



Mellanox OFED for Linux Release Notes

Rev 3.4-2.1.8.0



NOTE:

THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "ASIS" WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCT(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR 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 FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.



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

© Copyright 2017. Mellanox Technologies Ltd. All Rights Reserved.

Mellanox®, Mellanox logo, Accelio®, BridgeX®, CloudX logo, CompustorX®, Connect-IB®, ConnectX®, CoolBox®, CORE-Direct®, EZchip®, EZchip logo, EZappliance®, EZdesign®, EZdriver®, EZsystem®, GPUDirect®, InfiniHost®, InfiniBridge®, InfiniScale®, Kotura®, Kotura logo, Mellanox CloudRack®, Mellanox CloudXMellanox®, Mellanox Federal Systems®, Mellanox HostDirect®, Mellanox Multi-Host®, Mellanox Open Ethernet®, Mellanox OpenCloud®, Mellanox OpenCloud Logo®, Mellanox PeerDirect®, Mellanox ScalableHPC®, Mellanox StorageX®, Mellanox TuneX®, Mellanox Connect Accelerate Outperform logo, Mellanox Virtual Modular Switch®, MetroDX®, MetroX®, MLNX-OS®, NP-1c®, NP-2®, NP-3®, Open Ethernet logo, PhyX®, PlatformX®, PSIPHY®, SiPhy®, StoreX®, SwitchX®, Tiler®, Tiler logo, TestX®, TuneX®, The Generation of Open Ethernet logo, UFM®, Unbreakable Link®, Virtual Protocol Interconnect®, Voltaire® and Voltaire logo are registered trademarks of Mellanox Technologies, Ltd.

All other trademarks are property of their respective owners.

For the most updated list of Mellanox trademarks, visit <http://www.mellanox.com/page/trademarks>

Table of Contents

Table of Contents	3
Chapter 1 Overview	6
1.1 Content of Mellanox OFED for Linux.	6
1.2 Supported Platforms and Operating Systems	7
1.2.1 Supported Non-Linux Virtual Machines	7
1.3 Hardware and Software Requirements	7
1.4 Supported HCAs Firmware Versions	8
1.5 Compatibility Matrix.	8
1.6 RoCE Modes Matrix	8
Chapter 2 Changes and New Features in Rev 3.4-2.1.8.0.	9
2.1 Future API Changes in MLNX_OFED	9
Chapter 3 Known Issues	10
3.1 Driver Installation/Loading/Unloading/Start Known Issues	10
3.1.1 Installation Known Issues.	10
3.1.2 Driver Unload Known Issues.	11
3.1.3 Driver Start Known Issues.	11
3.1.4 System Time Known Issues.	14
3.1.5 UEFI Secure Boot Known Issues.	14
3.2 Performance Known Issues	14
3.3 HCAs Known Issues.	17
3.3.1 mlx5 Driver Known Issues.	17
3.4 Ethernet Network.	18
3.4.1 Ethernet Known Issues	18
3.4.2 Port Type Management Known Issues	21
3.4.3 Flow Steering Known Issues.	22
3.4.4 Quality of Service Known Issues	22
3.4.5 Ethernet Performance Counters Known Issues	22
3.5 InfiniBand Network.	23
3.5.1 iPoIB Known Issues	23
3.5.2 eIPoIB Known Issues	27
3.5.3 XRC Known Issues	28
3.5.4 Verbs Known Issues	29
3.5.5 RoCE Known Issues	30
3.5.6 iSCSI over iPoIB Known Issues	32
3.5.7 SDP Known Issues	33
3.6 Storage Protocols Known Issues	33

3.6.1	Storage Known Issues	33
3.6.2	NVMeoF Host/Target Known Issues	33
3.6.3	SRP Known Issues	34
3.6.4	SRP Interop Known Issues	34
3.6.5	DDN Storage Fusion 10000 Target Known Issues	34
3.6.6	Oracle Sun ZFS Storage 7420 Known Issues	34
3.6.7	iSER Initiator Known Issues	35
3.6.8	iSER Target Known Issues	35
3.6.9	ZFS Appliance Known Issues	36
3.6.10	Erasure Coding Verbs Known Issues	36
3.7	Virtualization	37
3.7.1	SR-IOV Known Issues	37
3.8	Resiliency	42
3.8.1	Reset Flow Known Issues	42
3.9	Miscellaneous Known Issues	43
3.9.1	General Known Issues	43
3.9.2	ABI Compatibility Known Issues	44
3.9.3	Connection Manager (CM) Known Issues	44
3.9.4	Fork Support Known Issues	45
3.9.5	Uplinks Known Issues	45
3.9.6	Resources Limitation Known Issues	45
3.9.7	Accelerated Verbs Known Issues	46
3.10	InfiniBand Fabric Utilities Known Issues	47
3.10.1	Performance Tools Known Issues	47
3.10.2	Diagnostic Utilities Known Issues	47
3.10.3	Tools Known Issues	47
Chapter 4 Bug Fixes History		48
Chapter 5 Change Log History		61



Release Update History

Release	Date	Description
Rev 3.4-2.1.8.0	April 23, 2017	Initial release of this version.

1 Overview

These are the release notes of Mellanox OFED for Linux Driver, Rev 3.4-2.1.8.0 which operates across all Mellanox network adapter solutions supporting the following uplinks to servers:

Uplink/HCAs	Uplink Speed
ConnectX®-4	InfiniBand: SDR, QDR, FDR, FDR10, EDR

1.1 Content of Mellanox OFED for Linux

Mellanox OFED for Linux software contains the following components:

Components	Description
OpenFabrics core and ULPs	<ul style="list-style-type: none"> • InfiniBand and Ethernet HCA drivers (mlx4, mlx5) • core • Upper Layer Protocols: IPoIB, SRP, iSER and iSER Initiator and Target
OpenFabrics utilities	<ul style="list-style-type: none"> • OpenSM: IB Subnet Manager with Mellanox proprietary Adaptive Routing • Diagnostic tools • Performance tests • SSA (SLES12): libopensmssa plugin for OpenSM, ibssa, ibacm
MPI	<ul style="list-style-type: none"> • Open MPI stack 1.10.5 and later supporting the InfiniBand interface • MPI benchmark tests (OSU benchmarks, Intel MPI benchmarks, Presta)
PGAS	<ul style="list-style-type: none"> • HPC-X OpenSHMEM v2.2 supporting InfiniBand, MXM and FCA • HPC-X UPC v2.2 supporting InfiniBand, MXM and FCA
HPC Acceleration packages	<ul style="list-style-type: none"> • Mellanox MXM v3.5.3093 (p2p transport library acceleration over Infiniband) • Mellanox FCA v2.5.2431 (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	<ul style="list-style-type: none"> • ibutils2 • ibdump • MFT
Sources of all software modules (under conditions mentioned in the modules' LICENSE files) except for MFT, OpenSM plugins, ibutils2, and ibdump	
Documentation	

1.2 Supported Platforms and Operating Systems

The following are the supported OSs in MLNX_OFED Rev 3.4-2.1.8.0:

Table 1 - Supported Platforms and Operating Systems

Operating System	Platform
RHEL7.2	x86_64
RHEL7.3	x86_64



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 Manual for instructions.



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

1.2.1 Supported Non-Linux Virtual Machines

The following are the supported Non-Linux (InfiniBand only) Virtual Machines in MLNX_OFED Rev 3.4-2.1.8.0:

- Windows Server 2012 R2

1.3 Hardware and Software Requirements

The following are the hardware and software requirements of MLNX_OFED Rev 3.4-2.1.8.0.

- 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:

Operating System	Required Packages Installation Command
RHEL/OL/Fedora	<code>yum install perl pciutils python gcc-gfortran libxml2-python tcsh libnl.i686 libnl expat glib2 tcl libstdc++ bc tk gtk2 atk cairo numactl pkgconfig ethtool</code>

1.4 Supported HCAs Firmware Versions

MLNX_OFED Rev 3.4-2.1.8.0 supports the following Mellanox network adapter cards firmware versions:

Table 2 - Supported HCAs Firmware Versions

HCA	Recommended Firmware Rev.	Additional Firmware Rev. Supported
ConnectX®-4	12.17.2046 ^a	N/A

a. This firmware version is only supported for the following OPNs: MCX455A-ECAT and MCX456A-ECAT.

For the official firmware versions, please see:

http://www.mellanox.com/content/pages.php?pg=firmware_download

1.5 Compatibility Matrix

MLNX_OFED Rev 3.4-2.1.8.0 is compatible with the following:

Table 3 - Compatibility Matrix

Mellanox Product	Description/Version
MLNX-OS®	MSX6036 w/w MLNX-OS® version 3.4.3202 ^a
Grid Director™	4036 w/w Grid Director™ version 3.9.1-985
Unified Fabric Manager (UFM®)	v5.8
MXM	v3.5.3093
HPC-X UPC	v2.22
HPC-X OpenSHMEM	v1.8.3
FCA	v2.5.2431
OpenMPI	v1.10.5a1

a. MLNX_OFED Rev 1.0 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 4 - RoCE Modes Matrix

Software Stack / Inbox Distribution	RoCEv1 (IP Based GIDs) Supported as of Version		RoCEv2 Supported as of Version		RoCEv1 & RoCEv2 (RoCE per GID) Supported as of Version
	ConnectX-3 & ConnectX- 3 Pro	ConnectX-4 & ConnectX- 4 Lx	ConnectX- 3 Pro	ConnectX-4 & ConnectX-4 Lx	ConnectX-3 Pro, ConnectX- 4 & ConnectX-4 Lx
MLNX- _OFED	2.1-x.x.x	3.0-x.x.x	2.3-x.x.x	3.0-x.x.x	3.0-x.x.x
Kernel.org	3.14	4.4	4.4	4.4	4.4
RHEL	6.6, 7.0	-	-	-	-

Table 4 - RoCE Modes Matrix

Software Stack / Inbox Distribution	RoCEv1 (IP Based GIDs) Supported as of Version		RoCEv2 Supported as of Version		RoCEv1 & RoCEv2 (RoCE per GID) Supported as of Version
	ConnectX-3 & ConnectX- 3 Pro	ConnectX-4 & ConnectX- 4 Lx	ConnectX- 3 Pro	ConnectX-4 & ConnectX-4 Lx	ConnectX-3 Pro, ConnectX- 4 & ConnectX-4 Lx
SLES	12	-	-	-	-
Ubuntu	14.04.4, 16.04, 15.10	-	-	-	-

2 Changes and New Features in Rev 3.4-2.1.8.0

The following are the changes and/or new features that have been added to this version of MLNX_OFED.

Table 5 - Changes and New Features in Rev 3.4-2.1.8.0

Feature/Change	Description
Bug fixes	See Section 4, “Bug Fixes History” , on page 48.

2.1 Future API Changes in MLNX_OFED



Note that the following APIs will be deprecated and replaced with the new APIs as of MLNX_OFED version 4.0, as listed in the table below.

Feature	Type	Current API	New API
Rereg MR	Verb	ibv_exp_rereg_mr	ibv_rereg_mr
Memory Window	Verb	ibv_exp_bind_mw	ibv_bind_mw
	Structure	ibv_exp_send_wr -> bind_mw	ibv_send_wr -> bind_mw
	Opcodes	IBV_EXP_WR_SEND_WITH_INV	IBV_WR_SEND_WITH_INV
		IBV_EXP_WR_LOCAL_INV	IBV_WR_LOCAL_INV
		IBV_EXP_WR_BIND_MW	IBV_WR_BIND_MW
	Capability	IBV_EXP_DEVICE_MEM_WINDOW	IBV_DEVICE_MEM_WINDOW
Completion	IBV_EXP_WC_WITH_INV	IBV_WC_WITH_INV	

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 6 - Installation Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#918880: The driver version shown in mod-info and ethtool outputs is 3.4-1.0.6 instead of 3.4-2.0.0.	-
2.	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 (<code>/etc/init.d/openibd stop</code>) before upgrading to this new version.
3.	When using bonding on Ubuntu OS, the "ifenslave" package must be installed.	-
4.	On PPC systems, the <code>ib_srp</code> module is not installed by default since it breaks the <code>ibm-vscsi</code> module.	If your system does not require the <code>ibm-vscsi</code> module, run the <code>mlnxofedinstall</code> script with the " <code>--with-srp</code> " flag.
5.	#679801: Updating MLNX_OFED via Yum (e.g. running " <code>yum update mlnx-ofed-all</code> ") can fail with the following error: --> Finished Dependency Resolution Error: Package: <code>mpitests_openmpi__1_8_8-3.2.16-fe5387c.x86_64</code> (installed) Requires: <code>liboshmem.so.3()(64bit)</code> Removing: <code>openmpi-1.8.8-1.x86_64</code> (installed) <code>liboshmem.so.3()(64bit)</code> Updated By: <code>openmpi-1.10.2rc4-1.32008.x86_64</code> (mlnx_ofed) <code>~liboshmem.so.9()(64bit)</code>	Remove the <code>mpitests</code> packages manually: <pre># rpm -e --allmatches \$(rpm -qa grep mpitests_)</pre>
6.	#690799: OpenSM package removal fails with the following error on Ubuntu12.04: Removing <code>opensm</code> ... <code>/sbin/insserv: No such file or directory</code>	1. Create the missing link by running this command: <pre># ln -s /usr/lib/insserv/insserv /sbin/insserv</pre> 2. Remove the package.

Table 6 - Installation Known Issues

Index	Internal Reference Number: Description	Workaround
7.	#764204: Weak Updates (KMP) support is broken on RHEL PPC64LE with errata kernels. MLNX_OFED installation will pass, but no links will be created under the weak-updates directory for the new kernel. Therefore, the driver load will fail.	<ul style="list-style-type: none"> As of MLNX_OFED v3.3, use the <code>mlnx_add_kernel_support.sh</code> script, or simply provide the <code>--add-kernel-support</code> flag to <code>mlnxofedinstall</code> script. Update the <code>kmop</code> package using the following link: https://rhn.redhat.com/errata/RHBA-2016-1832.html
8.	#785119: When upgrading ConnectX-4/ConnectX-4 Lx firmware version from v12/14.14.2036 to a newer one (for example:12/14.16.1xxx), power cycle is necessary to enable working in Pass-Through mode. Using <code>mlxfwreset</code> instead of power cycle will print messages similar to the following when Passing-Through the device to Virtual Machine: "-device vfio-pci,host=04:00.0,id=host-dev0,bus=pci.0,addr=0x7: vfio: Error: Failed to setup INTx fd: No such device 2016-05-22T06:46:39.164786Z qemu-kvm: -device vfio-pci,host=04:00.0,id=host-dev0,bus=pci.0,addr=0x7: Device initialization failed."	-

3.1.2 Driver Unload Known Issues

Table 7 - Driver Unload Known Issues

Index	Internal Reference Number: Description	Workaround
1.	"openibd stop" can sometime fail with the error: Unloading <code>ib_cm</code> [FAILED] ERROR: Module <code>ib_cm</code> is in use by <code>ib_i-poib</code>	Re-run "openibd stop"

3.1.3 Driver Start Known Issues

Table 8 - Driver Start Known Issues

Index	Internal Reference Number: Description	Workaround
1.	Failure to load a 4K page on ARM architecture.	Enlarge the CMA area by adding <code>cma=256M</code> or more to <code>grub.conf</code> .

Table 8 - Driver Start Known Issues

Index	Internal Reference Number: Description	Workaround
2.	“Out-of-memory” issues may rise during drivers load depending on the values of the driver module parameters set (e.g. log_num_cq).	-
3.	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"
4.	Occasionally, when trying to repetitively reload the NES hardware driver on SLES11 SP2, a soft lockups occurs that required reboot.	-
5.	When downgrading from MLNX_OFED 3.0-x.x.x, driver reload might fail with the following error 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
6.	#773774: When downgrading from MLNX_OFED 3.3-x.x.x, driver reload might fail with the following error in dmeg: rmmod: ERROR: Module mlx_compat is in use by: ib_netlink	The issues will be resolved automatically after system reboot or by invoking the following commands: rmmod ib_netlink depmod -a /etc/init.d/openibd restart
7.	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.	-

Table 8 - Driver Start Known Issues

Index	Internal Reference Number: Description	Workaround
8.	“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.	-
9.	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.	<ol style="list-style-type: none"> 1. Unmount any mounted Lustre storages: # umount<lustre_mount_point> 2. Unload all Lustre modules: # lustre_rmmod
10.	Driver unload fails with the following error message: Unloading rdma_cm [FAILED] rmmod: ERROR: Module rdma_cm is in use by: xprtrdma	<p>Make sure that there are no mount points over NFS/RDMA prior to unloading the driver and run:</p> <pre># modprobe -r xprtrdma</pre> <p>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.</p> <p>Create an executable file "/etc/infiniband/pre-stop-hook.sh" with the following content:</p> <pre>#!/bin/bash modprobe -r xprtrdma</pre>
11.	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.hp.com/hpsc/doc/public/display?docId=emr_na-c04446026&lang=en-us&cc=us) The messages are harmless, and may be ignored.	<p>If you are <i>*not*</i> 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.</p>
12.	False alarms errors may be printed to dmesg	-
13.	On RHEL 7.1, after updating to kernel version 3.10.0-229.14.1.el7 or later, driver load fails with unknown symbols errors in dmesg.	Use the mlnx_add_kernel_support.sh script to compile MLNX_OFED drivers against the new kernel.

3.1.4 System Time Known Issues

Table 9 - System Time Known Issues

Index	Internal Reference Number: 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 OL6.x this may change the local system time.	-

3.1.5 UEFI Secure Boot Known Issues

Table 10 - UEFI Secure Boot Known Issues

Index	Internal Reference Number: 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 11 - Performance Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#765777: Low VxLAN throughput due to broken GRO offload in most kernels older than kernel v4.6.	Use kernel version 4.6 or above.
2.	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]

Table 11 - Performance Known Issues

Index	Internal Reference Number: Description	Workaround
3.	Out-of-the-box throughput performance in Ubuntu 14.04 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.
4.	UDP receiver throughput may be lower than 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. <code>ethtool -C <eth>X adaptive-rx off rx-usecs 64 rx-frames 24</code> Values above may need tuning, depending the system, configuration and link speed.
5.	Performance degradation might occur when bonding Ethernet interfaces.	-
6.	In RHEL7.0, when the irqbalance service is started or restarted, it incorrectly re-balances the IRQs, including the banned ones.	-
7.	In RH7.0/RH7.1, performance issue with ConnectX-4 cards over 100GbE link might occur when the process of forwarding the packets between the ports, which is done by the kernel, fib_table_lookup() function is called. For further information, please refer to: http://comments.gmane.org/gmane.linux.network/344243	Use RH7.2 to avoid such performance issues.
8.	#754646: The default RX coalescing values yield to high CPU utilization when using VXLAN on VMs over PV.	Increase the RX microseconds and frames coalescing parameters for a better utilization using the ethtool -C command.

Table 11 - Performance Known Issues

Index	Internal Reference Number: Description	Workaround
9.	#783496: When using a VF over RH7.X KVM, low throughput is expected.	Install the following packages using the link below: <ul style="list-style-type: none"> • qemu-img-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm • qemu-kvm-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm • qemu-kvm-common-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm http://people.redhat.com/~alwillia/bz1299846/

3.3 HCAs Known Issues

3.3.1 mlx5 Driver Known Issues

Table 12 - mlx5 Driver Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#860311: An allocation of high-order page in <code>mlx5e_alloc_striding_rx_wqe</code> fails with a call-trace.	No action is required on user's end. A fragmented fallback flow will handle this failure.
2.	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)	-
3.	EEH events that arrive while the mlx5 driver is loading may cause the driver to hang.	-
4.	The mlx5 driver can handle up to 5 EEH events per hour.	If more events are received, cold reboot the machine.
5.	: When working with Connect-IB® firmware v10.10.5054, the following message would appear in driver start. <code>command failed, status bad system state(0x4), syndrome 0x408b33</code> The message can be safely ignored.	Upgrade Connect-IB firmware to the latest available version.
6.	Changing the link speed is not supported in Ethernet driver when connected to a ConnectX-4 card.	-
7.	Bonding "active-backup" mode does not function properly.	-
8.	Rate, speed and width using IB sysfs/tools are available in RoCE mode in ConnectX-4 only after port physical speed configuration is done.	-
9.	Since MLNX_OFED's <code>openibd</code> does not unload modules while OpenSM is running, removing <code>mlx5_core</code> manually while OpenSM is running, may cause it to be out of sync when probed again.	Restart OpenSM
10.	ConnectX-4 port GIDs table shows a duplicated RoCE v2 default GID.	-

3.4 Ethernet Network

3.4.1 Ethernet Known Issues

Table 13 - Ethernet Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#843306: [ConnectX-4/ConnectX-4 Lx] When configuring ETS, bandwidth values are limited between 1-100, and 0 is an invalid value.	-
2.	#704750: [ConnectX-4/ConnectX-4 Lx] First ICMP6 packet may be lost as a result of first IP fragment loss when packets size is significantly bigger than MTU.	-
3.	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: vhr command ALLOC_RES (0xf00) slave:0 in_param 0x7e in_mod=0x107, op_mod=0x1 failed with error:0, status -28	-
4.	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()
5.	In PPC systems when QoS is enabled a harmless Kernel DMA mapping error messages might appear in kernel log (IOMMU related issue).	-
6.	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)	-
7.	Mixing ETS and strict QoS policies for TCs in 40GbE ports may cause inaccurate results in bandwidth division among TCs.	-

Table 13 - Ethernet Known Issues

Index	Internal Reference Number: Description	Workaround
8.	Creating a VLAN with user priority ≥ 4 on ConnectX®-2 HCA is not supported.	-
9.	Affinity hints are not supported in Xen Hypervisor (an irqblancer issue). This causes a non-optimal IRQ affinity.	To overcome this issues, run: <code>set_irq_affinity.sh eth<x></code>
10.	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"	-
11.	In ConnectX®-2, RoCE UD QP does not include VLAN tags in the Ethernet header	
12.	VXLAN may not be functional when configured over Linux bridge in RH7.0 faceor Ubuntu14.04. The issue is within the bridge modules in those kernels. In Vanilla kernels above 3.16 issues is fixed.	-
13.	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.	-
14.	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: <code>ethtool -K ethX lro off</code>
15.	On SLES12, the bonding interface over Mellanox Ethernet slave interfaces does not get IP address after reboot.	<ol style="list-style-type: none"> 1. Set "STARTMODE=hotplug" in the bonding slave's ifcfg files. More details can be found in the SUSE documentation 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 hot-plugging: Open the "/etc/wicked/common.xml" file. Change: "<code><use-nanny>>false</use-nanny></code>" to "<code><use-nanny>>true</use-nanny></code>" 3. Run: <code># systemctl restart wickedd.service wicked</code>

Table 13 - Ethernet Known Issues

Index	Internal Reference Number: Description	Workaround
16.	ethtool -x command does not function in SLES OS.	-
17.	Ethertype proto 0x806 not supported by eth-tool	-
18.	ETS configuration is not supported in the following kernels: <ul style="list-style-type: none"> • 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 	
19.	ETS is not supported in kernels that do not have MQPRIO as QDISC_KIND option in the tc tool.	-
20.	When NC-SI is ON, the port's MTU cannot be set to lower than 1500.	-
21.	GRO is not functional when using VXLAN in ConnectX-3 adapter cards.	-
22.	ethtool -X: The driver supports only the 'equal' mode and cannot be set by using weight flags.	-
23.	Q-in-Q infrastructure in the kernel is supported only in kernel version 3.10 and up.	-
24.	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.	-
25.	When using a hardware that has Time Stamping enabled, the system time might be higher than the expected variance.	-
26.	In Q-in-Q, ping failed when sending traffic with package size > 1468	-
27.	Call trace may occur when configuring VXLAN or under high traffic stress.	-
28.	HW LRO does not function in ConnectX-4 adapter cards.	-

Table 13 - Ethernet Known Issues

Index	Internal Reference Number: Description	Workaround
29.	ethtool header does not currently support the link speeds of 25/50/100. Therefore, these speeds cannot be seen as advertised/supported.	-
30.	#835239: While running Q-in-Q packets with stag offloading, tcpdump/wireshark on host may show svlan ethertype as 0x8100 instead of 0x88A8.	Check the wire or a switch between the hosts, the wireshark will show 0x88A8 ethertype as expected.

3.4.2 Port Type Management Known Issues

Table 14 - Port Type Management Known Issues

Index	Internal Reference Number: Description	Workaround
1.	After changing port type using <code>connectx-_port_config</code> interface ports' names can be changed. For example. <code>ib1</code> -> <code>ib0</code> if port1 changed to be Ethernet port and port2 left IB.	Use <code>udev</code> rules for persistent naming configuration. For further information, please refer to the User Manual
2.	OpenSM must be stopped prior to changing the port protocol from InfiniBand to Ethernet.	-
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 <code>::ffff:<Multicast IPv4 Address></code>	-
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	<code>connectx_port_config</code> configurations is not saved after unbind/bind.	Re-run " <code>connectx_port_config</code> "

3.4.3 Flow Steering Known Issues

Table 15 - Flow Steering Known Issues

Index	Internal Reference Number: Description	Workaround
1.	Flow Steering is disabled by default in firm-ware version < 2.32.5100.	To enable it, set the parameter below as follow: <code>log_num_mgm_entry_size</code> 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 16 - Quality of Service Known Issues

Index	Internal Reference Number: 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, <code>mlnx_qos</code> tool may crash.	-
3.	QoS default settings are not returned after configuring QoS.	-

3.4.5 Ethernet Performance Counters Known Issues

Table 17 - Ethernet Performance Counters Known Issues

Index	Internal Reference Number: 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.	-

Table 17 - Ethernet Performance Counters Known Issues

Index	Internal Reference Number: Description	Workaround
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 18 - IPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#854235: IPoIB bonding interface has to be restarted in order to work on some operating systems.	Toggle bonding interface to state “down” and then to state “up”.
2.	When user increases receive/send a buffer, it might consume all the memory when few child's interfaces are created.	-
3.	The size of send queue in Connect-IB® cards cannot exceed 1K.	-
4.	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.	-
5.	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 <i>than NetworkManager.conf</i>

Table 18 - IPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
6.	IPoIB interface does not function properly if a third party application changes the PKey table.	We recommend modifying PKey tables via OpenSM.
7.	Fallback to the primary slave of an IPoIB bond does not work with ARP monitoring. (https://bugs.openfabrics.org/show_bug.cgi?id=1990)	-
8.	Out-of memory issue might occur due to overload of interfaces created.	To calculate the allowed memory per each IPoIB interface check the following: <ul style="list-style-type: none"> • 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
9.	Connect-IB does not reach the bidirectional line rate	Optimize the IPoIB performance in Connect-IB: <pre>cat /sys/class/net/<interface>/device/local_cpus > /sys/class/net/<interface>/queues/rx-0/rps_cpus</pre>
10.	If the CONNECTED_MODE 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.

Table 18 - IPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
11.	<p>Whenever the IOMMU parameter is enabled in the kernel it can decrease the number of child interfaces on the device according to resource limitation.</p> <p>The driver will stuck after unknown amount of child interfaces creation.</p> <p>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</p>	<p>To avoid such issue:</p> <ul style="list-style-type: none"> • 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 <p>For KVM users: Run: <pre>echo 1 > /sys/module/kvm/parameters/allow_unsafe_assigned_interrupts</pre></p> <p>To make this change persist across reboots, add the following to the <code>/etc/modprobe.d/kvm.conf</code> file (or create this file, if it does not exist): <pre>options kvm allow_unsafe_assigned_interrupts=1 kernel parameters</pre></p>
12.	<p>System might crash in <code>skb_checksum_help()</code> while performing TCP retransmit involving packets with 64k packet size. A similar out to the below will be printed:</p> <pre>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</pre>	<p>Use UD mode in IPoIB</p>

Table 18 - IPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
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 <code>/etc/udev/rules.d/70-persistent-net.rules</code> file: <pre>SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="*<Port GID>", NAME="ibN"</pre> <p>Where N is the IPoIB required interface index</p>
14.	After releasing a bond interface that contains IPoIB slaves, a call trace might be printed to the <code>dmesg</code> .	-
15.	IPoIB interfaces are loaded without an IP address on SLES 12.	1. Open the <code>/etc/wicked/common.xml</code> file. 2. Change: <pre>"<use-nanny>>false</use-nanny>" to "<use-nanny>>true</use-nanny>"</pre> 3. Run: <pre># systemctl restart wickedd.service wicked # ifup all</pre>
16.	In RHEL7.0, running <code>ifdown</code> then <code>ifup</code> on an interface after changing <code>CONNECTED_MODE</code> in its <code>ifcfg</code> file, will cause the interface bring up to fail.	Reload the driver <code>"/etc/init.d/openibd restart"</code> 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.	<code>eth_ipoib</code> module is not loaded after reloading the <code>ib_ipoib</code> module.	To restart the IPoIB driver, run <code>"/etc/init.d/openibd restart"</code> . Do not restart it by manually restarting each module.
19.	In Ubuntu and Debian, the default of the <code>recv_queue_size</code> and <code>send_queue_size</code> is 128 according to the <code>io_mmu</code> issue.	-

Table 18 - iPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
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.	Kernel Oops may occur after reboot.	-
23.	Kernel panic may occur while re-assigning LIDs.	-
24.	ICMP traffic might be lost after Vnic restart	-
25.	Spikes may occur while running PTP protocol over ConnectX-3/ConnectX-3 Pro.	-
26.	: ifdown fails on SLES12SP0/SP1 with the following errors # ifdown ib0 wicked: ifdown: no matching interfaces The error indicates that there are active interfaces using the interface you are trying to bring down, and you must ifdown all dependent interfaces.	To see the list of all dependent interfaces, run: # wicked --debug all ifdown ib0 wicked: skipping ib0 interface: unable to ifdown due to lowerdev dependency to: ib0.8001 wicked: ifdown: no matching interfaces wicked: Exit with status: 0
27.	: Occasionally, in kernel 3.10, under heavy load, the kernel fails to get free page. For more details, please refer to: https://bugs.centos.org/view.php?id=10245	-

3.5.2 eIPoIB Known Issues

Table 19 - eIPoIB Known Issues

Index	Internal Reference Number: 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.

Table 19 - eIPoIB Known Issues

Index	Internal Reference Number: Description	Workaround
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 20 - XRC Known Issues

Index	Internal Reference Number: 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 21 - Verbs Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#855362: A compilation error will occur in kernel space application when setting the <code>wr_id</code> field upon initializing any of the following structures: <code>ib_wc</code> , <code>ib_send_wr</code> , or <code>ib_recv_wr</code> . This is caused due to <code>wr_id</code> insertion into an anonymous union.	Assign the enum field explicitly. For example: <code>wr.wr_id = MY_WR_ID;</code>
2.	Using <code>libnl1_1_3~26</code> or earlier, requires <code>ibv_create_ah</code> protection by a lock for multi-threaded applications.	-
3.	In <code>MLNX_OFED v2.4-1.0.0</code> , 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.
4.	<code>ibv_task_pingpong</code> over ConnectX-2 adapter cards is not supported.	-
5.	According to the verbs header (<code>/usr/include/infiniband/verbs.h</code>), the static rate field in the Address handler can take value from 0 to 18. The values 11 to 18 (inclusive) are not supported for Connect-X 4 and Connect-X 3.	Run QP command query to verify the value.

3.5.5 RoCE Known Issues

Table 22 - RoCE Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#869158: Occasionally, UC UD traffic over default GIDs, with high iterations may get stuck.	-
2.	#854517: Driver restart while having RDMA-CM running applications may hang.	-
3.	#866072: [mlx5] RoCE v2 multicast traffic using RDMA-CM with IPv4 address will not be received.	-
4.	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
5.	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.	-
6.	On rare occasions, the driver reports a wrong GID table (read from /sys/class/infiniband/mlx4_*/ports/*/gids/*). This may cause communication problems.	-
7.	MLNX_OFED v2.1-1.0.0 and onwards is not interoperable with older versions of MLNX_OFED.	-
8.	<p>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:</p> <ul style="list-style-type: none"> • "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) <p>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.</p>	-

Table 22 - RoCE Known Issues

Index	Internal Reference Number: Description	Workaround
9.	A working IP connectivity between the RoCE devices is required when creating an address handle or modifying a QP with an address vector.	-
10.	IPv4 multicast over RoCE requires the MGID format to be as follow ::ffff:<Multicast IPv4 Address>	-
11.	IP RoCEv2 does not support Multicast Listener Discovery (MLD) therefore, multicast traffic over IPv6 may not work as expected.	-
12.	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 of the presence of the IPv6 address.	-
13.	Using IPv6 link local address (GID0) when VLANs are configured is not supported.	-
14.	Using GID index 0 (the default GID) on port 2 is currently not supported on kernel 3.14 and below.	-
15.	Dynamically Connected (DC) in RoCE in ConnectX®-4 is currently not supported.	-
16.	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.
17.	ibv_create_ah_from_wc is not supported for multicast messages.	-
18.	Occasionally, when the Bonding Mode is set to other than active/backup mode (mode 1), the GID table is not populated correctly.	Add slave devices to the master before giving it an IP address.
19.	when the port speed is lower than 10Gbps, the IB tools will present a higher rate.	-

Table 22 - RoCE Known Issues

Index	Internal Reference Number: Description	Workaround
20.	#778492: RoCE requires that when a bonding module enslaves 2 Ethernet interfaces, the GID for any IP address on bond0 will appear only on the port of the active interface Due to kernel limitations, the information about active slave is unknown, therefore, any IP address on bond0 will appear on both ports.	Work in fail_over_mac mode (bonding).
21.	#781383: Creating Address Handler (AH) may run slow or may hang under a heavy load on all nodes' cores (for example: MPI All2All cases).	-

3.5.6 iSCSI over IPoIB Known Issues

Table 23 - iSCSI over IPoIB Known Issues

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

3.5.7 SDP Known Issues

Table 24 - SDP Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#664110: SDP is currently not supported in mlx5 driver (Connect-IB and Connect-X 4 adapter cards)	-

3.6 Storage Protocols Known Issues

3.6.1 Storage Known Issues

Table 25 - Storage Known Issues

Index	Internal Reference Number: Description	Workaround
1.	Older versions of <code>rescan_scsi_bus.sh</code> may not recognize some newly created LUNs.	If encountering such issues, it is recommended to use the '-c' flag.

3.6.2 NVMeoF Host/Target Known Issues

Table 26 - NVMeoF Host Known Issues

Index	Internal Reference Number: Description	Workaround
1.	NVMeoF support is available only for kernels of version 4.8.x.	

3.6.3 SRP Known Issues

Table 27 - SRP Known Issues

Index	Internal Reference Number: Description	Workaround
1.	SRP daemon does not start at boot.	Add “service srpd start” to rc.local or start it manually.
2.	srp_daemon fails to connect on ConnectX-4 VF.	-
3.	MLNX_OFED SRP installation breaks the ibmvstgt and ibmvscsi symbol resolution in RHEL7.0	-

3.6.4 SRP Interop Known Issues

Table 28 - SRP Interop Known Issues

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

3.6.5 DDN Storage Fusion 10000 Target Known Issues

Table 29 - DDN Storage Fusion 10000 Target Known Issues

Index	Internal Reference Number: Description	Workaround
1.	DDN does not accept non-default P_Key connection establishment.	-

3.6.6 Oracle Sun ZFS Storage 7420 Known Issues

Table 30 - Oracle Sun ZFS Storage 7420 Known Issues

Index	Internal Reference Number: 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.	-

3.6.7 iSER Initiator Known Issues

Table 31 - iSER Initiator Known Issues

Index	Internal Reference Number: Description	Workaround
1.	On SLES OSs, the <code>ib_iser</code> module does not load on boot.	Add a dummy interface using <code>iscsiadm</code> : <ul style="list-style-type: none"> <code># iscsiadm -m iface -I ib_iser -o new</code> <code># iscsiadm -m iface -I ib_iser -o update -n iface.transport_name -v ib_iser</code>
2.	Ubuntu12 requires update of user space <code>open-iscsi</code> 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.	Ubuntu14.04: Stress login/logout might cause block layer to invoke a WARN trace.	-
8.	iSER small read IO (< 8k) performance degrades compared to previous versions. iSER performs memory registration for each IO and avoids sending a global memory key to the target. Sending the global memory key to the wire should only be done in a trusted environment and is not recommended to use over the Internet protocol.	Set module param <code>always_register=N</code> <code>\$ modprobe ib_iser always_register=N</code>

3.6.8 iSER Target Known Issues

Table 32 - iSER Target Known Issues

Index	Internal Reference Number: Description	Workaround
1.	iSER Target currently supports only the following OSs (distribution kernel): <ul style="list-style-type: none"> RHEL 7.0/7.1/7.2 SLES12/12.1 Ubuntu14.04, Ubuntu15.04 	-

Table 32 - iSER Target Known Issues

Index	Internal Reference Number: Description	Workaround
2.	RHEL/CentOS 7.0: Discovery over RDMA is not supported.	-
3	ib_isert is unavailable on custom kernels after running the <code>mlnx_add_kernel_support.sh</code> script.	<ol style="list-style-type: none"> 1. Add "isert=y" to the <code>mlnx_add_kernel_support.sh</code> script after "cat << EOF > ofed.conf". 2. Use the updated script to build MLNX_OFED for the custom kernel.

3.6.9 ZFS Appliance Known Issues

Table 33 - ZFS Appliance Known Issues

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

3.6.10 Erasure Coding Verbs Known Issues

Table 34 - Erasure Coding Verbs Known Issues

Index	Internal Reference Number: Description	Workaround
1.	The Erasure-coding logical block size must be aligned to 64 bytes	-
2.	Only $w=1,2,3,4$ are supported (w corresponds to the Galois symbol size - $GF(2^w)$)	-
3.	<p><code>ibv_exp_ec_mem</code> must pass with the following restrictions:</p> <ul style="list-style-type: none"> • <code>num_data_sge</code> must be equal to K (property of the EC calc) • <code>num_code_sge</code> must be equal to M (property of the EC calc) 	-

3.7 Virtualization

3.7.1 SR-IOV Known Issues

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
1.	#858628: PCI error handling is not supported during driver reload. This might cause a kernel panic or calltrace.	-
2.	#860385: Creating 127 VFs may cause kernel panic in SLES11 SP4 KVM with Kernel 3.0.101-63 because of a IOMMU kernel bug.	-
3.	#866875: During VM shutdown, kernel panic may occur as a result of using the <code>ndo_get_phys_port_id</code> callback during shutdown.	-
4.	#822781: SR-IOV is not supported in systems with a page size greater than 16KB since this is the maximal VF uar size supported.	-
5.	#795697: [mlx4] While spoof-check filters the incoming traffic to a VM, when this feature is disabled, traffic still does not reach the VM.	The driver must be restarted for the disablement of the feature to take effect and all traffic to be reached to the VM.
6.	#791101: [mlx4] Spoof-check may be turned on for MAC address 00:00:00:00:00:00	-
7.	#835065: [mlx5] When working with InfiniBand QoS, the bandwidth for VFs that are attached to VMs might not be spread according to the QoS configuration if not enough cores are assigned to the VM.	-
8.	#784940: Currently, the firmware cannot process many page requests in parallel as the driver processes page requests serially. Therefore, enabling/disabling a large number of VFs will often cause an driver slowdown.	-
9.	#784954: When SR-IOV is disabled, the VF driver receives <code>pci_err_detected</code> event and a teardown flow will be started. During the teardown flow, all firmware commands will fail because the function is already deleted.	-

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
10.	#819595: [ConnectX-3 Pro] In case a VF is set to VST mode on the same port following QinQ configuration, that VF will insert C-VLAN not only to untagged packets, but also to tagged packets. The packets that are tagged twice will be dropped by the switch or by the destination host since they have two C-VLANs.	-
11.	#775944: Bonding VFs on the same physical port using bonding mode 0 requires configuration of <code>fail_over_mac=1</code> .	-
12.	When using legacy VMs with MLNX_OFED 2.x hypervisor, you may need to set the <code>'enable_64b_cqe_eqe'</code> parameter to zero on the hypervisor. It should be set in the same way that other module parameters are set for <code>mlx4_core</code> at module load time. For example, add <code>"options mlx4_core enable_64b_cqe_eqe=0"</code> as a line in the file <code>/etc/modprobe.d/mlx4_core.conf</code> .	-
13.	#381754: <code>mlx4_port1_mtu</code> sysfs entry shows a wrong MTU number in the VM.	-
14.	#426988: When at least one port is configured as InfiniBand, and the <code>num_vfs</code> is provided but the <code>probe_vf</code> is not, HCA initialization fails.	Use both the <code>num_vfs</code> and the <code>probe_vf</code> in the modprobe line.
15.	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 <code>mlx4_ib</code> and reload it in the VM.
16.	Attaching or detaching a Virtual Function on SLES11 SP3 to a guest Virtual Machine while the <code>mlx4_core</code> driver is loaded in the Virtual Machine may cause a kernel panic in the hypervisor.	Unload the <code>mlx4_core</code> module in the hypervisor before attaching or detaching a function to or from the guest.
17.	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.

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
18.	Enabling SR-IOV requires appending the "intel_iommu=on" option to the relevant OS in file /boot/grub/grub.conf or /boot/grub2/grub.cfg, depending on the OS installed. Without that SR-IOV cannot be loaded.	-
19.	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.
20.	The known PCI BDFs for all VFs in kernel command line should be specified by adding <code>xen-pciback.hide</code> For further information, please refer to http://wiki.xen.org/wiki/Xen_PCI_Passthrough	-
21.	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.	-
22.	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.	-
23.	Attempting to attach a PF to a VM when SR-IOV is already enabled on that PF may result in a kernel panic.	-
24.	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.	-
25.	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.	-

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
26.	QPerf test is not supported on SR-IOV guests in Connect-IB cards.	-
27.	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.	-
28.	Occasionally, the <code>lspci grep Mellanox</code> command shows incorrect or partial information due to the current <code>pci.ids</code> file on the machine.	<ol style="list-style-type: none"> 1. Locate the file: <code>\$locate pci.ids</code> 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 (using the following command: <code>update-pci-ids</code>).
29.	SR-IOV is not supported in AMD architecture.	-
30.	Setting 1 Mbit/s rate limit on Virtual Functions (Qos Per VF feature) may cause TX queue transmit timeout.	-
31.	DC transport type is not supported on SR-IOV VMs.	-
32.	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.	-
33.	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.	-
34.	Since the guest MAC addresses are configured to be all zeroes by default, in ConnectX-4 the administrator must explicitly set the VFs' MAC addresses. otherwise the Guest VM will see MAC zero and traffic is not passed.	-
35.	Restarting the PF (Hypervisor) driver while Virtual Functions are assigned is not allowed in RH7 and above due to a <code>vfio-pci</code> bug.	-
36.	Due to an issue with SR-IOV loopback, prevention "Duplicate IPv6 detected" are seen in the VF driver.	-

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
37.	: [ConnectX-4/Connect-IB] Failed to enable SR-IOV due to errors in PCI or BIOS.	1. Add <code>pci=realloc=on</code> to the grub command line. 2. Add more memory to the server. 3. Upgrade BIOS version.
38.	Kernel panic may occur while running IPv6 UDP on SR-IOV ConnectX-4 environment	-
39.	Bind/Unbind over ConnectX-4 Hypervisor may cause system lockup.	-
40.	Occasionally, IPv6 might not function properly and cause lockup on SR-IOV ConnectX-4 environment.	-
41.	In ConnectX-3 adapter cards, the extended counter <code>port_rcv_data_64</code> on the VF may not be updated in some flows.	-
42.	690674 When the physical link is down, any traffic from the PF to any VF on the same port will be dropped.	-
43.	When in LAG mode and the Virtual Functions are present (VF LAG), the IP address given to the bonding interface (in the hypervisor) cannot be used for RoCE as well.	Probe one of the VFs in the hypervisor and use for RoCE.
44.	Ethernet SR-IOV in ConnectX-4 requires firmware version 12.14.1100 and higher	-
45.	VF vport statistics are not cleared upon <code>ifconfig up/down</code> .	-
46.	In SLES11 SP4, user cannot open all VFs announced in <code>sriov_totalvfs</code> . However he can set the <code>num_vfs</code> up to maximum <code>sriov_totalvfs-1</code> vfs.	-
47.	#784127: While disabling SR-IOV, all firmware teardown flow commands are expected to fail and error messages will be reported in the <code>dmesg</code> .	-
48.	#784146: Creating/destroying as many as 64 VFs may sometimes take longer time than usual on some setups.	-

Table 35 - SR-IOV Known Issues

Index	Internal Reference Number: Description	Workaround
49.	#766105: Due to a bug in some QEMU versions, interrupts do not function properly for Virtual Functions. This causes the driver initialization to fail, and such error message will be printed: "mlx4_core 0000:0b:00.0: command 0x31 timed out (go bit not cleared) mlx4_core 0000:0b:00.0: NOP command failed to generate interrupt (IRQ 57), aborting".	Upgrade to the latest version of QEMU in the hypervisor.
50.	When working with InfiniBand QoS, the bandwidth for Virtual Functions (VF) that are attached to Virtual Machines (VM) might not be spread according to the QoS configuration if not enough cores are assigned to the Virtual Machine.	-

3.8 Resiliency

3.8.1 Reset Flow Known Issues

Table 36 - Reset Flow Known Issues

Index	Internal Reference Number: Description	Workaround
51.	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
52.	Upon Reset Flow or after running restart driver, Ethernet VLANs are lost.	Reset the VLANs using the <code>ifup</code> command.
53.	Restarting the driver or running <code>connectx-port_config</code> when Reset Flow is running might result in a kernel panic	-
54.	Networking configuration (e.g. VLANs, IPv6) should be statically defined in order to have them set after Reset Flow as of after restart driver.	-
55.	After recovering from an EEH event, <code>mlx-5_core/mlx4_core</code> 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 37 - General Known Issues

Index	Internal Reference Number: Description	Workaround
1.	<p>#856033: The following PCIe bus error on Qualcomm ARM processor might appear when mapping a large number of DMA addresses:</p> <pre> "AER: Corrected error received: id=0000 PCIe Bus Error: severity=Corrected, type=Transaction Layer, id=0000(Receiver ID) device [17cb:0400] error status/ mask=00002000/00004000 [13] Advisory Non-Fatal mlx5_warn:mlx5_0:dump_cqe:257:(pid 0): dump error cqe 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 00000000 12007806 25000063 8728c8d3" </pre>	<ol style="list-style-type: none"> Edit the kernel parameters (in grub) and add <code>qiommu.identity_map_qiommu=PCIE0_MMU,PCIE4_MMU</code> (The bus numbers depend on the ConnectX-4 slot.) Reboot the server.
3.	<p>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 <code>ibstat</code>). <code>Mlxburn/flint</code> return <code>0xffff</code> 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.</p>	<p>N/A. Please use the GUID value returned by the fabric/driver utilities (not <code>0xffff</code>).</p>
4.	<p>On rare occasions, under extremely heavy MAD traffic, MAD (Management Datagram) storms might cause soft-lockups in the UMAD layer.</p>	-
5.	<p>Packets are dropped on the SM server on big clusters.</p>	<p>Increase the <code>recv_queue_size</code> of <code>ib_mad</code> module parameter for SM server to 8K. The <code>recv_queue_size</code> default size (4K)</p>
6.	<p>On ConnectX-4/ConnectX-4 Lx, when running "<code>lspci</code>" in RH7.0/7.1, the device information is displayed incorrect or the device is unnamed.</p>	<p>Run <code>update-pciids</code></p>

Table 37 - General Known Issues

Index	Internal Reference Number: Description	Workaround
7.	Resetting hardware counters after netdev goes up can break statistics scripts.	-

3.9.2 ABI Compatibility Known Issues

Table 38 - ABI Compatibility Known Issues

Index	Internal Reference Number: 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 39 - Connection Manager (CM) Known Issues

Index	Internal Reference Number: 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.
2.	#781382: The number of local ports that rdma_cm ID can bind to is limited. This limitation depends on the OS dynamics.	<p>Modify the range of available ports for binding, run:</p> <pre>sysctl net.ipv4.ip_local_port_range="MIN MAX"</pre> <p>The MIN and MAX values can range from 0 to 65535.</p> <p>Note: Modifying the range also affects the range of available ports for socket applications (TCP/IP) even though the pool is not mutual between the RDMA stack and the TCP/IP stack.</p>

3.9.4 Fork Support Known Issues

Table 40 - Fork Support Known Issues

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

3.9.5 Uplinks Known Issues

Table 41 - Uplinks Known Issues

Index	Internal Reference Number: 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

3.9.6 Resources Limitation Known Issues

Table 42 - Resources Limitation Known Issues

Index	Internal Reference Number: 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 user-space.	-
2.	<code>mlx4_core</code> 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 42 - Resources Limitation Known Issues

Index	Internal Reference Number: 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.	<p>To free memory to allow it to be allocated in a user process, run the <code>drop_caches</code> 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.</p> <p>Note: As this is a non-destructive operation and dirty objects are not freeable, run <code>`sync'</code> first.</p> <ul style="list-style-type: none"> To free the pagecache: <pre>echo 1 > /proc/sys/vm/drop_caches</pre> To free dentries and inodes: <pre>echo 2 > /proc/sys/vm/drop_caches</pre> To free pagecache, dentries and inodes: <pre>echo 3 > /proc/sys/vm/drop_caches</pre>

3.9.7 Accelerated Verbs Known Issues

Table 43 - Accelerated Verbs Known Issues

Index	Internal Reference Number: Description	Workaround
1.	<p>On ConnectX@-4 Lx, the following may not be supported when using Multi-Packet WR flag (<code>IBV_EXP_QP_BURST_CREATE_ENABLE_MULTI_PACKET_SEND_WR</code>) on QP-burst family creation:</p> <ul style="list-style-type: none"> 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 44 - Performance Tools Known Issues

Index	Internal Reference Number: 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 45 - Diagnostic Utilities Known Issues

Index	Internal Reference Number: Description	Workaround
1.	When running the ibdiagnet check <code>nodes_info</code> 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 <code>ibdiagnet --skip nodes_info</code>
2.	<code>ibdump</code> 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.	#736136: The maximum number of HCAs shown by <code>ibstat</code> is 32 HCAs.	-

3.10.3 Tools Known Issues

Table 46 - Tools Known Issues

Index	Internal Reference Number: Description	Workaround
1.	Running <code>ibdump</code> in InfiniBand mode with firmware older than v2.33.5000, may cause the server to hang due to a firmware issue.	Run <code>ibdump</code> with firmware v2.33.5000 and higher

4 Bug Fixes History

This table lists the bugs fixed in this release.

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
1.	ODP	#980367: Fixed the issue where a misleading error message about failure to handle ODP page fault used to appear in Syslog.	3.4-2.1.4.1	3.4-2.1.8.0
2.		#778363: Fixed an issue in implicit ODP garbage collection mechanism, which enabled access to a released memory region.	3.4-2.1.4.1	3.4-2.1.8.0
3.		#975614: Fixed an issue in memory region length calculation which caused an overflow in ODP memory registration.	3.4-2.1.4.1	3.4-2.1.8.0
4.		#974087: Fixed an issue where during driver unload, the QP used for ODP memory registrations remained active after being destroyed.	3.4-2.1.4.1	3.4-2.1.8.0
5.	biosdav-ename	#873538: Fixed the issue where biosdav-ename running on Redhat 6.x with OFED may show the same name to ConnectX-3 Eth port 1 and ConnectX-3 Eth port 2.	3.4-1.0.0.0	3.4-2.0.0.0
6.	mlx5 Driver	#876329: Fixed the issue of when the error flow was re-factored, the reading of the device caps was excluded from the error recovery flow.	3.4-1.0.0.0	3.4-2.0.0.0
7.	mlx4_en	#876419: Fixed the issue where kernel panic was observed on openibd stop as a result of querying non-existent bond slave.	3.3-2.0.0.0	3.4-2.0.0.0
8.		#868665: Fixed the issue where kernel panic in <code>mlx4_en_get_phys_port_id</code> may occur during server reboot.	3.3-1.0.0.0	3.4-2.0.0.0
9.		#824855: Fixed the issue of packet drop on UDP unidirectional multi streams, which used to be reported with the <code>vport_rx_dropped</code> counter.	3.0-1.0.1	3.4-2.0.0.0
10.		#882227: Fixed the issue of when EEH was injected and the <code>mlx4</code> tear down code was called, the eqs were not released, causing a page fault.	3.4-1.0.0.0	3.4-2.0.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
11.	mlx5_core	#887348: Fixed the issue of when <code>prof_sel</code> was invalid, <code>mlx5_core</code> failed upon debug print.	3.4-1.0.0.0	3.4-2.0.0.0
12.		#898161: Fixed the issue where a compilation error in kernels of v4.6 or above used to occur due to a large stack size in the <code>get_numa_phys_mask</code> function.	3.4-1.0.0.0	3.4-2.0.0.0
13.	ib_core	#880269: Fixed the issue of when OFED was run on kernel v4.6 or higher, in which a memory management subsystem change was embedded, a kernel failure used to occur.	3.4-1.0.0.0	3.4-2.0.0.0
14.	ib_isert	#887245: Fixed the issue where the system used to pick the dummy <code>ib_isert</code> module instead of the real module on RHEL with errata kernel.	3.4-1.0.0.0	3.4-2.0.0.0
15.	irqbalancer	#854344: Fixed the issue where <code>mlnx_affinity</code> script on RHEL/CentOS7.x host did not disable or enable <code>irqbalancer</code> .	3.3-1.0.0.0	3.4-1.0.0.0
16.	QoS	#824736: Fixed wrong <code>skprio2UP</code> mapping by removing it and its scripts, such as <code>tc_wrap</code> , from the driver. This mapping should now be done using the kernel's <code>set_egress_map</code> commands. Note: Only for RDMACM over old kernels, the original <code>skprio2UP</code> mapping in <code>tc_wrap</code> remains valid as these kernels do not support <code>set_egress_map</code> .	3.3-1.0.0.0	3.4-1.0.0.0
17.	ibacm	#824775: Fixed the issue where starting <code>ibacm</code> daemon failed on Debian based distributions with the following message: “/etc/init.d/ibacm: line 37: /sbin/start_daemon: No such file or directory”.	3.3-1.0.0.0	3.4-1.0.0.0
18.	IPoIB	#799004: Fixed the issues of when establishing IPoIB CM connection, a race could occur if there were many CM connections taking place while the driver was going up and down. This race in the IPoIB driver could have caused memory corruption.	3.0-2.0.0.0	3.4-1.0.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
19.	ARM	#777733/778099: Fixed the issue where in ARM architecture, multiple kernel panics of mlx4 and mlx5 drivers were observed as a result of undefined behavior of vmap(virt_to_page(dma_alloc_coherent)) call sequence on driver load, by allocating contiguous memory instead of vmapping it.	3.3-1.0.0.0	3.4-1.0.0.0
20.	mlx4_en	#826686: Fixed the issue where server reboot could get stuck because of kernel panic in mlx4_en_get_drvinfo() that is called from asynchronous event handler.	3.3-1.0.0.0	3.4-1.0.0.0
21.		#824130: Fixed the issue where ethtool self test used to fail on interrupt test after timeout if mlx4_ib module was not loaded.	3.3-1.0.0.0	3.4-1.0.0.0
22.	mlx4 driver	#855311: Fixed the issue of when using RDMA READ with a higher value than 30 SGEs in the WR, this might have lead to “local length error”.	3.0-1.0.1	3.4-1.0.0.0
23.	mlx5 driver	#786720: Fixed a crash that used to occur when trying to bring the interface up in a kernel that did not support accelerated RFS (aRFS).	3.3-1.0.0.0	3.4-1.0.0.0
24.	SR-IOV	#781747: Fixed the issue of when attempting to disable SR-IOV while there are any VF net-devs open, the operation would fail and the driver would hang.	3.3-1.0.0.0	3.4-1.0.0.0
25.		#568602: Fixed the issue of when repeating change of the mlx5_num_vfs value from 0 to non-zero might have caused kernel panic in the PF driver.	3.0-2.0.0	3.4-1.0.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
26.	IPoIB	#814941: Fixed the issue where as a result of the <code>ndo_set_mac_address</code> support in IPoIB, a memory corruption issue was exposed in the bonding driver over IPoIB interface. This memory corruption issue could have caused unpredictable behavior, such as ports becoming dysfunctional in RedHat, kernel panicking in Ubuntu devices, and other behaviors.	3.3-1.0.0	3.3-1.0.4.0
27.	IB Core	#781183: Fixed a potential security breach in the InfiniBand stack that was caused due to wrong reliance on the write system call. For more information, please see CVE-2016-4565 tracker.	3.2-2.0.0	3.3-1.0.0
28.	RoCE	#592652: Fixed the issue where the InfiniBand error counters found under <code>/sys/class/infiniband/<mlx5_dev>/ports/<port>/</code> did not function properly in ConnectX-4 adapter cards.	3.1-0.0.7	3.3-1.0.0
29.	Installation	#765162: Fixed the issue where <code>dapl</code> package was missing in <code>MLNX_OFED</code> for Ubuntu PPC64LE.	2.0.2.0.5	3.3-1.0.0.0
30.	TX Queue Counter	#748308: Changed TX queue counter format to: <code>xq_[tc]*[ring/channel]</code> .	3.2-2.0.0	3.3-1.0.0.0
31.	RDMA Sniffer	#751097: Fixed RDMA sniffer functionality issues.	3.2-2.0.0	3.3-1.0.0.0
32.	IPoIB	Fixed IPoIB Connected Mode in ConnectX-3 functionality issues.	3.2-2.0.0	3.3-1.0.0.0
33.		#769688: Fixed the issue where in order to change the IPoIB mode (connected/datagram), the interface had to be taken down (via <code>ifconfig ibX down</code> or <code>ifdown ibX</code>). Now, the mode can be changed regardless of the interface's state ("up" or "down").	3.2-2.0.0	3.3-1.0.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
34.	mlx4_en	Added DCB PFC support through CEE netlink commands to prevent Priority Flow Control mode functionality issues on the host side.	3.2-2.0.0	3.3-1.0.0.0
35.	SR-IOV	Fixed an issue which added error messages to the dmesg when a VF used ethtool facilities.	3.1-1.0.5	3.3-1.0.0.0
36.		Fixed an issue which caused any traffic from PF to any VF on the same port to drop when the physical link was down.	3.2-1.0.1.1	3.3-1.0.0.0
37.	mlx5 driver	Fixed kernel's back-ports of XPS and affinity that did not have CONFIG_CPUMASK_OFFSTACK	3.2-1.0.1.1	3.2-2.0.0.0
38.		Added support for Rate Limit 0 to enable unlimited rate limiter and to prevent max rate zero traffic lose.	3.2-1.0.1.1	3.2-2.0.0.0
39.	SR-IOV	Fixed an issue which enabled SR-IOV on RHEL 6.7 although SR-IOV was already enabled. A check was added to make sure SR-IOV is not enabled before enabling it.	3.2-1.0.1.1	3.2-2.0.0.0
40.	eIPoIB	Fixed race between the udev that changes the interface name of eth_ipoib driver and the eIPoIB daemon that configured the same interface.	3.0-1.0.1	3.2-2.0.0.0
41.	Ethernet traffic	from running Ethernet traffic on Big Endian arch machines.	3.2-1.0.1.1	3.2-2.0.0.0
42.	Performance	close NUMA node as default for RSS.	3.2-1.0.1.1	3.2-2.0.0.0
43.	mlx4_en	Fixed an issue where the ARP request packets destined for a proxy VXLAN interface were not handled correctly when GRO was enabled.	3.2-1.0.1.1	3.2-2.0.0.0
44.	Counters	Fixed an issue which prevented the calculated software counters (the correct ones) from being shown and provided the error counters that were previously inactive.	3.0-1.0.1	3.2-2.0.0.0
45.	Virtualization	Fixed an issue which prevented the driver from reaching VLAN when the VLAN was created over a Linux bridge.	3.1-1.0.3	3.2-1.0.1.1

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
46.	mlx5 driver	# 656298: Fixed an issue in the driver (in ConnectX-4) that discarded s-tag VLAN packets when in Promiscuous Mode.	3.1-1.0.3	3.2-1.0.1.1
47.		# 647865: Fixed an issue which prevented PORT_ERR event to be propagated to the user-space application when the port state was changed from Active to Initializing.	3.0-1.0.1	3.2-1.0.1.1
48.	HPC Acceleration packages	# 663975: Fixed a rare issue which allowed the knem package to run depmod on the wrong kernel version.	3.1-1.0.3	3.2-1.0.1.1
49.	IB/Core	# 666992: Fixed a race condition in the IB/umad layer that caused NULL pointer dereference.	3.0-2.0.1	3.2-1.0.1.1
50.	IPoIB	# 657718: Fixed an IPoIB issue that caused connectivity lost after server's restart in a cluster.	3.1-1.0.3	3.2-1.0.1.1
51.	Driver un-installation	# 619272: Fixed an issue causing MLNX_OFED to remove the "mutt" package upon driver uninstall.	3.1-1.0.3	3.2-1.0.1.1
52.	PFC	# 613514: Added a warning message in dmesg, notifying the user that the PFC RX/TX cannot be enabled simultaneously with Global Pauses. In this case Global Pauses will be disabled.	3.1-1.0.3	3.2-1.0.1.1
53.	IB MAD	Fixed an issue causing MADs to drop in large scale clusters.	3.1-1.0.0	3.1-1.0.3
54.	SR-IOV	Fixed InfiniBand counters which were unavailable in the VM.	2.1-1.0.0	3.1-1.0.0
55.	RoCE	Fixed InfiniBand traffic counters that are found under <code>/sys/class/infiniband/<mlx-5_dev>/ports/<port>/</code> which do not function properly in ConnectX-4 adapter cards.	3.0-1.0.1	3.1-1.0.0
56.	Virtualization	Fixed VXLAN functionality issues.	3.0-2.0.1	3.1-1.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
57.	Performance	TCP/UDP latency on ConnectX®-4 was higher than expected.	3.0-2.0.1	3.1-1.0.0
58.		TCP throughput on ConnectX®-4 achieved full line rate.	3.0-2.0.1	3.1-1.0.0
59.		Fixed an issue causing inconsistent performance with ConnectX-3 and PowerKVM 2.1.1.	3.0-2.0.1	3.1-1.0.0
60.		Fixed ConnectX-4 traffic counters.	3.0-2.0.1	3.1-1.0.0
61.	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
62.	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
63.	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
64.	openibd	Fixed an issue which prevented openibd from starting correctly during boot.	3.0-2.0.1	3.1-1.0.0
65.	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
66.	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
67.	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
68.	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
69.		Added the option to toggle LRO ON/OFF using the "-K" flags.	3.0-2.0.1	3.1-1.0.0
70.		Fixed race when updating counters.	3.0-2.0.1	3.1-1.0.0
71.		Fixed scheduling while sending atomic dmesg warning during bonding configuration.	3.0-2.0.1	3.1-1.0.0
72.		Added set_rx_csum callback implementation.	3.0-2.0.1	3.1-1.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
73.	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
74.	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
75.		Removed BUG_ON assert when checking if the ring is full.	3.0-1.0.1	3.1-1.0.0
76.	libvma	Added libvma support for Debian 8.0 x86_64 and Ubuntu 15.04	3.0-2.0.1	3.1-1.0.0
77.	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
78.	Configura- tion	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
79.	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
80.		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
81.		Fixed double-free memory corruption in case where SR-IOV enabling failed (error flow).	3.0-1.0.1	3.0-2.0.0
82.	Start-up sequence	Fixed a crash in EQ's initialization error flow.	3.0-1.0.1	3.0-2.0.0
83.	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

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
84.	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
85.		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
86.		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
87.	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
88.	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
89.	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
90.	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
91.		Resolved soft lock on massive amount of user memory registrations	2.3-2.0.1	2.4-1.0.0
92.	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

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
93.	mlx4_en	LRO fixes and improvements for jumbo MTU.	2.3-2.0.1	2.4-1.0.0
94.		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
95.		Fixed ping issues with IP fragmented data-grams in MTUs 1600-1700.	2.2-1.0.1	2.4-1.0.0
96.		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
97.	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
98.		Fixed an issue in verbs API: fallback to glibc on contiguous memory allocation failure	2.3-2.0.1	2.3-2.0.5
99.	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
100.	IB MAD	Fixed an issue to prevent process starvation due to MAD packet storm.	2.3-2.0.1	2.3-2.0.5
101.	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
102.		Returned the CQ to its original state (armed) to prevent traffic from stopping	2.3-1.0.1	2.3-2.0.0
103.		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
104.	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

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
105.	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
106.		Fixed an issue that caused an application to fail when attaching Shared Memory.	2.3-1.0.1	2.3-2.0.0
107.	mlx4_en	Fixed dmesg warnings: "NOHZ: local_soft-irq_pending 08".	2.3-1.0.1	2.3-2.0.0
108.		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
109.	mlx5_core	Fixed race when async events arrived during driver load.	2.3-1.0.1	2.3-2.0.0
110.		Fixed race in <code>mlx5_eq_int</code> when events arrived before <code>eq->dev</code> was set.	2.3-1.0.1	2.3-2.0.0
111.		Enabled all pending interrupt handlers completion before freeing EQ memory.	2.3-1.0.1	2.3-2.0.0
112.	mlnx.conf	Defined <code>mlnx.conf</code> as a configuration file in <code>mlnx-ofa_kernel RPM</code>	2.1-1.0.0	2.3-2.0.0
113.	SR-IOV	Fixed counter index allocation for VFs which enables Ethernet port statistics.	2.3-1.0.1	2.3-2.0.0
114.	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
115.	RoCE v2	RoCE v2 was non-functional on big Endian machines.	2.3-1.0.1	2.3-2.0.0
116.	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
117.	Installation	Using both ' <code>-c --config</code> ' and ' <code>--add-kernel-support</code> ' flags simultaneously when running the <code>mlnxofedinstall.sh</code> script caused installation failure with the following on screen message " <code>--config does not exist</code> ".	2.2-1.0.1	2.3-2.0.0
118.	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

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
119.	XRC	XRC over ROCE in SR-IOV mode is not functional	2.0-3.1.0	2.2-1.0.1
120.	mlx4_en	Fixed wrong calculation of packet true-size reporting in LRO flow.	2.1-1.0.0	2.2-1.0.1
121.		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
122.		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
123.	IB Core	Fixed the QP attribute mask upon smac resolving	2.1-1.0.0	2.1-1.0.6
124.	mlx5_ib	Fixed a send WQE overhead issue	2.1-1.0.0	2.1-1.0.6
125.		Fixed a NULL pointer de-reference on the debug print	2.1-1.0.0	2.1-1.0.6
126.		Fixed arguments to kzalloc	2.1-1.0.0	2.1-1.0.6
127.	mlx4_core	Fixed the locks around completion handler	2.1-1.0.0	2.1-1.0.6
128.	mlx4_core	Restored port types as they were when recovering from an internal error.	2.0-2.0.5	2.1-1.0.0
129.		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
130.	SR-IOV	Fixed memory leak in SR-IOV flow.	2.0-2.0.5	2.0-3.0.0
131.		Fixed communication channel being stuck	2.0-2.0.5	2.0-3.0.0
132.	mlx4_en	Fixed ALB bonding mode failure when enslaving Mellanox interfaces	2.0-3.0.0	2.1-1.0.0
133.		Fixed leak of mapped memory	2.0-3.0.0	2.1-1.0.0
134.		Fixed TX timeout in Ethernet driver.	2.0-2.0.5	2.0-3.0.0
135.		Fixed ethtool stats report for Virtual Functions.	2.0-2.0.5	2.0-3.0.0
136.		Fixed an issue of VLAN traffic over Virtual Machine in paravirtualized mode.	2.0-2.0.5	2.0-3.0.0
137.		Fixed ethtool operation crash while interface down.	2.0-2.0.5	2.0-3.0.0

Table 47 - Bug Fixes History

#	Issue	Internal Reference Number: Description	Discovered in Release	Fixed in Release
138.	IPoIB	Fixed memory leak in Connected mode.	2.0-2.0.5	2.0-3.0.0
139.		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 48 - Change Log History

Release	Category	Description
3.4-2.1.4.1	On Demand Paging (ODP) Implicit Memory Region (MR)	Added support to provide an implicit lkey that represents the complete address space.
3.4-2.0.0.0	NVMeoF	Added support for NVMeoF in host/target systems over RDMA.
3.4-1.0.0.0	VST Q-in-Q	[ConnectX@-3/ConnectX@-3 Pro] Added support for Q-in-Q encapsulation per VF in Linux (VST) for ConnectX-3 Pro adapter cards.
	Package Content	[ConnectX@-3/ConnectX@-3 Pro] SR-IOV enabled firmware binaries for ConnectX-3 has been removed from MLNX_OFED package (the installation flag “--enable-sriov” has been deprecated). To configure SR-IOV, please use the “mlxconfig” or “mst-config” utilities.
		[ConnectX@-3/ConnectX@-3 Pro] MLNX_OFED repository metadata files has been moved to the folder holding the binary packages (named “RPMS” in rpm based OS, and “DEBS” in Debian based OS). Please update your repository configuration file accordingly (refer to the MLNX_OFED User Manual for more details about setting up MLNX_OFED as a repository).

Table 48 - Change Log History

Release	Category	Description
	Raw Ethernet Programming	<p>[ConnectX@-4/ConnectX@-4 Lx] Added new APIs for enhanced raw Ethernet programming:</p> <ul style="list-style-type: none"> • Packet Pacing • TCP Segmentation Offload (TSO) • ToS based steering • Flow ID based steering (beta) • VxLAN based steering (beta) <p>For further information, refer to the “Programming” section in OFED User Manual.</p>
	Enhanced PCIe Error Recovery	<p>[ConnectX@-4/ConnectX@-4 Lx] Enhanced PCIe error recovery by adding the following behaviors to the flow:</p> <ul style="list-style-type: none"> • In case SR-IOV is enabled during the recovery process, it will not get automatically disabled and will require the administrator that enabled it to disable it. • When the driver goes down, VF PCI function will not be removed. • Ethernet interface attributes (MTU, state, ring size, etc...) will be recovered after the error recovery stage is completed. • The net device kernel layer will not be aware of any ongoing PCI error recovery process.
	SR-IOV Max Rate Limit Ethernet/RoCE (beta level)	<p>[ConnectX@-4/ConnectX@-4 Lx] Added the ability to rate-limit traffic per Virtual Function in SR-IOV mode.</p>
	Dynamically tuned Interrupt Moderation (DIM)	<p>[ConnectX@-4/ConnectX@-4 Lx] Added support for dynamically controlling the interrupts per channel to ensure maximum packet rate with minimum interrupt rate. This feature is enabled by default.</p>
	Dump Configuration	<p>[ConnectX@-4/ConnectX@-4 Lx] Added support for dump configuration which helps dumping driver and firmware configuration using ethtool. It creates a backup of the configuration files into a specified dump file.</p>
	Ethernet Counters	<p>[ConnectX@-4/ConnectX@-4 Lx] Updated the list of counters the can be retrieved via ethtool for mlx5 driver, changed counters names and added new counters.</p>

Table 48 - Change Log History

Release	Category	Description
	Mellanox PeerDirect Async (beta level)	<p>[ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4/ConnectX®-4 Lx]</p> <p>The experimental PeerDirect Async APIs have been changed to make the implementation of peer clients simpler. Note the following:</p> <ul style="list-style-type: none"> • These changes are not backward compatible. • The code adds CQ polling support for peer devices that do not support the NOR operation, replacing it with a GEQ operation. <p>To see the API changes, refer to the man page.</p>
	ABI Incompatibility	<p>[ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4/ConnectX®-4 Lx]</p> <p>Added the ability to fix the issue of preventing the load of MLNX_EN modules when a new kernel is not compatible with these modules.</p>
	Mellanox Scalable Hierarchical Aggregation Protocol (SHARP™)	<p>[Connect-IB/ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4]</p> <p>IB only: This technology improves the performance of MPI operation by offloading collective operations from the CPU and dispatching to the switch network, and eliminating the need to send data multiple times between endpoints. This approach decreases the amount of data traversing the network as aggregation nodes are reached, and dramatically reduces the MPI operation time.</p> <p>For further information on SHARP and its configuration, see SHARP Deployment Guide.</p>

Table 48 - Change Log History

Release	Category	Description
3.3-1.0.0.0	VF MAC Address Anti-Spoofing	[ConnectX-4/ConnectX-4 Lx] Also known as MAC spoof-check, the VF MAC Address Anti-Spoofing prevents malicious VFs from faking their MAC addresses.
	VF All-multi Mode	[ConnectX-4/ConnectX-4 Lx] Added support for the VF to enter all-multi RX mode, meaning that in addition to the traffic originally targeted to the VF, it will receive all the multicast traffic sent from/to the other functions on the same physical port. Note: Only privileged/trusted VFs can enter the all-multi RX mode.
	VF Promiscuous Mode	[ConnectX-4/ConnectX-4 Lx] Added support for the VF to enter promiscuous RX mode, meaning that in addition to the traffic originally targeted to the VF, it will receive the unmatched traffic and all the multicast traffic that reaches the physical port. The unmatched traffic is any traffic's DMAC that does not match any of the VFs' or PFs' MAC addresses. Note: Only privileged/trusted VFs can enter the promiscuous RX mode.
	Privileged VF	[ConnectX-4/ConnectX-4 Lx] Added support for determining privileged/trusted VFs so security sensitive features can be enabled for these VFs, such as entering promiscuous and all-multi RX modes.
	DCBX	[ConnectX-4/ConnectX-4 Lx] Added support for standard DCBX CEE API.
	Per Priority Counters	[ConnectX-4/ConnectX-4 Lx] Exposed performance counters per priority.
	IB Error Counters	[ConnectX-4/ConnectX-4 Lx] Exposed IB sysfs error counters for mlx5 driver.
	Accelerated Receive Flow Steering (aRFS)	[ConnectX-4/ConnectX-4 Lx] Boosts the speed of RFS by adding hardware assistance. RFS is an in-kernel-logic responsible for load balancing between CPUs by attaching flows to CPUs that are used by flow's owner applications.
	Packet Pacing for UDP/TCP	[ConnectX-4/ConnectX-4 Lx] Performs rate limit per UDP/TCP connection.
	OFED Scripts	Renamed the UP name that appears in mlnx_perf report to "TC", as the mlnx_perf script counts the packets and calculates the bandwidth on rings that belong to the same Traffic Class (TC).

Table 48 - Change Log History

Release	Category	Description
	Physical Memory Allocation	Added support for Physical Address Memory Region (PA-MR) which allows managing physical memory used for posting send and receive requests.
	MAD Congestion Control	Added an SA MAD congestion control mechanism that is configurable using sysfs entries.
	IB Router	Added the ability to send traffic between two or more sub-nets.
	PeerDirect Async	Mellanox PeerDirect Async™ sub-system gives peer hardware devices, such as GPU cards, and dedicated AS accelerators the ability to take control over HCA in critical path offloading CPU.
	Physical MR	Allows the user to use physical addresses instead of virtual addresses in critical path. Thus enhances performance since there is no need in addresses translation.
	RoCE v1 (Layer 2) Compatibility	Added the option to connect between nodes running MLNX_OFED and nodes running RoCE with Layer 2 GID format.
3.2-2.0.0.0	API Changes	<ul style="list-style-type: none"> • Support FCS scattering for Raw Packet QPs and WQs. • Indication of L4 packet type on the receive side completions • Support CVLAN insertion for WQs
	IPoIB	<ul style="list-style-type: none"> • Added support for the following IPoIB UD QP offloads: <ul style="list-style-type: none"> • RX check summing (AKA RX csu) • Large Send Offloads (AKA LSO) <p>To see the new IPoIB UD mode, run: "ethtool -k <interface>"</p>

Table 48 - Change Log History

Release	Category	Description
3.2-1.0.1.1	VXLAN Hardware Stateless Offloads	[ConnectX-4 / ConnectX-4 Lx] Provides scalability and security challenges solutions.
	Priority Flow Control (PFC)	[ConnectX-4 / ConnectX-4 Lx] Applies pause functionality to specific classes of traffic on the Ethernet link.
	Offloaded Traffic Sniffer/TCP Dump	[ConnectX-4 / ConnectX-4 Lx] Allows bypass kernel traffic (such as, RoCE, VMA, DPDK) to be captured by existing packet analyzer such as tcpdump.
	Ethernet Time Stamping	[ConnectX-4 / ConnectX-4 Lx] Keeps track of the creation of a packet. A time-stamping service supports assertions of proof that a datum existed before a particular time.
	Custom RoCE Counters	[ConnectX-4 / ConnectX-4 Lx] Provide a clear indication on RDMA send/receive statistics and errors.
	LED Beaconing	[ConnectX-4 / ConnectX-4 Lx] Enables visual identification of the port by LED blinking.
	Enhanced Transmission Selection standard (ETS)	[ConnectX-4 / ConnectX-4 Lx] Exploits the time periods in which the offered load of a particular Traffic Class (TC) is less than its minimum allocated bandwidth.
	Strided WQE User Space	[ConnectX-4 / ConnectX-4 Lx] Striding RQ is a receive queue comprised by work queue elements (i.e. WQEs), where multiple packets of LRO segments (i.e. message) are written to the same WQE.
	VLAN Stripping in Linux Verbs	[ConnectX-4 / ConnectX-4 Lx] Adds access to the device's ability to offload the Customer VLAN (cVLAN) header stripping from an incoming packet.
	iSER: Remote invalidation support (target and initiator)	[ConnectX-4 / ConnectX-4 Lx] Improves performance by enabling the hardware to perform implicit memory region invalidation.
	iSER: Zero-Copy ImmediateData	[ConnectX-4 / ConnectX-4 Lx] Reduces the latency of small writes by avoiding an extra memory copy in the iSER target stack.
iSER: Indirect Memory Registration	[ConnectX-4 / ConnectX-4 Lx] Uses ConnectX®-4 adapter card's Indirect Memory Registration capabilities to avoid bounce buffer strategy implementation and to reduce the latency of highly unaligned vectored IO operations, and also in cases of BIO merging.	

Table 48 - Change Log History

Release	Category	Description
	Vector Calculation/Erasure coding offload	[ConnectX-4 / ConnectX-4 Lx] Uses the HCA for offloading erasure coding calculations.
	Virtual Guest Tagging (VGT+)	[ConnectX-3 / ConnectX-3 Pro] VGT+ is 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.
	Link Aggregation for Virtual Functions	[ConnectX-3 / ConnectX-3 Pro] Protects a VM with an attached ConnectX-3 VF from VF port failure, when VFs are present and RoCE Link Aggregation is configured in the Hypervisor.
3.1-1.0.3	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 48
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.

Table 48 - Change Log History

Release	Category	Description
3.0-1.0.1	HCA's	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.

Table 48 - Change Log History

Release	Category	Description
	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 <code>query_gid</code> 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.
	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: <ul style="list-style-type: none"> • <code>ibv_exp_create_res_domain</code> • <code>ibv_exp_destroy_res_domain</code> • <code>ibv_exp_query_intf</code> • <code>ibv_exp_release_intf</code> Added the following interface families: <ul style="list-style-type: none"> • <code>ibv_exp_qp_burst_family</code> • <code>ibv_exp_cq_family</code>

Table 48 - Change Log History

Release	Category	Description
2.4-1.0.4	Bug Fixes	See “Bug Fixes History” on page 48.
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: <code>ethtool -k txvlan [on/off]</code>
		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 messages.
	libmlx5	Added ConnectX®-4 device (4114) to the list of supported devices (<code>hca_table</code>),
	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	<p>Added the following files under <code>/sys/class/infiniband/mlx5_0/mr_cache/</code>:</p> <ul style="list-style-type: none"> <code>rel_timeout</code>: Defines the minimum allowed time between the last MR creation to the first MR released from the cache. When <code>rel_timeout = -1</code>, MRs are not released from the cache <code>rel_imm</code>: Triggers the immediate release of excess MRs from the cache when set to 1. When all excess MRs are released from the cache, <code>rel_imm</code> is reset back to 0.
	Bug Fixes	See “Bug Fixes History” on page 48.
2.3-2.0.1	Bug Fixes	See “Bug Fixes History” on page 48.

Table 48 - 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 <code>mr_cache</code> subtree from <code>debugfs</code> to <code>mlx5_ib</code> while preserving all its semantics.
	InfiniBand Utilities	Updated the <code>ibutils</code> package. Added to the <code>ibdiagnet</code> tool the " <code>ibdiagnet2.mlx_cntrs</code> " option to enable reading of Mellanox diagnostic counters.
	Bug Fixes	See "Bug Fixes History" on page 48.

Table 48 - Change Log History

Release	Category	Description
2.3-1.0.1	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 <code>CONFIG_MLX4_EN_VXLAN=y</code> 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
	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.	

Table 48 - Change Log History

Release	Category	Description
	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

Table 48 - Change Log History

Release	Category	Description
Rev 2.2-1.0.1	Reset Flow	Reset Flow is not activated by default. It is controlled by the <code>mlx4_core</code> ' <code>internal_err_reset</code> ' module parameter.
	<code>mlnxofedinstall</code>	32-bit libraries are no longer installed by default on 64-bit OS. To install 32-bit libraries use the ' <code>--with-32bit</code> ' installation parameter.
	<code>openibd</code>	Added pre/post start/stop scripts support. For further information, please refer to section “ <i>openibd Script</i> ” in the MLNX_OFED User Manual.
	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 <code>num_cpu</code> for port ConnectX®-3 has 1024 MSI-X vectors, 28 MSI-X vectors are reserved. <ul style="list-style-type: none"> Physical Function - gets the number of MSI-X vectors according to the <code>pf_msix_table_size</code> (multiple of 4 - 1) INI parameter Virtual Functions – the remaining MSI-X vectors are spread equally between all VFs, according to the <code>num_vfs</code> <code>mlx4_core</code> 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 <code>libibverbs</code> . The Experimental API is an extended API therefore, 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: <ul style="list-style-type: none"> 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

Table 48 - Change Log History

Release	Category	Description
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.
	mlnxfedinstall	Added support for the following parameters: <ul style="list-style-type: none"> '--umad-dev-na' '--without-<package>'
	Content Packages Updates	The following packages were updated: <ul style="list-style-type: none"> 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.	

Table 48 - Change Log History

Release	Category	Description
Rev 2.0-3.0.0	Operating Systems	Additional OS support: <ul style="list-style-type: none"> • SLES11SP3 • Fedora16, Fedora17
	Hardware	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.		
Storage	iSER (at GA level)	

Table 48 - Change Log History

Release	Category	Description
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: <ul style="list-style-type: none"> • Checksum TX/RX • Flow Steering
	InfiniBand Network	Contiguous pages: <ul style="list-style-type: none"> • 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.

API Change Log History

Table 49 - API Change Log History

Release	Name	Description
Rev 3.2-2.0.0.0	libibverbs	<ul style="list-style-type: none"> • Support FCS scattering for Raw Packet QPs and WQs. <ul style="list-style-type: none"> • Query: <code>ibv_exp_query_device</code> reports <code>IBV_EXP_DEVICE_SCATTER_FCS</code> when it is supported. • Enablement of this feature is done in the creation: <ol style="list-style-type: none"> 1. For Raw Packet QPs: Set <code>IBV_EXP_QP_CREATE_SCATTER_FCS</code> in <code>exp_create_flags</code>. 2. For WQs: Set <code>IBV_EXP_CREATE_WQ_FLAG_SCATTER_FCS</code> in flags of <code>ibv_exp_wq_init_attr</code>. • Indication of L4 packet type on the receive side completions: <ul style="list-style-type: none"> • Query: <code>ibv_exp_query_device</code> reports <code>IBV_EXP_DEVICE_RX_TCP_UDP_PKT_TYPE</code> when it is supported. <code>bv_exp_cq_family_flags</code> was extended with two flags <code>IBV_EXP_CQ_RX_TCP_PACKET</code> and <code>IBV_EXP_CQ_RX_UDP_PACKET</code> to support L4 packet type when using <code>poll_length_flags()</code>. • Support CVLAN insertion for WQs: <ul style="list-style-type: none"> • Query: <code>IBV_EXP_RECEIVE_WQ_CVLAN_INSERTION</code> is set in <code>ibv_exp_vlan_offloads</code> when CVLAN insertion is supported. • Enablement: The <code>ibv_exp_qp_burst_family</code> was extended to support CVLAN insetion: <ol style="list-style-type: none"> 1. <code>send_pending_vlan</code>: Put one message in the provider send queue and insert <code>vlan_tci</code> to header. 2. <code>send_pending_inline_vlan</code>: Put one inline message in the provider send queue and insert <code>vlan_tci</code> to header. 3. <code>send_pending_sg_list_vlan</code>: Put one scatter-gather(sg) message in the provider send queue and insert <code>vlan_tci</code> to header.

Table 49 - API Change Log History

Release	Name	Description
3.2-1.0.1.1	libibverbs	<ul style="list-style-type: none"> • [Beta] Added API and primitives for Erasure Coding calculations. <ul style="list-style-type: none"> • Verbs: <ul style="list-style-type: none"> • Ibv_exp_alloc_ec_calc • Ibv_exp_dealloc_ec_calc • Ibv_exp_ec_encode_sync • Ibv_exp_ec_decode_sync • Ibv_exp_ec_encode_async • Ibv_exp_ec_decode_async • Ibv_exp_ec_encode_send <p>For further information, please refer to the manual page of the verbs.</p> • Structs: <ul style="list-style-type: none"> • Ibv_exp_ec_calc • Ibv_exp_ec_mem • Ibv_exp_ec_stripe • Ibv_exp_ec_comp • Added version 1 of the CQ family with support for: <ul style="list-style-type: none"> • Multi-Packet RQ (also called striding RQ) • Cvlan stripping offload • Added enhanced masked-atomic device capability • Added a flag to the create QP/WQ option to enable end of RX message padding
Rev 3.1-1.0.3	libibverbs	<ul style="list-style-type: none"> • Added <code>ibv_exp_wq_family</code> 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 <code>ibv_exp_query_intf</code> to notify that common-flags/family-flags are not supported. • Added <code>ibv_exp_query_gid_attr</code> verb. For further information, please refer to the manual page of the verb.

Table 49 - API Change Log History

Release	Name	Description
Rev 3.0-1.0.0	libibverbs	<ul style="list-style-type: none"> Added the following new APIs: <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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	<p>Added the following verbs interfaces:</p> <ul style="list-style-type: none"> ibv_create_flow ibv_destroy_flow ibv_exp_use_priv_env ibv_exp_setenv
Rev 2.3-1.0.1	libibverbs	<ul style="list-style-type: none"> ibv_exp_rereg_mr - Added new API for memory region re-integration (For further information, please refer to MLNX_OFED User Manual) Added to the experimental API <code>ibv_exp_post_send</code> the following opcodes: <ul style="list-style-type: none"> IBV_EXP_WR_EXT_MASKED_ATOMIC_CMP_AND_SWP IBV_EXP_WR_EXT_MASKED_ATOMIC_FETCH_AND_ADD IBV_EXP_WR_NOP <p>and these completion opcodes:</p> <ul style="list-style-type: none"> IBV_EXP_WC_MASKED_COMP_SWAP IBV_EXP_WC_MASKED_FETCH_ADD

Table 49 - API Change Log History

Release	Name	Description
Rev 2.2-1.0.1	libibverbs	<p>The following verbs changed to align with upstream libibverbs:</p> <ul style="list-style-type: none"> • <code>ibv_reg_mr</code> - <code>ibv_access_flags</code> changed. • <code>ibv_post_send</code> - opcodes and send flags changed and <code>wr</code> fields removed (<code>task</code>, <code>op</code>, <code>dc</code> and <code>bind_mw</code>) • <code>ibv_query_device</code> - capability flags changed. • <code>ibv_poll_cq</code> - opcodes and <code>wc</code> flags changed. • <code>ibv_modify_qp</code> - mask bits changed • <code>ibv_create_qp_ex</code> - <code>create_flags</code> field removed. <p>The following verbs removed to align with upstream libibverbs:</p> <ul style="list-style-type: none"> • <code>ibv_bind_mw</code> • <code>ibv_post_task</code> • <code>ibv_query_values_ex</code> • <code>ibv_query_device_ex</code> • <code>ibv_poll_cq_ex</code> • <code>ibv_reg_shared_mr_ex</code> • <code>ibv_reg_shared_mr</code> • <code>ibv_modify_cq</code> • <code>ibv_create_cq_ex</code> • <code>ibv_modify_qp_ex</code>
Rev 2.2-1.0.1	Verbs Experimental API	<p>The following experimental verbs added (replacing the removed extended verbs):</p> <ul style="list-style-type: none"> • <code>ibv_exp_bind_mw</code> • <code>ibv_exp_post_task</code> • <code>ibv_exp_query_values</code> • <code>ibv_exp_query_device</code> • <code>ibv_exp_poll_cq</code> • <code>ibv_exp_reg_shared_mr</code> • <code>ibv_exp_modify_cq</code> • <code>ibv_exp_create_cq</code> • <code>ibv_exp_modify_qp</code> <p>New experimental verbs:</p> <ul style="list-style-type: none"> • <code>ibv_exp_arm_dct</code> • <code>ibv_exp_query_port</code> • <code>ibv_exp_create_flow</code> • <code>ibv_exp_destroy_flow</code> • <code>ibv_exp_post_send</code> • <code>ibv_exp_reg_mr</code> • <code>ibv_exp_get_provider_func</code>

Table 49 - API Change Log History

Release	Name	Description
Rev 2.1-1.0.0	Dynamically Connected (DC)	<p>The following verbs were added:</p> <ul style="list-style-type: none"> • <code>struct ibv_dct *ibv_exp_create_dct(struct ibv_context *context, struct ibv_exp_dct_init_attr *attr)</code> • <code>int ibv_exp_destroy_dct(struct ibv_dct *dct)</code> • <code>int ibv_exp_query_dct(struct ibv_dct *dct, struct ibv_exp_dct_attr *attr)</code>
	Verbs Extension API: Verbs extension API defines OFA APIs extension scheme to detect ABI compatibility and enable backward and forward compatibility support.	<ul style="list-style-type: none"> • <code>ibv_post_task</code> • <code>ibv_query_values_ex</code> • <code>ibv_query_device_ex</code> • <code>ibv_create_flow</code> • <code>ibv_destroy_flow</code> • <code>ibv_poll_cq_ex</code> • <code>ibv_reg_shared_mr_ex</code> • <code>ibv_open_xrzd</code> • <code>ibv_close_xrzd</code> • <code>ibv_modify_cq</code> • <code>ibv_create_srq_ex</code> • <code>ibv_get_srq_num</code> • <code>ibv_create_qp_ex</code> • <code>ibv_create_cq_ex</code> • <code>ibv_open_qp</code> • <code>ibv_modify_qp_ex</code>
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.	<ul style="list-style-type: none"> • <code>ibv_exp_create_qp</code> • <code>ibv_exp_query_device</code> • <code>ibv_exp_create_dct</code> • <code>ibv_exp_destroy_dct</code> • <code>ibv_exp_query_dct</code>
Rev 2.0-3.0.0	XRC	<p>The following verbs have become deprecated:</p> <ul style="list-style-type: none"> • <code>struct ibv_xrc_domain *ibv_open_xrc_domain</code> • <code>struct ibv_srq *ibv_create_xrc_srq</code> • <code>int ibv_close_xrc_domain</code> • <code>int ibv_create_xrc_rcv_qp</code> • <code>int ibv_modify_xrc_rcv_qp</code> • <code>int ibv_query_xrc_rcv_qp</code> • <code>int ibv_reg_xrc_rcv_qp</code> • <code>int ibv_unreg_xrc_rcv_qp</code>

Table 49 - API Change Log History

Release	Name	Description
Rev 2.0-2.0.5	Libibverbs - Extended speeds	<ul style="list-style-type: none"> Missing the <code>ext_active_speed</code> attribute from the struct <code>ibv_port_attr</code> Removed function <code>ibv_ext_rate_to_int</code> Added functions <code>ibv_rate_to_mbps</code> and <code>mbps_to_ibv_rate</code>
	Libibverbs - Raw QPs	QP types <code>IBV_QPT_RAW_PACKET</code> and <code>IBV_QPT_RAW_ETH</code> are not supported
	Libibverbs - Contiguous pages	<ul style="list-style-type: none"> Added Contiguous pages support Added function <code>ibv_reg_shared_mr</code>
	Libmverbs	<ul style="list-style-type: none"> The enumeration <code>IBV_M_WR_CALC</code> was renamed to <code>IBV_M_WR_CALC_SEND</code> The enumeration <code>IBV_M_WR_WRITE_WITH_IMM</code> was added In the structure <code>ibv_m_send_wr</code>, the union <code>wr.send</code> was renamed to <code>wr.calc_send</code> and <code>wr.rdma</code> was added The enumerations <code>IBV_M_WQE_CAP_CALC_RDMA_WRITE_WITH_IMM</code> was added The following enumerations were renamed: <ul style="list-style-type: none"> From <code>IBV_M_WQE_SQ_ENABLE_CAP</code> to <code>IBV_M_WQE_CAP_SQ_ENABLE</code> From <code>IBV_M_WQE_RQ_ENABLE_CAP</code> to <code>IBV_M_WQE_CAP_RQ_ENABLE</code> From <code>IBV_M_WQE_CQE_WAIT_CAP</code> to <code>IBV_M_WQE_CAP_CQE_WAIT</code> From <code>IBV_M_WQE_CALC_CAP</code> to <code>IBV_M_WQE_CAP_CALC_SEND</code>