a third party application changes the PKey table. We recommend modifying PKey tables via OpenSM.	-
6.	Fallback to the primary slave of an IPoIB bond does not work with ARP monitoring. (https://bugs.open-fabrics.org/show_bug.cgi?id=1990)	-
7.	Out-of memory issue might occur due to overload of interfaces created.	To calculate the allowed memory per each IPoIB interface check the following: • Num-rings = min(num-cores-on-that-device, 16) • Ring-size = 512 (by default, it is module parameter) • UD memory: 2 * num-rings * ring-size * 8K • CM memory: ring-size * 64k • Total memory = UD mem + CM mem
8.	Connect-IB does not reach the bidirectional line rate	Optimize the IPoIB performance in Connect-IB: `cat /sys/class/net/<interface>/device/local_cpus > /sys/class/net/<interface>/queues/rx-0/rps_cpus`
9.	If the `CONNECTED_MODE=no` parameter is set to "`no`" or missing from the ifcfg file for Connect-IB® IPoIB interface then the `"service network restart"` will hang.	Set the `CONNECTED_MODE=yes parameter` in the ifcfg file for Connect-IB® interface.
10.	Joining a multicast group in the SM using the RDMA_CM API requires IPoIB to first join the broadcast group.	-

Table 23 - IPoIB Known Issues (Continued)

Index	Description	Workaround

11.	Whenever the IOMMU parameter is enabled in the kernel it can decrease the number of child interfaces on the device according to resource limitation.
	The driver will stuck after unknown amount of child interfaces creation.
	
	For further information, please see:
	https://access.redhat.com/site/articles/66747
	http://support.citrix.com/article/CTX136517
	http://www.novell.com/support/kb/doc.php?id=7012337
	https://bugzilla.redhat.com/show_bug.cgi?id=1044595
		To avoid such issue:
		• Decrease the amount of the RX receive buffers (module parameter, the default is 512)
		• Decrease the number of RX rings (sys/fs or ethtool in new kernels)
		• Avoid using IOMMU if not required
		
		For **KVM** users:
		Run:
		`echo 1 > /sys/module/kvm/parameters/allow_unsafe_assigned_interrupts`
		
		To make this change persist across reboots, add the following to the `/etc/modprobe.d/kvm.conf` file (or create this file, if it does not exist):
		`options kvm allow_unsafe_assigned_interrupts=1 kernel parameters`

12.	System might crash in `skb_checksum_help()` while performing TCP retransmit involving packets with 64k packet size.
	A similar out to the below will be printed:
	`kernel BUG at net/core/dev.c:1707!`
	`invalid opcode: 0000 [#1] SMP`
	`RIP: 0010:[<ffffffff81448988>] skb_checksum_help+0x148/0x160`
	`Call Trace:`
	`<IRQ>`
	`[<ffffffff81448d83>] dev_hard_start_xmit+0x3e3/0x530`
	`[<ffffffff8144c805>] dev_queue_xmit+0x205/0x550`
	`[<ffffffff8145247d>] neigh_connected_output+0xbd/0x1`
	…
	…
		Use UD mode in IPoIB

13.	When InfiniBand ports are removed from the host (e.g when changing port type from IB to Eth or removing a card from the PCI bus) the remaining IPoIB interface might be renamed.	To avoid it and have persistent IPoIB network devices names for ConnectX ports, add to the `/etc/udev/rules.d/70-persistent-net.rules` file: `SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="*<Port GID>", NAME="ibN"` Where N is the IPoIB required interface index
14.	After releasing a bond interface that contains IPoIB slaves, a call trace might be printed into the dmesg.	-

Table 23 - IPoIB Known Issues (Continued)

Index	Description	Workaround
15.	IPoIB interfaces are loaded without an IP address on SLES 12.	1. Open the "/etc/wicked/common.xml" file. 2. Change: "<use-nanny>false</use-nanny>" to "<use-nanny>true</use-nanny>" 3. Run: # systemctl restart wickedd.service wicked # ifup all
16.	In RHEL7.0, running ifdown then ifup on an interface after changing CONNECTED_MODE in its ifcfg file, will cause the interface bring up to fail.	Reload the driver "/etc/init.d/openibd restart" or reboot the system.
17.	Clone interfaces receive a duplicated IPv6 address when a child interface with the same PKey (a.k.a clone interface) is used for all the clones.	-
18.	eth_ipoib module is not loaded after reloading the ib_ipoib module.	To restart the IPoIB driver, run "/etc/init.d/openibd restart". Do not restart it by manually restarting each module.
19.	In Ubuntu and Debian, the default of the recv_queue_size and send_queue_size is 128 according to the io_mmu issue.	-
20.	In RHEL6.7, when the Network Manager service is enabled and an IPoIB interface is configured using the "nm-connection-editor" tool, the generated ifcfg file is missing the "DEVICE=<interface name>" parameter. Hence, changing the CONNECTED_MODE in the ifcfg file, will cause a failure in the interface bring up.	Either disable the Network Manager, or add "DEVICE=<interface name>" to the interface's ifcfg file.
21.	ifdown command does not function in RH7.x	-
22.	In RHEL7.2, when creating a child interface for the IPoIB interface, in some cases you might get a trace or a panic as the following: RIP: 0010:[<ffffffffa05e204c>] [<ffffffffa05e204c>] ipoib_get_iflink+0x1c/0x30 [ib_ipoib] Call Trace: register_netdevice+0x140/0x430 __ipoib_vlan_add+0x10d/0x290 [ib_ipoib] ipoib_vlan_add+0x1b5/0x240 [ib_ipoib] create_child+0x64/0x90 [ib_ipoib]	-

3.5.2 eIPoIB Known Issues

Table 24 - eIPoIB Known Issues

Index	Description	Workaround
1.	On rare occasions, upon driver restart the following message is shown in the dmesg: `'cannot create duplicate filename '/class/net/eth_ipoib_interfaces'`	-
2.	No indication is received when eIPoIB is non functional.	Run `'ps -ef | grep ipoibd'` to verify its functionality.
3.	eIPoIB requires libvirtd, python	-
4.	eIPoIB supports only active-backup mode for bonding.	-
5.	eIPoIB supports only VLAN Switch Tagging (VST) mode on guests.	-
6.	IPv6 is currently not supported in eIPoIB	-
7.	eIPoIB cannot run when Flow Steering is enabled	-
8.	eIPoIB daemon requires the following libs in order to run: python-libxml2, libvirt-bin, libvirt0	-
9.	The eIPoIB driver in ConnectX®-3 and Connect-IB is currently at beta level.	-

3.5.3 XRC Known Issues

Table 25 - XRC Known Issues

Index	Description	Workaround

1.	Legacy API is deprecated, thus when recompiling applications over MLNX_OFED v2.0-3.x.x, warnings such as the below are displayed.
	`rdma.c:1699: warning: 'ibv_open_xrc_domain' is deprecated (declared at /usr/include/infiniband/ofa_verbs.h:72)`
	`rdma.c:1706: warning: 'ibv_create_xrc_srq' is deprecated (declared at /usr/include/infiniband/ofa_verbs.h:89)`
	These warnings can be safely ignored.
		-

2.	XRC is not functional in heterogeneous clusters containing non Mellanox HCAs.	-
3.	XRC options do not work when using qperf tool.	Use perftest instead
4.	Out-of memory issue might occur due to overload of XRC receive QP with non zero receive queue size created. XRC QPs do not have receive queues.	-

3.5.4 Verbs Known Issues

Table 26 - Verbs Known Issues

Index	Description	Workaround
1.	Using libnl1_1_3~26 or earlier, requires ibv_create_ah protection by a lock for multi-threaded applications.	-
2.	In MLNX_OFED v2.4-1.0.0, if several CQEs are received on a CQ, they will be coalesced and a user-space event will be triggered only once.	When getting an event, poll the CQ until it is empty.
3.	ibv_task_pingpong over ConnectX-2 adapter cards in not supported.	-

3.5.5 RoCE Known Issues

Table 27 - RoCE Known Issues

Index	Description	Workaround
1.	Not configuring the Ethernet devices or independent VMs with a unique IP address in the physical port, may result in RoCE GID table corruption.	Restart the driver
2.	If RDMA_CM is not used for connection management, then the source and destination GIDs used to modify a QP or create AH should be of the same type - IPv4 or IPv6.	-
3.	On rare occasions, the driver reports a wrong GID table (read from /sys/class/infiniband/mlx4_*/ports/*/gids/*). This may cause communication problems.	-
4.	MLNX_OFED v2.1-1.0.0 and onwards is not interoperable with older versions of MLNX_OFED.	-
5.	Since the number of GIDs per port is limited to 128, there cannot be more than the allowed IP addresses configured to Ethernet devices that are associated with the port. Allowed number is: • "127" for a single function machine • "15" for a hypervisor in a multifunction machine • "(127-15)/n" for a guest in a multifunction machine (where n is the number of virtual functions) Note also that each IP address occupies 2 entries when RoCE mode is set to 4 (RoCEv1 + RoCE v2). This further reduces the number of allowed IP addresses.	-
6.	A working IP connectivity between the RoCE devices is required when creating an address handle or modifying a QP with an address vector.	-
7.	IPv4 multicast over RoCE requires the MGID format to be as follow ::ffff:<Multicast IPv4 Address>	-

Table 27 - RoCE Known Issues (Continued)

Index	Description	Workaround
8.	IP RoCEv2 does not support Multicast Listener Discovery (MLD) therefore, multicast traffic over IPv6 may not work as expected.	-
9.	Using GID index 0 (the default GID) is possible only if the matching IPv6 link local address is configured on the net device of the port. This behavior is possible even though the default GID is configured regardless the presence of the IPv6 address.	-
10.	Using IPv6 link local address (GID0) when VLANs are configured is not supported.	-
11.	Using GID index 0 (the default GID) on port 2 is currently not supported on kernel 3.14 and below.	-
12.	Dynamically Connected (DC) in RoCE in ConnectX®-4 is currently not supported.	-
13.	Enslaving a Mellanox device to a bond with already configured IPs (or configured upper devices), prevents these IPs from being configured as GIDs.	1. Enslave the Mellanox device. 2. Configure IP devices.
14.	`ibv_create_ah_from_wc` is not supported for multicast messages.	-
15.	Infiniband error counters that are found under `/sys/class/infiniband/<mlx5_dev>/ports/<port>/` which dis not function properly in ConnectX-4 adapter cards.	-

3.5.6 ISCSI over IPoIB Known Issues

Table 28 - ISCSI over IPoIB Known Issues

Index	Description	Workaround
1.	When working with ISCSI over IPoIB, LRO must be disabled (even if IPoIB is set to connected mode) due to a a bug in older kernels which causes a kernel panic.	-

3.6 Storage Protocols Known Issues

3.6.1 Storage Known Issues

Table 29 - Storage Known Issues

Index	Description	Workaround
1.	Older versions of `rescan_scsi_bus.sh` may not recognize some newly created LUNs.	If encountering such issues, it is recommended to use the '-c' flag.

3.6.2 SRP Known Issues

Table 30 - SRP Known Issues

Index	Description	Workaround
1.	MLNX_OFED SRP installation breaks the ibmvstgt and ibmvscsi symbol resolution in RHEL7.0	-

3.6.3 SRP Interop Known Issues

Table 31 - SRP Interop Known Issues

Index	Description	Workaround
1.	The driver is tested with Storage target vendors recommendations for multipath.conf extensions (ZFS, DDN, TMS, Nimbus, NetApp).	-

3.6.4 DDN Storage Fusion 10000 Target Known Issues

Table 32 - DDN Storage Fusion 10000 Target Known Issues

Index	Description	Workaround
1.	DDN does not accept non-default P_Key connection establishment.	-

3.6.5 Oracle Sun ZFS Storage 7420 Known Issues

Table 33 - Oracle Sun ZFS Storage 7420 Known Issues

Index	Description	Workaround
1.	Ungraceful power cycle of an initiator connected with Targets DDN, Nimbus, NetApp may result in temporary "stale connection" messages when initiator reconnects.	-

3.6.6 iSER Initiator Known Issues

Table 34 - iSER Initiator Known Issues

Index	Description	Workaround
1.	On SLES OSs, the `ib_iser` module does not load on boot.	Add a dummy interface using iscsiadm: • `# iscsiadm -m iface -I ib_iser -o new` • `# iscsiadm -m iface -I ib_iser -o update -n iface.transport_name -v ib_iser`
2	Ubuntu12 requires update of user space open-iscsi to v2.0.873	-
3	The initiator does not respect interface parameter while logging in.	Configure each interface on a different subnet.
4	iSCSID v2.0.873 can enter an endless loop on bind error.	-
5	iSCSID may hang if target crashes during logout sequence (reproducible with TCP).	-
6	SLES12: Logging in with PI disabled followed by a log out and re-log in with PI enabled, without flushing multipath might cause the block layer to panic.	-
7	Rarely, in InfiniBand devices, when a catastrophic error scenario occurs, iSCSI/iSER initiator might not fully recover and result in system hang.	-
8	Ubuntu14.04: Stress login/logout might cause block layer to invoke a WARN trace.	-

3.6.7 iSER Target Known Issues

Table 35 - iSER Target Known Issues

Index	Description	Workaround
1.	Currently only the following OSs are supported: RHEL/ContOS 7.0, SLES12, Ubuntu14.04.	-
2	Stress login/logout from multiple initiators may cause iSER target to panic.	-
3	RHEL/CentOS 7.0: Discovery over RDMA is not supported.	-
4	ib_isert is unavailable on custom kernels after running the `mlnx_add_kernel_support.sh` script.	1. Add `"isert=y"` to the `mlnx_add_kernel_support.sh` script after `"cat << EOF > ofed.conf"`. 2. Use the updated script to build MLNX_OFED for the custom kernel.

3.6.8 ZFS Appliance Known Issues

Table 36 - ZFS Appliance Known Issues

Index	Description	Workaround
1.	Connection establishment occurs twice which may cause iSER to log a stack trace.	-

3.7 Virtualization

3.7.1 SR-IOV Known Issues

Table 37 - SR-IOV Known Issues

Index	Description	Workaround
1.	When using legacy VMs with MLNX_OFED 2.x hypervisor, you may need to set the `'enable_64b_cqe_eqe'` parameter to zero on the hypervisor. It should be set in the same way that other module parameters are set for mlx4_core at module load time. For example, add `"options mlx4_core enable_64b_cqe_eqe=0"` as a line in the file `/etc/modprobe.d/mlx4_core.conf`.	-
2.	`mlx4_port1_mtu` sysfs entry shows a wrong MTU number in the VM.	-
3.	When at least one port is configured as InfiniBand, and the `num_vfs` is provided but the `probe_vf` is not, HCA initialization fails.	Use both the `num_vfs` and the `probe_vf` in the modprobe line.
4.	When working with a bonding device to enslave the Ethernet devices in active-backup mode and failover MAC policy in a Virtual Machine (VM), establishment of RoCE connections may fail.	Unload the module mlx4_ib and reload it in the VM.
5.	Attaching or detaching a Virtual Function on SLES11 SP3 to a guest Virtual Machine while the `mlx4_core` driver is loaded in the Virtual Machine may cause a kernel panic in the hypervisor.	Unload the `mlx4_core` module in the hypervisor before attaching or detaching a function to or from the guest.
6.	When detaching a VF without shutting down the driver from a VM and reattaching it to another VM with the same IP address for the Mellanox NIC, RoCE connections will fail	Shut down the driver in the VM before detaching the VF.
7.	Enabling SR-IOV requires appending the `"intel_iommu=on"` option to the relevant OS in file `/boot/grub/grub.conf`. Without that SR-IOV cannot be loaded.	-
8.	On various combinations of Hypervisor/OSes and Guest/OSes, an issue might occur when attaching/detaching VFs to a guest while that guest is up and running.	Attach/detach VFs to/from a VM only while that VM is down.
9.	The known PCI BDFs for all VFs in kernel command line should be specified by adding `xen-pciback.hide` For further information, please refer to http://wiki.xen.org/wiki/Xen_PCI_Passthrough	-
10.	The inbox qemu version (2.0) provided with Ubuntu 14.04 does not work properly when more than 2 VMs are run over an Ubuntu 14.04 Hypervisor.	-

Table 37 - SR-IOV Known Issues (Continued)

Index	Description	Workaround
11.	SR-IOV UD QPs are forced by the Hypervisor to use the base GID (i.e., the GID that the VF sees in its GID entry at its paravirtualized index 0). This is needed for security, since UD QPs use Address Vectors, and any GID index may be placed in such a vector, including `indices` not belonging to that VF.	-
12.	Attempting to attach a PF to a VM when SR-IOV is already enabled on that PF may result in a kernel panic.	-
13.	osmtest on the Hypervisor fails when SR-IOV is enabled. However, only the test fails, OpenSM will operate correctly with the host. The failure reason is that if an mcg is already joined by the host, a subsequent join request for that group succeeds automatically (even if the join parameters in the request are not correct). This success does no harm.	-
14.	If a VM does not support PCI hot plug, detaching an mlx4 VF and probing it to the hypervisor may cause the hypervisor to crash.	-
15.	QPerf test is not supported on SR-IOV guests in Connect-IB cards.	-
16.	On ConnectX®-3 HCAs with firmware version 2.32.5000 and later, SR-IOV VPI mode works only with Port 1 = ETH and Port 2 = IB.	-
17.	Occasionally, the `lspci | grep Mellanox` command shows incorrect or partial information due to the current `pci.ids` file on the machine.	1. Locate the file: `$locate pci.ids` 2. Manually update the file according to the latest version available online at: https://pci-ids.ucw.cz/v2.2/pci.ids This file can also be downloaded.
18	SR-IOV is not supported in AMD architecture.	-
19	Setting 1 Mbit/s rate limit on Virtual Functions (Qos Per VF feature) may cause TX queue transmit timeout.	-
20	DC transport type is not supported on SR-IOV VMs.	-
21	Attaching a VF to a VM before unbinding it from the hypervisor and then attempting to destroy the VM, may cause the system to hang for a few minutes.	-
22	When using SR-IOV make sure to set interface to down and unbind BEFORE unloading driver/removing VF/restarting VM or kernel will lock. (reboot needed) Basically, clean-up might not work perfectly so user should do it manually.	-
23	Repeating change of the `mlx5_num_vfs` value from 0 to non-zero, might cause kernel panic in the PF driver.	-

3.8 Resiliency

3.8.1 Reset Flow Known Issues

Table 38 - Resiliency Known Issues

Index	Description	Workaround
1.	SR-IOV non persistent configuration (such as VGT, VST, Host assigned GUIDs, and QP0-enabled VFs) may be lost upon Reset Flow.	Reset Admin configuration post Reset Flow
2.	Upon Reset Flow or after running restart driver, Ethernet VLANs are lost.	Reset the VLANs using the `ifup` command.
3.	Restarting the driver or running `connectx_port_config` when Reset Flow is running might result in a kernel panic	-
4.	Networking configuration (e.g. VLANs, IPv6) should be statically defined in order to have them set after Reset Flow as of after restart driver.	-
5.	After recovering from an EEH event, mlx5_core/ mlx4_core unload may fail due to a bug in some kernel versions. The bug is fixed in Kernel 3.15	-

3.9 Miscellaneous Known Issues

3.9.1 General Known Issues

Table 39 - General Known Issues

Index	Description	Workaround
1.	On ConnectX-2/ConnectX-3 Ethernet adapter cards, there is a mismatch between the GUID value returned by firmware management tools and that returned by fabric/driver utilities that read the GUID via device firmware (e.g., using ibstat). Mlxburn/ flint return 0xffff as GUID while the utilities return a value derived from the MAC address. For all driver/ firmware/software purposes, the latter value should be used.	N/A. Please use the GUID value returned by the fabric/driver utilities (not 0xffffff).
2.	On rare occasions, under extremely heavy MAD traffic, MAD (Management Datagram) storms might cause soft-lockups in the UMAD layer.	-
3.	Packets are dropped on the SM server on big clusters.	Increase the `recv_queue_size` of `ib_mad` module parameter for SM server to 8K. The `recv_queue_size` default size (4K)

3.9.2 ABI Compatibility Known Issues

Table 40 - ABI Compatibility Known Issues

Index	Description	Workaround
1.	MLNX_OFED v2.3-1.0.1 is not ABI compatible with previous MLNX_OFED/OFED versions.	Recompile the application over the new MLNX_OFED version

3.9.3 Connection Manager (CM) Known Issues

Table 41 - Connection Manager (CM) Known Issues

Index	Description	Workaround
1.	When 2 different ports have identical GIDs, the CM might send its packets on the wrong port.	All ports must have different GIDs.

3.9.4 Fork Support Known Issues

Table 42 - Fork Support Known Issues

Index	Description	Workaround
1.	Fork support from kernel 2.6.12 and above is available provided that applications do not use threads. `fork()` is supported as long as the parent process does not run before the child exits or calls `exec()`. The former can be achieved by calling `wait(childpid)`, and the latter can be achieved by application specific means. The Posix system() call is supported.	-

3.9.5 MLNX_OFED Sources Known Issues

Table 43 - MLNX_OFED Sources Known Issues

Index	Description	Workaround
1.	MLNX_OFED includes the OFED source RPM packages used as a build platform for kernel code but does not include the sources of Mellanox proprietary packages.	-

3.9.6 Uplinks Known Issues

Table 44 - Uplinks Known Issues

Index	Description	Workaround
1.	On rare occasions, ConnectX®-3 Pro adapter card may fail to link up when performing parallel detect to 40GbE.	Restart the driver

3.9.7 Resources Limitation Known Issues

Table 45 - Resources Limitation Known Issues

Index	Description	Workaround
1.	The device capabilities reported may not be reached as it depends on the system on which the device is installed and whether the resource is allocated in the kernel or the userspace.	-
2.	mlx4_core can allocate up to 64 MSI-X vectors, an MSI-X vector per CPU.	-
3.	Setting more IP addresses than the available GID entries in the table results in failure and the "update_gid_table error message is displayed: `GID table of port 1 is full. Can't add <address>" message.`	-
4.	Registering a large amount of Memory Regions (MR) may fail because of DMA mapping issues on RHEL 7.0.	-

Table 45 - Resources Limitation Known Issues (Continued)

Index	Description	Workaround
5.	Occasionally, a user process might experience some memory shortage and not function properly due to Linux kernel occupation of the system's free memory for its internal cache.	To free memory to allow it to be allocated in a user process, run the `drop_caches` procedure below. Performing the following steps will cause the kernel to flush and free pages, dentries and inodes caches from memory, causing that memory to become free. **Note:** As this is a non-destructive operation and dirty objects are not freeable, run `` `sync' `` first. • To free the pagecache: `echo 1 > /proc/sys/vm/drop_caches` • To free dentries and inodes: `echo 2 > /proc/sys/vm/drop_caches` • To free pagecache, dentries and inodes: `echo 3 > /proc/sys/vm/drop_caches`

3.9.8 Accelerated Verbs Known Issues

Table 46 - Accelerated Verbs Known Issues

Index	Description	Workaround
1.	On ConnectX®-4 Lx, the following may not be supported when using Multi-Packet WR flag (IBV_EXP_QP_BURST_CREATE_ENABLE_MULTI_PACKET_SEND_WR) on QP-burst family creation: • ACLs • SR-IOV (eSwitch offloads) • priority and dscp forcing • Loopback decision. • VLAN insertion • encapsulation (encap/decap) • sniffer • Signature	-

3.10 InfiniBand Fabric Utilities Known Issues

3.10.1 Performance Tools Known Issues

Table 47 - Performance Tools Known Issues

Index	Description	Workaround
1.	perftest package in MLNX_OFED v2.2-1.0.1 and onwards does not work with older versions of the driver.	-

3.10.2 Diagnostic Utilities Known Issues

Table 48 - Diagnostic Utilities Known Issues

Index	Description	Workaround
1.	When running the ibdiagnet check `nodes_info` on the fabric, a warning specifying that the card does not support general info capabilities for all the HCAs in the fabric will be displayed.	Run `ibdiagnet --skip nodes_info`
2.	ibdump does not work when IPoIB device managed Flow Steering is OFF and at least one of the ports is configured as InfiniBand.	Enable IPoIB Flow Steering and restart the driver. For further information, please refer to MLNX-_OFED User Manual section Enable/Disable Flow Steering.

3.10.3 Tools Known Issues

Table 49 - Tools Known Issues

Index	Description	Workaround
1.	Running ibdump in InfiniBand mode with firmware older than v2.33.5000, may cause the server to hang due to a firmware issue.	Run ibdump with firmware v2.33.5000 and higher

4 Bug Fixes History

Table 50 lists the bugs fixed in this release.

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
1.	IB MAD	Fixed an issue causing MADs to drop in large scale clusters.	3.1-1.0.0	3.1-1.0.3
2.	SR-IOV	Fixed InfiniBand counters which were unavailable in the VM.	2.1-1.0.0	3.1-1.0.0
3.	RoCE	Fixed InfiniBand traffic counters that are found under `/sys/class/infiniband/<mlx5_dev>/ports/<port>/` which dis not function properly in ConnectX-4 adapter cards.	3.0-1.0.1	3.1-1.0.0
4.	Virtualization	Fixed VXLAN functionality issues.	3.0-2.0.1	3.1-1.0.0
5.	Performance	TCP/UDP latency on ConnectX®-4 was higher than expected.	3.0-2.0.1	3.1-1.0.0
6.		TCP throughput on ConnectX®-4 achieved full line rate.	3.0-2.0.1	3.1-1.0.0
7.		Fixed an issue causing inconsistent performance with ConnectX-3 and PowerKVM 2.1.1.	3.0-2.0.1	3.1-1.0.0
8.		Fixed ConnectX-4 traffic counters.	3.0-2.0.1	3.1-1.0.0
9.	`num_entries`	Updated the desired `num_entries` in each iteration, and accordingly updated the offset of the WC in the given WC array.	3.0-1.0.1	3.1-1.0.0
10.	mlx5 driver	Fixed incorrect port rate and port speed values in RoCE mode in ConnectX-4.	3.0-2.0.1	3.1-1.0.0
11.	IPoIB	In RedHat7.1 kernel 3.10.0-299, when sending ICMP/TCP/UDP traffic over Connect-IB/ ConnectX-4 in UD mode, the packets were dropped with the following error: `UDP: bad checksum...`	3.0-2.0.1	3.1-1.0.0
12.	openibd	Fixed an issue which prevented openibd from starting correctly during boot.	3.0-2.0.1	3.1-1.0.0
13.	Ethernet	Added a new module parameter to control the number of IRQs allocated to the device.	3.0-2.0.1	3.1-1.0.0
14.	mlx5 driver	Fixed an issue on PPC servers which prevented PCI from reloading after EEH error recovery.	3.0-2.0.1	3.1-1.0.0
15.	OpenSM	Fixed an issue which prevented the OpenSM package from being fully removed when uninstalling MLNX_OFED v3.0-2.0.1	3.0-2.0.1	3.1-1.0.0

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
16.	mlx5_en	Added the option to toggle LRO ON/OFF using the "-K" flags. The priv flag hw_lro will determine the type of LRO to be used, if the flag is ON, the hardware LRO will be used, otherwise the software LRO will be used.	3.0-2.0.1	3.1-1.0.0
17.		Added the option to toggle LRO ON/OFF using the "-K" flags.	3.0-2.0.1	3.1-1.0.0
18.		Fixed race when updating counters.	3.0-2.0.1	3.1-1.0.0
19.		Fixed scheduling while sending atomic dmesg warning during bonding configuration.	3.0-2.0.1	3.1-1.0.0
20.		Added `set_rx_csum` callback implementation.	3.0-2.0.1	3.1-1.0.0
21.	mlx4_ib	Fixed mismatch between SL and VL in outgoing QP1 packets, which caused buffer overruns in attached switches at high MAD rates.	3.0-1.0.1	3.1-1.0.0
22.	SR-IOV/RoCE	Fixed a problem on VFs where the RoCE driver registered a zero MAC into the port's MAC table (during QP1 creation) because the ETH driver had not yet generated a non-zero random MAC for the ETH port.t	2.3-1.0.1	3.1-1.0.0
23.		Removed BUG_ON assert when checking if the ring is full.	3.0-1.0.1	3.1-1.0.0
24.	libvma	Added libvma support for Debian 8.0 x86_64 and Ubuntu 15.04	3.0-2.0.1	3.1-1.0.0
25.	IPoIB	Fixed an issue which prevented the failure to destroy QP upon IPoIB unload on debug kernel.	3.0-1.0.1	3.0-2.0.0
26.	Configuration	Fixed an issue which prevented the driver version to be reported to the Remote Access Controller tools (such as iDRAC)	3.0-1.0.1	3.0-2.0.0
27.	SR-IOV	Passed the correct port number in port-change event to single-port VFs, where the actual physical port used is port 2.	2.4-1.0.0	3.0-2.0.0
28.		Enabled OpenSM, running over a ConnectX-3 HCA, to manage a mixed ConnectX-3/ConnectX-4 network (by recognizing the "Well-known GID" in mad demux processing).	3.0-1.0.1	3.0-2.0.0
29.		Fixed double-free memory corruption in case where SR-IOV enabling failed (error flow).	3.0-1.0.1	3.0-2.0.0
30.	Start-up sequence	Fixed a crash in EQ's initialization error flow.	3.0-1.0.1	3.0-2.0.0

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
31.	Installation	MLNX_OFED v3.0-1.0.1 installation using yum fails on RH7.1	3.0-1.0.1	3.0-2.0.0
32.	mlx5 driver	In PPC systems, when working with ConnectX®-4 adapter card configured as Ethernet, driver load fails with BAD INPUT LENGTH. dmesg: `command failed, status bad input length(0x50), syndrome 0x9074aa`	3.0-1.0.1	3.0-2.0.0
33.		Error counters such as: CRC error counters, RX out range length error counter, are missing in the ConnectX-4 Ethernet driver.	3.0-1.0.1	3.0-2.0.0
34.		Changing the RX queues number is not supported in Ethernet driver when connected to a ConnectX-4 card.	3.0-1.0.1	3.0-2.0.0
35.	Ethernet	Hardware checksum call trace may appear when receiving IPV6 traffic on PPC systems that uses CHECKSUM COMPLETE method.	3.0-1.0.1	3.0-2.0.0
36.	mlx4_en	Fixed ping/traffic issue occurred when RXVLAN offload was disabled and CHECKSUM COMPLETE was used on ingress packets.	2.4-1.0.4	3.0-1.0.1
37.	Security	CVE-2014-8159 Fix: Prevented integer overflow in IB-core module during memory registration.	2.0-2.0.5	2.4-1.0.4
38.	mlx5_ib	Fixed the return value of max inline received size in the created QP.	2.3-2.0.1	2.4-1.0.0
39.		Resolved soft lock on massive amount of user memory registrations	2.3-2.0.1	2.4-1.0.0
40.	InfiniBand Counters	Occasionally, `port_rcv_data` and `port_xmit_data` counters may not function properly.	2.3-1.0.1	2.4-1.0.0
41.	mlx4_en	LRO fixes and improvements for jumbo MTU.	2.3-2.0.1	2.4-1.0.0
42.		Fixed a crash occurred when changing the number of rings (ethtool set-channels) when interface connected to netconsole.	2.2-1.0.1	2.4-1.0.0
43.		Fixed ping issues with IP fragmented datagrams in MTUs 1600-1700.	2.2-1.0.1	2.4-1.0.0
44.		The default priority to TC mapping assigns all priorities to TC0. This configuration achieves fairness in transmission between priorities but may cause undesirable PFC behavior where pause request for priority "n" affects all other priorities.	2.3-1.0.1	2.4-1.0.0

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
45.	mlx5_ib	Fixed an issue related to large memory regions registration. The problem mainly occurred on PPC systems due to the large page size, and on non PPC systems with large pages (contiguous pages).	2.3-2.0.1	2.3-2.0.5
46.		Fixed an issue in verbs API: fallback to glibc on contiguous memory allocation failure	2.3-2.0.1	2.3-2.0.5
47.	IPoIB	Fixed a memory corruption issue in multi-core system due to intensive IPoIB transmit operation.	2.3-2.0.1	2.3-2.0.5
48.	IB MAD	Fixed an issue to prevent process starvation due to MAD packet storm.	2.3-2.0.1	2.3-2.0.5
49.	IPoIB	Fixed an issue which prevented the spread of events among the closet NUMA CPU when only a single RX queue existed in the system.	2.3-1.0.1	2.3-2.0.0
50.		Returned the CQ to its original state (armed) to prevent traffic from stopping	2.3-1.0.1	2.3-2.0.0
51.		Fixed a TX timeout issue in CM mode, which occurred under heavy stress combined with ifup/ifdown operation on the IPoIB interface.	2.1-1.0.0	2.3-2.0.0
52.	mlx4_core	Fixed "sleeping while atomic" error occurred when the driver ran many firmware commands simultaneously.	2.3-1.0.1	2.3-2.0.0
53.	mlx4_ib	Fixed an issue related to spreading of completion queues among multiple MSI-X vectors to allow better utilization of multiple cores.	2.1-1.0.0	2.3-2.0.0
54.		Fixed an issue that caused an application to fail when attaching Shared Memory.	2.3-1.0.1	2.3-2.0.0
55.	mlx4_en	Fixed dmesg warnings: `"NOHZ: local_softirq_pending 08"`.	2.3-1.0.1	2.3-2.0.0
56.		Fixed erratic report of hardware clock which caused bad report of PTP hardware Time Stamping.	2.1-1.0.0	2.3-2.0.0
57.	mlx5_core	Fixed race when async events arrived during driver load.	2.3-1.0.1	2.3-2.0.0
58.		Fixed race in `mlx5_eq_int` when events arrived before `eq->dev` was set.	2.3-1.0.1	2.3-2.0.0
59.		Enabled all pending interrupt handlers completion before freeing EQ memory.	2.3-1.0.1	2.3-2.0.0
60.	mlnx.conf	Defined mlnx.conf as a configuration file in mlnx-ofa_kernel RPM	2.1-1.0.0	2.3-2.0.0

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
61.	SR-IOV	Fixed counter index allocation for VFs which enables Ethernet port statistics.	2.3-1.0.1	2.3-2.0.0
62.	iSER	Fixed iSER DIX sporadic false DIF errors caused in large transfers when block merges were enabled.	2.3-1.0.1	2.3-2.0.0
63.	RoCE v2	RoCE v2 was non-functional on big Endian machines.	2.3-1.0.1	2.3-2.0.0
64.	Verbs	Fixed registration memory failure when fork was enabled and contiguous pages or ODP were used.	2.3-1.0.1	2.3-2.0.0
65.	Installation	Using both `'-c|--config'` and `'--add-kernel-support'` flags simultaneously when running the `mlnxofedinstall.sh` script caused installation failure with the following on screen message `"--config does not exist"`.	2.2-1.0.1	2.3-2.0.0
66.	IPoIB	Changing the GUID of a specific SR-IOV guest after the driver has been started, causes the ping to fail. Hence, no traffic can go over that InfiniBand interface.	2.1-1.0.0	2.3-1.0.1
67.	XRC	XRC over ROCE in SR-IOV mode is not functional	2.0-3.1.0	2.2-1.0.1
68.	mlx4_en	Fixed wrong calculation of packet true-size reporting in LRO flow.	2.1-1.0.0	2.2-1.0.1
69.		Fixed kernel panic on Debian-6.0.7 which occurred when the number of TX channels was set above the default value.	2.1-1.0.0	2.2-1.0.1
70.		Fixed a crash incidence which occurred when enabling Ethernet Time-stamping and running VLAN traffic.	2.0-2.0.5	2.2-1.0.1
71.	IB Core	Fixed the QP attribute mask upon smac resolving	2.1-1.0.0	2.1-1.0.6
72.	mlx5_ib	Fixed a send WQE overhead issue	2.1-1.0.0	2.1-1.0.6
73.		Fixed a NULL pointer de-reference on the debug print	2.1-1.0.0	2.1-1.0.6
74.		Fixed arguments to kzalloc	2.1-1.0.0	2.1-1.0.6
75.	mlx4_core	Fixed the locks around completion handler	2.1-1.0.0	2.1-1.0.6
76.	mlx4_core	Restored port types as they were when recovering from an internal error.	2.0-2.0.5	2.1-1.0.0
77.		Added an N/A port type to support port_type_array module param in an HCA with a single port	2.0-2.0.5	2.1-1.0.0

Table 50 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
78.	SR-IOV	Fixed memory leak in SR-IOV flow.	2.0-2.0.5	2.0-3.0.0
79.		Fixed communication channel being stuck	2.0-2.0.5	2.0-3.0.0
80.	mlx4_en	Fixed ALB bonding mode failure when enslaving Mellanox interfaces	2.0-3.0.0	2.1-1.0.0
81.		Fixed leak of mapped memory	2.0-3.0.0	2.1-1.0.0
82.		Fixed TX timeout in Ethernet driver.	2.0-2.0.5	2.0-3.0.0
83.		Fixed ethtool stats report for Virtual Functions.	2.0-2.0.5	2.0-3.0.0
84.		Fixed an issue of VLAN traffic over Virtual Machine in paravirtualized mode.	2.0-2.0.5	2.0-3.0.0
85.		Fixed ethtool operation crash while interface down.	2.0-2.0.5	2.0-3.0.0
86.	IPoIB	Fixed memory leak in Connected mode.	2.0-2.0.5	2.0-3.0.0
87.		Fixed an issue causing IPoIB to avoid pkey value 0 for child interfaces.	2.0-2.0.5	2.0-3.0.0

5 Change Log History

Table 51 - Change Log History

Release	Category	Description
3.1-1.0.0	Wake-on-LAN (WOL)	Wake-on-LAN (WOL) is a technology that allows a network professional to remotely power on a computer or to wake it up from sleep mode.
	Hardware Accelerated 802.1ad VLAN (Q-in-Q Tunneling)	Q-in-Q tunneling allows the user to create a Layer 2 Ethernet connection between two servers. The user can segregate a different VLAN traffic on a link or bundle different VLANs into a single VLAN.
	ConnectX-4 ECN	ECN in ConnectX-4 enables end-to-end congestions notifications between two end-points when a congestion occurs, and works over Layer 3.
	RSS Verbs Support for ConnectX-4 HCAs	Receive Side Scaling (RSS) technology allows spreading incoming traffic between different receive descriptor queues. Assigning each queue to different CPU cores allows better load balancing of the incoming traffic and improve performance.
	Minimal Bandwidth Guarantee (ETS)	The amount of bandwidth (BW) left on the wire may be split among other TCs according to a minimal guarantee policy.
	SR-IOV Ethernet	SR-IOV Ethernet at Beta level
3.0-2.0.1	Virtualization	Added support for SR-IOV for ConnectX-4/Connect-IB adapter cards.
3.0-1.0.1	HCAs	Added support for ConnectX®-4 Single/Dual-Port Adapter supporting up to 100Gb/s.
	RoCE per GID	RoCE per GID provides the ability to use different RoCE versions/ modes simultaneously.
	RoCE Link Aggregation (RoCE LAG): ConnectX-3/ConnectX-3 Pro only	RoCE Link Aggregation (available in kernel 4.0 only) provides failover and link aggregation capabilities for mlx4 device physical ports. In this mode, only one IB port that represents the two physical ports, is exposed to the application layer.
	Resource Domain Experimental Verbs	Resource domain is a verb object which may be associated with QP and/or CQ objects on creation to enhance data-path performance.
	Alias GUID Support in InfiniBand	Enables the `query_gid` verb to return the admin desired value instead of the value that was approved by the SM, to prevent a case where the SM is unreachable or a response is delayed, or if the VF is probed into a VM before their GUID is registered with the SM.

Table 51 - Change Log History

Release	Category	Description
3.0-1.0.1 (cont.)	Denial Of Service (DOS) MAD Prevention	Denial Of Service MAD prevention is achieved by assigning a threshold for each agent's RX. Agent's RX threshold provides a protection mechanism to the host memory by limiting the agents' RX with a threshold.
	QoS per VF (Rate Limit per VF)	Virtualized QoS per VF, (supported in ConnectX-3/ConnectX-3 Pro adapter cards only with firmware v2.33.5100 and above), limits the chosen VFs' throughput rate limitations (Maximum throughput). The granularity of the rate limitation is 1Mbits.
	Ignore Frame Check Sequence (FCS) Errors	Upon receiving packets, the packets go through a checksum validation process for the FCS field. If the validation fails, the received packets are dropped. Using this feature, enables you to choose whether or not to drop the frames in case the FCS is wrong and use the FCS field for other info.
	Sockets Direct Protocol (SDP)	Sockets Direct Protocol (SDP) is a byte-stream transport protocol that provides TCP stream semantics. and utilizes InfiniBand's advanced protocol offload capabilities.
	Scalable Subnet Administration (SSA)	The Scalable Subnet Administration (SSA) solves Subnet Administrator (SA) scalability problems for Infiniband clusters. It distributes the needed data to perform the path-record-calculation needed for a node to connect to another node, and caches these locally in the compute (client) nodes. SSA[a] requires AF_IB address family support (3.12.28-4 kernel and later).
	SR-IOV in ConnectX-3 cards	Changed the Alias GUID support behavior in InfiniBand.
	LLR max retransmission rate	Added LLR max retransmission rate as specified in Vendor Specific MAD V1.1, Table 110 - PortLLRStatistics MAD Description ibdiagnet presents the LLR max_retransmission_rate counter as part of the PM_INFO in db_csv file.
	Experimental Verbs	Added the following verbs: • `ibv_exp_create_res_domain` • `ibv_exp_destroy_res_domain` • `ibv_exp_query_intf` • `ibv_exp_release_intf` Added the following interface families: • `ibv_exp_qp_burst_family` • `ibv_exp_cq_family`
2.4-1.0.4	Bug Fixes	See “Bug Fixes History” on page 36.

Table 51 - Change Log History

Release	Category	Description
2.4-1.0.0	mlx4_en net-device Ethtool	Added support for Ethtool speed control and advertised link mode.
		Added ethtool txvlan control for setting ON/OFF hardware TX VLAN insertion: `ethtool -k txvlan [on/off]`
		Ethtool report on port parameters improvements.
		Ethernet TX packet rate improvements.
	RoCE	RoCE uses now all available EQs and not only the 3 legacy EQs.
	InfiniBand	IRQ affinity hints are now set when working in InfiniBand mode.
	Virtualization	VXLAN fixes and performance improvements.
	libmlx4 & libmlx5	Improved message rate of short massages.
	libmlx5	Added ConnectX®-4 device (4114) to the list of supported devices (hca_table),
	Storage	Added iSER Target driver.
	Ethernet net-device	New adaptive interrupt moderation scheme to improve CPU utilization.
		RSS support of fragmented IP datagram.
	Connect-IB Virtual Function	Added Connect-IB Virtual Function to the list of supported devices.
2.3-2.0.5	mlx5_core	Added the following files under `/sys/class/infiniband/mlx5_0/mr_cache/`: • `rel_timeout`: Defines the minimum allowed time between the last MR creation to the first MR released from the cache. When `rel_timeout = -1`, MRs are not released from the cache • `rel_imm`: Triggers the immediate release of excess MRs from the cache when set to 1. When all excess MRs are released from the cache, `rel_imm` is reset back to 0.
	Bug Fixes	See “Bug Fixes History” on page 36.
2.3-2.0.1	Bug Fixes	See “Bug Fixes History” on page 36.

Table 51 - Change Log History

Release	Category	Description
2.3-2.0.0	Connect-IB	Added Suspend to RAM (S3).
	Reset Flow	Added Enhanced Error Handling for PCI (EEH), a recovery strategy for I/O errors that occur on the PCI bus.
	Register Contiguous Pages	Added the option to ask for a specific address when the register memory is using contiguous page.
	mlx5_core	Moved the `mr_cache` subtree from `debugfs` to `mlx5_ib` while preserving all its semantics.
	InfiniBand Utilities	Updated the ibutils package. Added to the ibdiagnet tool the "ibdiagnet2.mlnx_cntrs" option to enable reading of Mellanox diagnostic counters.
	Bug Fixes	See “Bug Fixes History” on page 36.

Table 51 - Change Log History

Release	Category	Description
2.3-1.0.1	OpenSM	Added Routing Chains support with Minhop/UPDN/FTree/DOR/ Torus-2QoS
		Added double failover elimination. When the Master SM is turned down for some reason, the Standby SM takes ownership over the fabric and remains the Master SM even when the old Master SM is brought up, to avoid any unnecessary re-registrations in the fabric. To enable this feature, set the "master_sm_priority" parameter to be greater than the "sm_priority" parameter in all SMs in the fabric. Once the Standby SM becomes the Master SM, its priority becomes equal to the "master_sm_priority". So that additional SM handover is avoided. Default value of the master_sm_priority is 14. To disable this feature, set the "master_sm_priority" in opensm.conf to 0.
		Added credit-loop free unicast/multicast updn/ftree routing
		Added multithreaded Minhop/UPDN/DOR routing
	RoCE	Added IP routable RoCE modes. For further information, please refer to the MLNX_OFED User Manual.
	Installation	Added apt-get installation support.
	Ethernet	Added support for arbitrary UDP port for VXLAN. From upstream 3.15-rc1 and onward, it is possible to use arbitrary UDP port for VXLAN. This feature requires firmware version 2.32.5100 or higher. Additionally, the following kernel configuration option CONFIG_MLX4_EN_VXLAN=y must be enabled.
		MLNX_OFED no longer changes the OS sysctl TCP parameters.
		Added Explicit Congestion Notification (ECN) support
		Added Flow Steering: A0 simplified steering support
		Added RoCE v2 support

Table 51 - Change Log History

Release	Category	Description
2.3-1.0.1 (cont.)	InfiniBand Network	Added Secure host to enable the device to protect itself and the subnet from malicious software.
		Added User-Mode Memory Registration (UMR) to enable the usage of RDMA operations and to scatter the data at the remote side through the definition of appropriate memory keys on the remote side.
		Added On-Demand-Paging (ODP), a technique to alleviate much of the shortcomings of memory registration.
		Added Masked Atomics operation support
		Added Checksum offload for packets without L4 header support
		Added Memory re-registration to allow the user to change attributes of the memory region.
	Resiliency	Added Reset Flow for ConnectX®-3 (+SR-IOV) support.
	SR-IOV	Added Virtual Guest Tagging (VGT+), an advanced mode of Virtual Guest Tagging (VGT), in which a VF is allowed to tag its own packets as in VGT, but is still subject to an administrative VLAN trunk policy.
	Ethtool	Added Cable EEPROM reporting support
		Disable/Enable ethernet RX VLAN tag striping offload via ethtool
		128 Byte Completion Queue Entry (CQE)
	Non-Linux Virtual Machines	Added Windows Virtual Machine over Linux KVM Hypervisor (SR-IOV with InfiniBand only) support
Rev 2.2-1.0.1	mlnxofedinstall	32-bit libraries are no longer installed by default on 64-bit OS. To install 32-bit libraries use the `'--with-32bit'` installation parameter.
	openibd	Added pre/post start/stop scripts support. For further information, please refer to section *"openibd Script"* in the MLNX_OFED User Manual.
	Reset Flow	Reset Flow is not activated by default. It is controlled by the mlx4_core`'internal_err_reset'` module parameter.

Table 51 - Change Log History

Release	Category	Description
Rev 2.2-1.0.1	InfiniBand Core	Asymmetric MSI-X vectors allocation for the SR-IOV hypervisor and guest instead of allocating 4 default MSI-X vectors. The maximum number of MSI-X vectors is `num_cpu` for port ConnectX®-3 has 1024 MSI-X vectors, 28 MSI-X vectors are reserved. • Physical Function - gets the number of MSI-X vectors according to the `pf_msix_table_size` (multiple of 4 - 1) INI parameter • Virtual Functions – the remaining MSI-X vectors are spread equally between all VFs, according to the `num_vfs mlx4_core` module parameter
	Ethernet	Ethernet VXLAN support for kernels 3.12.10 or higher
		Power Management Quality of Service: when the traffic is active, the Power Management QoS is enabled by disabling the CPU states for maximum performance.
		Ethernet PTP Hardware Clock support on kernels/OSes that support it
	Verbs	Added additional experimental verbs interface. This interface exposes new features which are not integrated yet in to the upstream libibverbs. The Experimental API is an extended API therefor, it is backward compatible, meaning old application are not required to be recompiled to use MLNX-OFED v2.2-1.0.1.
	Performance	Out of the box performance improvements: • Use of affinity hints (based on NUMA node of the device) to indicate the IRQ balancer daemon on the optimal IRQ affinity • Improvement in buffers allocation schema (based on the hint above) • Improvement in the adaptive interrupt moderation algorithm
Rev 2.1-1.0.6	IB Core	Added allocation success verification process to ib_alloc_device.
	dapl	dapl is recompiled with no FCA support.
	openibd	Added the ability to bring up child interfaces even if the parent's ifcfg file is not configured.
	libmlx4	Unmapped the hca_clock_page parameter from mlx4_uninit_context.
	scsi_transport_srp	scsi_transport_srp cannot be cleared up when rport reconnecting fails.
	mlnxofedinstall	Added support for the following parameters: • '--umad-dev-na' • '--without-<package>'

Table 51 - Change Log History

Release	Category	Description
Rev 2.1-1.0.6	Content Packages Updates	The following packages were updated: • bupc to v2.2-407 • mstflint to v3.5.0-1.1.g76e4acf • perftest to v2.0-0.76.gbf9a463 • hcoll to v2.0.472-1 • Openmpi to v1.6.5-440ad47 • dapl to v2.0.40
Rev 2.1-1.0.0	EoIB	EoIB is supported only in SLES11SP2 and RHEL6.4.
	eIPoIB	eIPoIB is currently at GA level.
	Connect-IB®	Added the ability to resize CQs.
	IPoIB	Reusing DMA mapped SKB buffers: Performance improvements when IOMMU is enabled.
	mlnx_en	Added reporting autonegotiation support.
		Added Transmit Packet Steering (XPS) support.
		Added reporting 56Gbit/s link speed support.
		Added Low Latency Socket (LLS) support.
		Added check for dma_mapping errors.
	eIPoIB	Added non-virtual environment support.
Rev 2.0-3.0.0	Operating Systems	Additional OS support: • SLES11SP3 • Fedora16, Fedora17
	Drivers	Added Connect-IB™ support
	Installation	Added ability to install MLNX_OFED with SR-IOV support.
		Added Yum installation support
	EoIB	EoIB (at beta level) is supported only in SLES11SP2 and RHEL6.4
	mlx4_core	Modified module parameters to associate configuration values with specific PCI devices identified by their bus/device/function value format
	mlx4_en	Reusing DMA mapped buffers: major performance improvements when IOMMU is enabled
		Added Port level QoS support
	IPoIB	Reduced memory consumption
		Limited the number TX and RX queues to 16
		Default IPoIB mode is set to work in Datagram, except for Connect-IB™ adapter card which uses IPoIB with Connected mode as default.

Table 51 - Change Log History

Release	Category	Description
Rev 2.0-3.0.0	Storage	iSER (at GA level)
Rev 2.0-2.0.5[b]	Virtualization	SR-IOV for both Ethernet and InfiniBand (at Beta level)
	Ethernet Network	RoCE over SR-IOV (at Beta level)
		eIPoIB to enable IPoIB in a Para-Virtualized environment (at Alpha level)
		Ethernet Performance Enhancements (NUMA related and others) for 10G and 40G
		Ethernet Time Stamping (at Beta level)
		Flow Steering for Ethernet and InfiniBand. (at Beta level)
		Raw Eth QPs: • Checksum TX/RX • Flow Steering
	InfiniBand Network	Contiguous pages: • Internal memory allocation improvements • Register shared memory • Control objects (QPs, CQs)
	Installation	YUM update support
	VMA	OFED_VMA integration to a single branch
	Storage	iSER (at Beta level) and SRP
	Operating Systems	Errata Kernel upgrade support
	API	VERSION query API: library and headers
	Counters	64bit wide counters (port xmit/recv data/packets unicast/mcast)

a. SSA is tested on SLES 12 only (x86-64 architecture).
b. SR-IOV, Ethernet Time Stamping and Flow Steering are ConnectX®-3 HCA capability.

6 API Change Log History

Table 52 - API Change Log History

Release	Name	Description
Rev 3.1-1.0.3	`libibverbs`	• Added `ibv_exp_wq_family` interface family (Supported only by ConnectX®-4) • Added flag to the QP-burst family to enable Multi-Packet WR • Added return error statuses to the `ibv_exp_query_intf` to notify that common-flags/family-flags are not supported. • Added `ibv_exp_query_gid_attr` verb. For further information, please refer to the manual page of the verb.

Rev 3.0-1.0.0	`libibverbs`	• Added the following new APIs:
		  • ibv_exp_create_res_domain - create resource domain
		  • ibv_exp_destroy_res_domain - destroy resource domain
		  • ibv_exp_query_intf - query for family of verbs interface for specific QP/CQ
		  • ibv_exp_release_intf - release the queried interface
		• Updated the following APIs:
		  • ibv_exp_create_qp - Add resource-domain to the verb parameters
		  • ibv_exp_create_cq - Add resource-domain to the verb parameters

Rev 2.4-1.0.0	libibverbs	Added the following verbs interfaces: • `ibv_create_flow` • `ibv_destroy_flow` • `ibv_exp_use_priv_env` • `ibv_exp_setenv`

Rev 2.3-1.0.1	libibverbs	• `ibv_exp_rereg_mr` - Added new API for memory region re-reintegration (For futher information, please refer to MLNX_OFED User Manual)
		• Added to the experimental API `ibv_exp_post_send` the following opcodes:
		  • `IBV_EXP_WR_EXT_MASKED_ATOMIC_CMP_AND_SWP`
		  • `IBV_EXP_WR_EXT_MASKED_-ATOMIC_FETCH_AND_ADD`
		  • `IBV_EXP_WR_NOP`
		and these completion opcodes:
		  • `IBV_EXP_WC_MASKED_COMP_SWAP`
		  • `IBV_EXP_WC_MASKED_FETCH_ADD`

Table 52 - API Change Log History

Release	Name	Description
Rev 2.2-1.0.1	libibverbs	The following verbs changed to align with upstream libibverbs: • `ibv_reg_mr - ibv_access_flags changed.` • `ibv_post_send - opcodes and send flags changed and wr fields removed (task, op, dc and bind_mw)` • `ibv_query_device - capability flags changed.` • `ibv_poll_cq - opcodes and wc flags changed.` • `ibv_modify_qp - mask bits changed` • `ibv_create_qp_ex - create_flags field removed.` The following verbs removed to align with upstream libibverbs: • `ibv_bind_mw` • `ibv_post_task` • `ibv_query_values_ex` • `ibv_query_device_ex` • `ibv_poll_cq_ex` • `ibv_reg_shared_mr_ex` • `ibv_reg_shared_mr` • `ibv_modify_cq` • `ibv_create_cq_ex` • `ibv_modify_qp_ex`
Rev 2.2-1.0.1	Verbs Experimental API	The following experimental verbs added (replacing the removed extended verbs): • ibv_exp_bind_mw • ibv_exp_post_task • ibv_exp_query_values • ibv_exp_query_device • ibv_exp_poll_cq • ibv_exp_reg_shared_mr • ibv_exp_modify_cq • ibv_exp_create_cq • ibv_exp_modify_qp New experimental verbs: • ibv_exp_arm_dct • ibv_exp_query_port • ibv_exp_create_flow • ibv_exp_destroy_flow • ibv_exp_post_send • ibv_exp_reg_mr • ibv_exp_get_provider_func

Table 52 - API Change Log History

Release	Name	Description
Rev 2.1-1.0.0	Dynamically Connected (DC)	The following verbs were added: • `struct ibv_dct *ibv_exp_create_dct(struct ibv_context *context, struct ibv_exp_dct_init_attr *attr)` • `int ibv_exp_destroy_dct(struct ibv_dct *dct)` • `int ibv_exp_query_dct(struct ibv_dct *dct, struct ibv_exp_dct_attr *attr)`
	Verbs Extension API: Verbs extension API defines OFA APIs extension scheme to detect ABI compatibility and enable backward and forward compatibility support.	• ibv_post_task • ibv_query_values_ex • ibv_query_device_ex • ibv_create_flow • ibv_destroy_flow • ibv_poll_cq_ex • ibv_reg_shared_mr_ex • ibv_open_xrcd • ibv_close_xrcd • ibv_modify_cq • ibv_create_srq_ex • ibv_get_srq_num • ibv_create_qp_ex • ibv_create_cq_ex • ibv_open_qp • ibv_modify_qp_ex
Rev 2.1-1.0.0	Verbs Experimental API: Verbs experimental API defines MLNX-OFED APIs extension scheme which is similar to the “Verbs extension API”. This extension provides a way to introduce new features before they are integrated into the formal OFA API and to the upstream kernel and libs.	• ibv_exp_create_qp • ibv_exp_query_device • ibv_exp_create_dct • ibv_exp_destroy_dct • ibv_exp_query_dct
Rev 2.0-3.0.0	XRC	The following verbs have become deprecated: • `struct ibv_xrc_domain *ibv_open_xrc_domain` • `struct ibv_srq *ibv_create_xrc_srq` • `int ibv_close_xrc_domain` • `int ibv_create_xrc_rcv_qp` • `int ibv_modify_xrc_rcv_qp` • `int ibv_query_xrc_rcv_qp` • `int ibv_reg_xrc_rcv_qp` • `int ibv_unreg_xrc_rcv_qp`

Table 52 - API Change Log History

Release	Name	Description
Rev 2.0-2.0.5	Libibverbs - Extended speeds	• Missing the `ext_active_speed attribute` from the struct `ibv_port_attr` • Removed function `ibv_ext_rate_to_int` • Added functions `ibv_rate_to_mbps` and `mbps_to_ibv_rate`
	Libibverbs - Raw QPs	QP types `IBV_QPT_RAW_PACKET` and `IBV_QPT_RAW_ETH` are not supported
	Libibverbs - Contiguous pages	• Added Contiguous pages support • Added function `ibv_reg_shared_mr`
	Libmverbs	• The enumeration `IBV_M_WR_CALC` was renamed to `IBV_M_WR_CALC_SEND` • The enumeration `IBV_M_WR_WRITE_WITH_IMM` was added • In the structure `ibv_m_send_wr`, the union `wr.send` was renamed to `wr.calc_send` and `wr.rdma` was added • The enumerations `IBV_M_WQE_CAP_CALC_RDMA_WRITE_WITH_IMM` was added

		• The following enumerations were renamed:
		  • From `IBV_M_WQE_SQ_ENABLE_CAP` to `IBV_M_WQE_CAP_SQ_ENABLE`
		  • From `IBV_M_WQE_RQ_ENABLE_CAP` to `IBV_M_WQE_CAP_RQ_ENABLE`
		  • From `IBV_M_WQE_CQE_WAIT_CAP` to `IBV_M_WQE_CAP_CQE_WAIT`
		  • From `IBV_M_WQE_CALC_CAP` to `IBV_M_WQE_CAP_CALC_SEND`