Mellanox OFED for Linux Release Notes

Rev 4.2-1.0.0.0

NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT (“PRODUCT(S)”) AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES “AS-IS” 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

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®, NPS®, Open Ethernet logo, PhyX®, PlatformX®, PSIPHY®, SiPhy®, StoreX®, SwitchX®, Tilera®, Tilera 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

Release Update History

Release	Date	Description
4.2-1.2.0.0	November 1, 2017	Initial release of this version.

1 Overview

These are the release notes of MLNX_OFED for Linux Driver, Rev 4.2-1.0.0.0 which operates across all Mellanox network adapter solutions supporting the following uplinks to servers:

Uplink/HCAs	Driver Name	Uplink Speed
ConnectX®-3/ ConnectX®-3 Pro	mlx4	• InfiniBand: SDR, QDR, FDR10, FDR • Ethernet: 10GigE, 40GigE and 56GigE[a]
ConnectX®-4	mlx5	• InfiniBand: SDR, QDR, FDR, FDR10, EDR • Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, 56GigE[a], and 100GigE
ConnectX®-4 Lx		• Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, and 50GigE
ConnectX®-5		• InfiniBand: SDR, QDR, FDR, FDR10, EDR • Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, and 100GigE
ConnectX®-5 Ex		• InfiniBand: SDR, QDR, FDR, FDR10, EDR • Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, and 100GigE
Innova™ IPsec EN		Ethernet: 10GigE, 40GigE
Connect-IB®		• InfiniBand: SDR, QDR, FDR10, FDR

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

1.1 Content of Mellanox OFED for Linux

Mellanox OFED for Linux software contains the following components:

Components	Description
OpenFabrics core and ULPs	• InfiniBand and Ethernet HCA drivers (mlx4, mlx5) • core • Upper Layer Protocols: IPoIB, SRP Initiator, iSER Initiator and Target, NVMEoF Host and Target **Note**: iSER Target (iSERT) supports the following distribution kernels only: RHEL 7.2/7.3, SLES 12.1/12.2, Ubuntu14.04/15.04/16.04/16.10
OpenFabrics utilities	• OpenSM: IB Subnet Manager with Mellanox proprietary Adaptive Routing • Diagnostic tools • Performance tests • SSA (SLES12): libopensmssa plugin for OpenSM, ibssa, ibacm

Components	Description
MPI	• Open MPI stack 1.6.5 and later supporting the InfiniBand interface • MPI benchmark tests (OSU benchmarks, Intel MPI benchmarks, Presta)
PGAS	• HPC-X OpenSHMEM v2.2 supporting InfiniBand, MXM and FCA • HPC-X UPC v2.2 supporting InfiniBand, MXM and FCA
HPC Acceleration packages	• Mellanox MXM v3.0 (p2p transport library acceleration over Infiniband) • Mellanox FCA v3.x (MPI/PGAS collective operations acceleration library over InfiniBand) • KNEM, Linux kernel module enabling high-performance intra-node MPI/PGAS communication for large messages
Extra packages	• ibutils2 • ibdump • MFT
Sources of all software modules (under conditions mentioned in the modules' LICENSE files) except for MFT, OpenSM plugins, ibutils2, and ibdump	
Documentation	

1.2 Supported Platforms and Operating Systems

The following are the supported OSs in MLNX_OFED Rev 4.2-1.0.0.0:

Table 1 - Supported Platforms and Operating Systems

Operating System	Platform
RHEL6.3/CentOS6.3	x86_64
RHEL6.6/CentOS6.6	x86_64
RHEL6.8/CentOS6.8	x86_64
RHEL6.9/CentOS6.9	x86_64/PPC64 (Power7)
RHEL7.2/CentOS7.2	x86_64/PPC64 (Power8)/PPC64LE (Power8)
CentOS7.3 with Kernel 4.9 for NVME-oF	Armv8(AMD) for Softiron
RHEL7.3/CentOS7.3	x86_64/PPC64 (Power8)/PPC64LE (Power8)
RHEL7.4/CentOS7.4	x86_64/PPC64LE (Power8)
Debian 8.7	x86_64
Debian 8.7 Kernel 4.1	x86_64
Debian 8.7 Kernel 4.4	x86_64

Table 1 - Supported Platforms and Operating Systems

Operating System	Platform
Debian 9.0	x86_64
Fedora 20	x86_64
Fedora 25	x86_64
Fedora 26	x86_64
OL 6.8	x86_64
OL 7.3	x86_64
SLES11 SP1	x86_64
SLES11 SP3	x86_64
SLES11 SP4	x86_64/PPC64 (Power 7)
SLES12 SP2	x86_64/PPC64 (Power 8)
SLES12 SP3	x86_64/PPC64 (Power8)
Ubuntu 14.04	x86_64
Ubuntu 16.04 with Kernel 4.9 - Bandera for Arm	Armv8 (Qualcomm) [**beta**]
Ubuntu 16.04.02	x86_64/PPC64LE (Power8)
Ubuntu 16.04.03	x86_64/PPC64LE (Power8)
Ubuntu 16.10	x86_64/PPC64LE (Power 8)
Ubuntu 17.04	x86_64/PPC64LE (Power 8)
Kernels 4.10-4.13	x86_64
CoreOS Kernel 4.11	x86_64
EulerOS 2.0 SP2	x86_64
XenServer 7.2	x86_64

32 bit platforms are no longer supported in MLNX_OFED.

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 4.2-1.0.0.0:

- Windows Server 2012 R2

1.2.2 Tested Hypervisors in Paravirtualized and SR-IOV Environments

Table 2 - Tested Hypervisors in Paravirtualized and SR-IOV Environments

Tested Hypervisors	HCAs	Operating System
SR-IOV	ConnectX-3/ ConnectX-3 Pro	RHEL6.9/CentOS6.9 KVM
		RHEL7.2/CentOS7.2 KVM
		RHEL7.3/CentOS7.3 KVM
		RHEL7.4/CentOS7.4 KVM
		Ubuntu 16.04.03 KVM
	ConnectX-4	RHEL6.9/CentOS6.9 KVM
		RHEL7.2/CentOS7.2 KVM
		RHEL7.3/CentOS7.3 KVM
		RHEL7.4/CentOS7.4 KVM
		Ubuntu 16.04.03 KVM
		Ubuntu 17.04 KVM
		XenServer 7.2
	ConnectX-4 Lx	RHEL6.9/CentOS6.9 KVM
		RHEL7.2/CentOS7.2 KVM
		RHEL7.3/CentOS7.3 KVM
		Ubuntu 14.04 KVM
		Ubuntu 16.04.03 KVM
		Ubuntu 17.04 PPC KVM
		XenServer 7.2
	ConnectX-5	RHEL6.9/CentOS6.9 KVM
		RHEL7.2/CentOS7.2 KVM
		RHEL7.3/CentOS7.3 KVM
		RHEL7.4/CentOS7.4 KVM
		Ubuntu 16.04.03 KVM

Table 2 - Tested Hypervisors in Paravirtualized and SR-IOV Environments

Tested Hypervisors	HCAs	Operating System
Paravirtualized	ConnectX-3	RHEL7.3/CentOS7.3 KVM
	ConnectX-3 Pro	Ubuntu 16.04.03 KVM
	ConnectX-4	RHEL7.3/CentOS7.3 PPC KVM
		RHEL7.4/CentOS7.4 PPC KVM
		Ubuntu 16.04.03 KVM
		Ubuntu 16.04.03 PPC KVM
	ConnectX-4 Lx	Ubuntu 16.04.03 KVM
		Ubuntu 16.04.03 PPC KVM
	ConnectX-5	RHEL7.3/CentOS7.3 PPC KVM
		RHEL7.4/CentOS7.4 PPC KVM

1.3 Hardware and Software Requirements

The following are the hardware and software requirements of MLNX_OFED Rev 4.2-1.0.0.0.

- 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	`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 lsof`
SLES 11 SP3	`zypper install perl pciutils python libnl-32bit libxml2-python tcsh libstdc++43 libnl expat glib2 tcl bc tk libcurl4 gtk2 atk cairo pkg-config ethtool lsof`
SLES 12	`zypper install pkg-config expat libstdc++6 libglib-2_0-0 libgtk-2_0-0 tcl libcairo2 tcsh python bc pciutils libatk-1_0-0 tk python-libxml2 lsof libnl3-200 ethtool lsof`
Ubuntu/Debian	`apt-get install perl dpkg autotools-dev autoconf libtool automake1.10 automake m4 dkms debhelper tcl tcl8.4 chrpath swig graphviz tcl-dev tcl8.4-dev tk-dev tk8.4-dev bison flex dpatch zlib1g-dev curl libcurl4-gnutls-dev python-libxml2 libvirt-bin libvirt0 libnl-dev libglib2.0-dev libgfortran3 automake m4 pkg-config libnuma logrotate ethtool lsof`

Operating System	Required Packages Installation Command
Debian 8	`apt-get install libnl-3-200 automake debhelper curl dkms logrotate libglib2.0-0 python-libxml2 graphviz tk tcl libvirt-bin coreutils pkg-config autotools-dev flex autoconf pciutils quilt module-init-tools libvirt0 libstdc++6 dpkg libgfortran3 procps lsof libltdl-dev gcc dpatch chrpath grep m4 gfortran bison libnl-route-3-200 swig perl make ethtool lsof`

1.4 Supported HCAs Firmware Versions

MLNX_OFED Rev 4.2-1.0.0.0 supports the following Mellanox network adapter cards firmware versions:

Table 3 - Supported HCAs Firmware Versions

HCA	Recommended Firmware Rev.	Additional Firmware Rev. Supported
ConnectX®-3	2.42.5000	2.40.7000
ConnectX®-3 Pro	2.42.5000	2.40.7000
ConnectX®-4	12.21.1000	12.20.1010
ConnectX®-4 Lx	14.21.1000	14.20.1010
ConnectX®-5	16.21.1000	16.20.1010
ConnectX®-5 Ex	16.21.1000	16.20.1010
Connect-IB®	10.16.1020	N/A

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

1.5 Compatibility Matrix

MLNX_OFED Rev 4.2-1.0.0.0 is compatible with the following:

Table 4 - Compatibility Matrix

Mellanox Product	Description/Version
MLNX-OS®	MSX6036 w/w MLNX-OS® version 3.6.4006[a]
Grid Director™	4036 w/w Grid Director™ version 3.9.1-985
Unified Fabric Manager (UFM®)	v5.9.6
MXM	v3.6.3103
HCOLL[b]	v3.9
OpenMPI	v3.0.0

[a] MLNX_OFED Rev 4.2-1.0.0.0 was tested with this switch. However, additional switches may be supported as well.

[b] HCOLL is now the default FCA version used in HPC-X, starting from HPC-X v1.8. This version replaces FCA v2.x.

1.6 RDMA CM and RoCE Modes

1.6.1 RoCE Modes Matrix

The following is RoCE modes matrix:

Table 5 - 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-5/ ConnectX-5 Ex[a]	ConnectX-3 Pro	ConnectX-4/ ConnectX-4 Lx/ ConnectX-5/ ConnectX-5 Ex[a]	ConnectX-3 Pro/ConnectX-4/ ConnectX-4 Lx/ConnectX-5/ ConnectX-5 Ex[a]
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	-	-	-	-
SLES	12	-	-	-	-
Ubuntu	14.04.4, 16.04, 15.10	-	-	-	-

[a] Note that support for ConnectX-5 and ConnectX-5 Ex adapter cards in MLNX_OFED starts from v4.0.

1.6.2 RDMA CM Default RoCE Mode

The default RoCE mode on which RDMA CM runs is RoCEv2 instead of RoCEv1, starting from MLNX_OFED v4.1. RDMA_CM session requires both the client and server sides to support the same RoCE mode. Otherwise, the client will fail to connect to the server.

For further information, refer to RDMA CM and RoCE Version Defaults Community post.

2 Changes and New Features in Rev 4.2-1.0.0.0

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

Table 6 - Changes and New Features in Rev 4.2-1.0.0.0

HCAs	Feature/Change	Description
mlx5 Driver	Physical Address Memory Allocation	Added support to register a specific physical address range.
Innova IPsec EN	Innova IPsec Adapter Cards	Added support for Mellanox Innova IPsec EN adapter card, that provides security acceleration for IPsec-enabled networks.
ConnectX-4/ ConnectX-4 Lx/ ConnectX-5	Precision Time Protocol (PTP)	Added support for PTP feature over PKEY interfaces. This feature allows for accurate synchronization between the distributed entities over the network. The synchronization is based on symmetric Round Trip Time (RTT) between the master and slave devices, and is enabled by default.
	1PPS Time Synchronization	Added support for One Pulse Per Second (1PPS) over IPoIB interfaces.
	Virtual MAC	Added support for Virtual MAC feature, which allows users to add up to 4 virtual MACs (VMACs) per VF. All traffic that is destined to the VMAC will be forwarded to the relevant VF instead of PF. All traffic going out from the VF with source MAC equal to VMAC will go to the wire also when Spoof Check is enabled. For further information, please refer to “Virtual MAC” section in MLNX_OFED User Manual.
	Receive Buffer	Added the option to change receive buffer size and cable length. Changing cable length will adjust the receive buffer's xon and xoff thresholds. For further information, please refer to “Receive Buffer” section in MLNX_OFED User Manual.
	GRE Tunnel Offloads	Added support for the following GRE tunnel offloads: • TSO over GRE tunnels • Checksum offloads over GRE tunnels • RSS spread for GRE packets
	NVMEoF	Added support for the host side (RDMA initiator) in Red-Hat 7.2 and above.
	Dropless Receive Queue (RQ)	Added support for the driver to notify the FW when SW receive queues are overloaded.

Table 6 - Changes and New Features in Rev 4.2-1.0.0.0

HCAs	Feature/Change	Description
	PFC Storm Prevention	Added support for configuring PFC stall prevention in cases where the device unexpectedly becomes unresponsive for a long period of time. PFC stall prevention disables flow control mechanisms when the device is stalled for a period longer than the default pre-configured timeout. Users now have the ability to change the default timeout by moving to auto mode. For further information, please refer to "PFC Stall Prevention" section in MLNX_OFEDUser Manual.
	Force DSCP	Added support for this feature that enables setting a global traffic_class value for all RC QPs.
ConnectX-5	Q-in-Q	Added support for Q-in-Q VST feature in ConnectX-5 adapter cards family.
	Device Memory Programming [**beta**]	Added support for on-chip memory allocation and usage in send/receive and RDMA operations at beta level.
	Virtual Guest Tagging (VGT+)	Added support for VGT+ in ConnectX-4/ConnectX-5 HCAs. This feature is s 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. The policy determines which VLAN IDs are allowed to be transmitted or received. The policy does not determine the user priority, which is left unchanged. For further information, please refer to "Virtual Guest Tagging (VGT+)" section in MLNX_OFED User Manual.
	Tag Matching Offload	Added support for hardware Tag Matching offload with Dynamically Connected Transport (DCT).
ConnectX-3/ ConnectX-3 Pro	Shared Memory Region (MR)	Removed support for Shared MR feature on ConnectX-3/ ConnectX-3 Pro adapter cards. As a result of this change, the following API/flags should not be used: • ibv_exp_reg_shared_mr • access shared flags for ibv_exp_reg_mr (IBV_-EXP_ACCESS_SHARED_MR_XXX)

Table 6 - Changes and New Features in Rev 4.2-1.0.0.0

HCAs	Feature/Change	Description
All	CRDUMP	Added support for the driver to take an automatic snapshot of the device's CR-Space in cases of critical failures. For further information, please refer to "CRDUMP" section in MLNX_OFED User Manual.
	Upstream Libraries	Added the option to install upstream libraries (based on upstream rdma-core) for DPDK users only. For further information, please refer to "Installing Upstream rdma-core Libraries" section in MLNX_OFED User Manual.
	DiSNI	Added the option to install libdisni package as part of MLNX_OFED. For further information, please refer to section "Installing libdisni Package" in MLNX_OFED User Manual.
	Service Scripts	Added the ability to disable the '`stop`' option in the openibd service script, by setting ALLOW_STOP=no in /etc/infiniband/openib.conf. Starting from the next release, '`stop`' option will be disabled by default, and in order to enable it, ALLOW_STOP should be set to 'yes' in the conf file, or `force-stop` should be run.
	Bug Fixes	See Section 4, "Bug Fixes History", on page 46.

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

2.1 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.

Table 7 - API Changes

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

2.2 Unsupported Functionalities/Features/HCAs

The following are the unsupported functionalities/features/HCAs in MLNX_OFED:

- ConnectX®-2 Adapter Card
- Ethernet IP over InfiniBand (eIPoIB)
- Relational Database Service (RDS)
- Ethernet over InfiniBand (EoIB) - mlx4_vnic
- Ethernet IP over InfiniBand (EIPoIB)
- mthca InfiniBand driver

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.

For the list of old known issues, please refer to Mellanox OFED Archived Known Issues file at: http://www.mellanox.com/pdf/prod_software/MLNX_OFED_Archived_Known_Issues.pdf

Table 8 - Known Issues

Internal Reference Number	Issue
-	**Description**: SR-IOV mode over Windows VM is only supported over InfiniBand/ Ethernet ConnectX-4/ConnectX-5 adapter cards family.
	Workaround: N/A
	Keywords: SR-IOV, VM
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: Packet Size (Actual Packet MTU) limitation for IPsec offload on Innova IPsec adapter cards: The current offload implementation does not support IP fragmentation. The original packet size should be such that it does not exceed the interface's MTU size after the ESP transformation (encryption of the original IP packet which increases its length) and the headers (outer IP header) are added: • Inner IP packet size <= I/F MTU - ESP additions (20) - outer_IP (20) - fragmentation issue reserved length (56) • Inner IP packet size <= I/F MTU - 96 This mostly affects forwarded traffic into smaller MTU, as well as UDP traffic. TCP does PMTU discovery by default and clamps the MSS accordingly.
	Workaround: N/A
	Keywords: Innova IPsec, MTU
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: No LLC/SNAP support on Innova IPsec adapter cards.
	Workaround: N/A
	Keywords: Innova IPsec, LLC/SNAP
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: No support for FEC on Innova IPsec adapter cards. When using switches, there may be a need to change its configuration.
	Workaround: N/A
	Keywords: Innova IPsec, FEC
	Discovered in Release: 4.2-1.0.0.0

Table 8 - Known Issues

Internal Reference Number	Issue
955929	**Description**: Heavy traffic may cause SYN flooding when using Innova IPsec adapter cards.
	Workaround: N/A
	Keywords: Innova IPsec, SYN flooding
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: Priority Based Flow Control is not supported on Innova IPsec adapter cards.
	Workaround: N/A
	Keywords: Innova IPsec, Priority Based Flow Control
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: Pause configuration is not supported when using Innova IPsec adapter cards. Default pause is global pause (enabled).
	Workaround: N/A
	Keywords: Innova IPsec, Global pause
	Discovered in Release: 4.2-1.0.0.0
1045097	**Description**: Connecting and disconnecting a cable several times may cause a link up failure when using Innova IPsec adapter cards.
	Workaround: N/A
	Keywords: Innova IPsec, Cable, link up
	Discovered in Release: 4.2-1.0.0.0
-	**Description**: On Innova IPsec adapter cards, supported MTU is between 512 and 2012 bytes. Setting MTU values outside this range might fail or might cause traffic loss.
	Workaround: Set MTU between 512 and 2012 bytes
	Keywords: Innova IPsec, MTU
	Discovered in Release: 4.2-1.0.0.0
1177196	**Description**: If the OpenSM version is 4.8.1 and above, the IB interfaces link remains Down while the "SRIOV_IB_ROUTING_MODE_P1=1" and "SRIOV_IB_ROUTING_MODE_P2=1" flags are enabled in the HCA.
	Workaround: N/A
	Keywords: OpenSM, SR-IOV, IB link
	Discovered in Release: 4.2-1.0.0.0

Table 8 - Known Issues

Internal Reference Number	Issue
1118530	**Description**: On kernel versions 4.10-4.13, when resetting sriov_numvfs to 0 on PowerPC systems, the following dmesg warning will appear: mlx5_core <BDF>: can't update enabled VF BAR0
	Workaround: Reboot the system to reset sriov_numvfs value.
	Keywords: SR-IOV, numvfs
	Discovered in Release: 4.2-1.0.0.0
1125184	**Description**: In old kernel versions, such as Ubuntu 14.04 and RedHat 7.1, VXLAN interface does not reply to ARP requests for a MAC address that exists in its own ARP table. This issue was fixed in the following newer kernel versions: Ubuntu 16.04 and RedHat 7.3.
	Workaround: N/A
	Keywords: ARP, VXLAN
	Discovered in Release: 4.2-1.0.0.0
1171764	**Description**: Connecting multiple ports on the same server to the same subnet (IP/IB) will cause all interfaces connected to that subnet to respond to ARP requests. As a result, wrong ARP replies might be received when trying to resolve IP addresses.
	Workaround: Run the following to make sure only the interface with the requested IP address responds to the ARP request: `sysctl -w net.ipv4.conf.all.arp_ignore=1`
	Keywords: IPoIB, librdmacm, ARP
	Discovered in Release: 4.2-1.0.0.0
1134323	**Description**: When using kernel versions older than version 4.7 with IOMMU enabled, performance degradations and logical issues (such as soft lockup) might occur upon high load of traffic. This is caused due to the fact that IOMMU IOVA allocations are centralized, requiring many synchronization operations and high locking overhead amongst CPUs.
	Workaround: Use kernel v4.7 or above, or a backported kernel that includes the following patches: • 2aac630429d9 iommu/vt-d: change intel-iommu to use IOVA frame numbers • 9257b4a206fc iommu/iova: introduce per-cpu caching to iova allocation • 22e2f9fa63b0 iommu/vt-d: Use per-cpu IOVA caching
	Keywords: IOMMU, soft lockup
	Discovered in Release: 4.2-1.0.0.0
1135738	**Description**: On 64k page size setups, DMA memory might run out when trying to increase the ring size/number of channels.
	Workaround: Reduce the ring size/number of channels.
	Keywords: DMA, 64K page
	Discovered in Release: 4.2-1.0.0.0

Table 8 - Known Issues

Internal Reference Number	Issue
1159650	**Description**: When configuring VF VST, VLAN-tagged outgoing packets will be dropped in case of ConnectX-4 HCAs. In case of ConnectX-5 HCAs, VLAN-tagged outgoing packets will have another VLAN tag inserted.
	Workaround: N/A
	Keywords: VST
	Discovered in Release: 4.2-1.0.0.0
1157770	**Description**: On Passthrough/VM machines with relatively old QEMU and libvirtd, CMD timeout might occur upon driver load. After timeout, no other commands will be completed and all driver operations will be stuck.
	Workaround: Upgrade the QEMU and libvirtd on the KVM server. Tested with (Ubuntu 16.10) are the following versions: • libvirt 2.1.0 • QEMU 2.6.1
	Keywords: QEMU
	Discovered in Release: 4.2-1.0.0.0
1147703	**Description**: Using dm-multipath for High Availability on top of NVMEoF block devices must be done with “directio” path checker.
	Workaround: N/A
	Keywords: NVMEoF
	Discovered in Release: 4.2-1.0.0.0
1152408	**Description**: RedHat v7.3 PPCLE and v7.4 PPCLE operating systems do not support KVM qemu out of the box. The following error message will appear when attempting to run `virt-install` to create new VMs: `Cant find qemu-kvm packge to install`
	Workaround: Acquire the following rpms from the beta version of 7.4ALT to 7.3/ 7.4 PPCLE (in the same order): • qemu-img-.el7a.ppc64le.rpm • qemu-kvm-common-.el7a.ppc64le.rpm • qemu-kvm-.el7a.ppc64le.rpm
	Keywords: Virtualization, PPC, Power8, KVM, RedHat, PPC64LE
	Discovered in Release: 4.2-1.0.0.0
1012719	**Description**: A soft lockup in the CQ polling flow might occur when running very high stress on the GSI QP (RDMA-CM applications). This is a transient situation after which the driver will recover from.
	Workaround: N/A
	Keywords: RDMA-CM, GSI QP, CQ
	Discovered in Release: 4.2-1.0.0.0

Table 8 - Known Issues

Internal Reference Number	Issue
1062940	**Description**: When running Network Manger on devices on which Enhanced IPoIB is enabled, CONNECTED_MODE can only be set to NO/AUTO. Setting it to YES will prevent the interface from being configured.
	Workaround: N/A
	Keywords: Enhanced IPoIB, network manager, connected_mode
	Discovered in Release: 4.2-1.0.0.0
1078630	**Description**: When working in RoCE LAG over kernel v3.10, a kernel crash might occur when unloading the driver as the Network Manager is running.
	Workaround: Stop the Network Manager before unloading the driver and start it back once the driver unload is complete.
	Keywords: RoCE LAG, network manager
	Discovered in Release: 4.2-1.0.0.0
1149557	**Description**: When setting VGT+, the maximal number of allowed VLAN IDs presented in the sysfs is 813 (up to the first 813).
	Workaround: N/A
	Keywords: VGT+
	Discovered in Release: 4.2-1.0.0.0
1122619	**Description**: On Arm setups, DMA memory resource is limited due to a default CMA limitation.
	Workaround: Increase the CMA limitation or cancel its use, using the kernel's CMD line parameters: • Add the parameter cma=256M to increase the CMA limit to 256MB • Add the parameter cma=0 to disable the use of CMA
	Keywords: IPoIB, CMA
	Discovered in Release: 4.2-1.0.0.0
1146837	**Description**: On SLES11 SP1 operating system, IPoIB interface renaming process may fail due to a broken udev rule, leaving interfaces with names like ib0_rename.
	Workaround: 1. Open the udev conf file "/etc/udev/rules.d/70-persistent-net.rules", and remove such lines as SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", =="" , NAME="eth0". 2. Reload the driver stack.
	Keywords: IPoIB
	Discovered in Release: 4.2-1.0.0.0

Table 8 - Known Issues

Internal Reference Number	Issue
965591	**Description**: Lustre support is limited to versions 2.9 and 2.10.
	Workaround: N/A
	Keywords: Lustre
	Discovered in Release: 4.1-1.0.2.0
-	**Description:** iSER Target currently supports the following OSs (distribution kernel) only: • RHEL 7.2/7.3 • SLES 12.1/12.2 • Ubuntu14.04/15.04/16.04/16.10
	Workaround: N/A
	Keywords: iSER Target
-	**Description:** NVMEoF support is available for the following: • SLES 12.3 and above • RHEL 7.2 and above (Host side only) • RHEL 7.4 and above (Host and Target side) • OS with distribution/custom kernel >= 4.8.x
	Workaround: N/A
	Keywords: NVMEoF Host/Target
995665	**Description**: Connection between NVMEoF host and target cannot be established in a hyper-threaded system with more than 64 CPUs on the NVMEoF host side.
	Workaround: On the host side, connect to NVMEoF subsystem using `--nr-io-queues <num_queues>` flag. Note that `num_queues` must be lower or equal to num_sockets multiplied with num_cores_per_socket.
	Keywords: NVMEoF
1039346	**Description**: Enabling multiple namespaces per subsystem while using NVMEoF target offload is not supported.
	Workaround: To enable more than one namespace, create a subsystem for each one.
	Keywords: NVMEoF Target Offload, namespace
1072347	**Description**: Ethtool -i <ibx> displays incorrect driver name for devices with enhanced IPoIB support.
	Workaround: N/A
	Keywords: Enhanced IPoIB, Ethtool

Table 8 - Known Issues

Internal Reference Number	Issue
1071457	**Description**: PKEY-related limitations in enhanced IPoIB: • Since the parent interface ib<x> and the child interface ib<x>.yyyy share the same receive resources, the parent interface's MTU cannot be less than the child interface's MTU • Interface counters and Ethtool control are not supported on child interfaces • Parent interface should be in UP state to enable child interface to receive traffic
	Workaround: N/A
	Keywords: PKEY, Enhanced IPoIB, MTU, Ethtool, Interface Counters
1059451	**Description**: When Enhanced IPoIB is enabled, the following module parameters will not be functional: • send_queue_size • recv_queue_size • max_nonsrq_conn_qp
	Workaround: N/A
	Keywords: Enhance IPoIB
1030301	**Description**: Creating virtual functions on a device that is in LAG mode will destroy the LAG configuration. The boding device over the Ethernet NICs will continue to work as expected.
	Workaround: N/A
	Keywords: LAG, SR-IOV
1047616	**Description**: When node GUID of a device is set to zero (0000:0000:0000:0000), RDMA_CM user space application may crash.
	Workaround: Set node GUID to a nonzero value.
	Keywords: RDMA_CM
1061298	**Description**: Since enhanced IPoIB does not support connected mode on RedHat operating systems, when using network manger and enhanced IPoIB capable devices, `CONNECTED_MODE` must be set to NO/AUTO. Setting `CONNECTED_MODE` to yes will cause the interface to not be configured.
	Workaround: N/A
	Keywords: Enhanced IPoIB
1068215	**Description**: When enhanced IPoIB mode is enabled, ring size limit is 8k. When it is disabled, ring size limit is decreased to 4k.
	Workaround: N/A
	Keywords: Enhanced IPoIB
1051701	**Description**: New versions of iproute which support new kernel features may misbehave on old kernels that do not support these new features.
	Workaround: N/A
	Keywords: iproute

Table 8 - Known Issues

Internal Reference Number	Issue
1006032	**Description**: Currently, iSER initiator supports T10-PI offload mechanism only in the following OSs: • Ubuntu 16.10, 16.04 • RedHat 7.3, 7.2 • SLES 12.02
	Workaround: N/A
	Keywords: iSER initiator
1007830	**Description**: When working on Xenserver hypervisor with SR-IOV enabled on it, make sure the following instructions are applied: 1. Right after enabling SR-IOV, unbind all driver instances of the virtual functions from their PCI slots. 2. It is not allowed to unbind PF driver instance while having active VFs.
	Workaround: N/A
	Keywords: SR-IOV
1008583	**Description**: A soft lockup in the CQ polling flow might occur when running very high stress on the GSI QP (RDMA-CM applications). This is a transient situation and the driver recovers from it after a while.
	Workaround: N/A
	Keywords: RDMA-CM
1007356	**Description**: Creating a PKEY interface using "`ip link`" is not supported.
	Workaround: Use sysfs to create a PKEY interface.
	Keywords: IPoIB, PKEY
1000197	**Description:** Displaying multicast groups using sysfs may not show all the entries on Fedora 23 OS.
	Workaround: N/A
	Keywords: IPoIB
1010148	**Description:** Upgrading from MLNX_OFED v3.x to v4.x using yum and apt-get repositories fails.
	Workaround: Remove MLNX_OFED v3.x using the `ofed_uninstall.sh` script, and only then install MLNX_OFED v4.x as usual.
	Keywords: Installation

Table 8 - Known Issues

Internal Reference Number	Issue

1005786	**Description:** When using ConnectX-5 adapter cards, the following error might be printed to dmesg, indicating temporary lack of DMA pages:
	`“mlx5_core ... give_pages:289:(pid x): Y pages alloc time exceeded the max permitted duration`
	`mlx5_core ... page_notify_fail:263:(pid x): Page allocation failure notification on func_id(z) sent to fw`
	`mlx5_core ... pages_work_handler:471:(pid x): give fail -12”`
	
	Example: This might happen when trying to open more than 64 VFs per port.

	Workaround: N/A
	Keywords: mlx5_core, DMA
1008066/ 1009004	**Description:** Performing some operations on the user end during reboot might cause call trace/panic, due to bugs found in the Linux kernel. For example: Running `get_vf_stats` (via iptool) during reboot.
	Workaround: N/A
	Keywords: mlx5_core, reboot

1009488	**Description:** Mounting MLNX_OFED to a path that contains special characters, such as parenthesis or spaces is not supported. For example, when mounting MLNX_OFED to “/media/CDROM(vcd)/”, installation will fail and the following error message will be displayed:
	`# cd /media/CDROM\(vcd\)/`
	`# ./mlnxofedinstall`
	`sh: 1: Syntax error: "(" unexpected`

	Workaround: N/A
	Keywords: Installation
982144	**Description:** When offload traffic sniffer is on, the bandwidth could decrease up to 50%.
	Workaround: N/A
	Keywords: Offload Traffic Sniffer
981045	**Description:** On kernels below v4.2, when removing a bonding module with devices different from ARPHRD_ETHER, a call trace may be received.
	Workaround: Remove the bond in the following order: Remove the slaves, delete the bond, and only then remove the bonding module.
	Keywords: Bonding
980066/981314	**Description:** Soft RoCE does not support Extended Reliable Connection (XRC).
	Workaround: N/A
	Keywords: Soft RoCE, XRC

Table 8 - Known Issues

Internal Reference Number	Issue
982534	**Description:** In ConnectX-3, when using a server with page size of 64K, the UAR BAR will become too small. This may cause one of the following issues: 1. mlx4_core driver does not load. 2. The mlx4_core driver does load, but calls to `ibv_open_device` may return ENOMEM errors.
	Workaround: 1. Add the following parameter in the firmware's ini file under [HCA] section: `log2_uar_bar_megabytes = 7` 2. Re-burn the firmware with the new ini file.
	Keywords: PPC
981362	**Description:** On several OSs, setting a number of TC is not supported via the tc tool.
	Workaround: Set the number of TC via the /sys/class/net/<interface>/qos/tc_num sysfs file.
	Keywords: Ethernet, TC
980257	**Description:** An issue in InfiniBand bond interfaces may cause memory corruption in Ubuntu v14.04 and v14.10 OSs. The memory corruption happens when attempting to reload the driver while the bond is up with InfiniBand salves.
	Workaround: Delete the bond before restarting the driver.
	Keywords: Bonding, IPoIB
980034/981311	**Description:** Soft RoCE counters located under /sys/class/infiniband/<rxe-inf>/ports/1/counters/ directory are not supported.
	Workaround: N/A
	Keywords: Soft RoCE
979907	**Description:** Only the following two experimental verbs are supported for Soft RoCE: • ibv_exp_query_device • ibv_exp_poll_cq.
	Workaround: N/A
	Keywords: Soft RoCE
979457	**Description:** When setting IOMMU=ON, a severe performance degradation may occur due to a bug in IOMMU.
	Workaround: Make sure the following patches are found in your kernel: • iommu/vt-d: Fix PASID table allocation • iommu/vt-d: Fix IOMMU lookup for SR-IOV Virtual Functions **Note**: These patches are already available in Ubuntu 16.04.02 and 17.04 OSs.
	Keywords: Performance, IOMMU

Table 8 - Known Issues

Internal Reference Number	Issue
977852	**Description:** `rdma_cm` running over IB ports does not support UD QPs on ConnectX-3 adapter cards.
	Workaround: N/A
	Keywords: SR-IOV, RDMA CM
955113/977990	**Description:** In RoCE LAG over ConnectX-4 adapter cards, the script ibdev2netdev may show a wrong port state for the bonded device. This means that although the IB device/port mlx5_bond_0/1 is up (as seen in ibstat), ibdev2netdev may report that it is down.
	Workaround: N/A
	Keywords: RoCE, LAG, bonding
942161	**Description:** On some kernels, there might be an issue in csum calculations of tunneled packets when the driver sets CHECKSUM_COMPLETE for the packet. This might print csum error messages to the dmesg log file.
	Workaround: Make sure your kernel version includes this fix.
	Keywords: Ethernet, checksum, tunneling
931574	**Description:** When using a kernel with Generic Receive Offload (GRO) support, UDP performance results will reveal degradation in comparison to the UDP performance results in MLNX_OFED v1.5.x.
	Workaround: Turn off the GRO feature to get better UDP performance. Run: `#ethtool -K <interface> gro off`
	Keywords: GRO, UDP, performance
920707	**Description:** In SLES12 SP2, you may get a memory low warning at the netlink layer when configuring a large number of VFs.
	Workaround: N/A
	Keywords: SR-IOV, SLES12
969467	**Description:** On SLES PPC64, the removal of packages with names starting with kernel-mft-mlnx might fail with such an error: `"Error: package kernel-mft-mlnx-kmp-default seems to contain modules for multiple kernel versions"`
	Workaround: Use the following command to remove the kernel-mft packages: `rpm -e --noscripts $(rpm -qa | grep kernel-mft-mlnx)`
	Keywords: Installation
918880	**Description:** The driver version shown in modinfo and ethtool outputs is 3.4-1.0.6 instead of 3.4-2.0.0.
	Workaround: N/A
	Keywords: Installation

Table 8 - Known Issues

Internal Reference Number	Issue

679801	**Description:** Updating MLNX_OFED via Yum (e.g. running "`yum update mlnx-ofed-all`") can fail with the following error:
	--> Finished Dependency Resolution
	`Error: Package: mpitests_openmpi__1_8_8-3.2.16-fe5387c.x86_64 (installed)`
	`Requires: liboshmem.so.3()(64bit)`
	`Removing: openmpi-1.8.8-1.x86_64 (installed)`
	`liboshmem.so.3()(64bit)`
	`Updated By: openmpi-1.10.2rc4-1.32008.x86_64 (mlnx_ofed)`
	`~liboshmem.so.9()(64bit)`

	Workaround: Remove the mpitests packages manually: `# rpm -e --allmatches $(rpm -qa | grep mpitests_)`
	Keywords: Installation

690799	**Description:** OpenSM package removal fails with the following error on Ubuntu12.04:
	`Removing opensm ...`
	`/sbin/insserv: No such file or directory`

	Workaround: 1. Create the missing link by running this command: `# ln -s /usr/lib/insserv/insserv /sbin/insserv` 2. Remove the package.
	Keywords: Installation
764204	**Description:** 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.
	Workaround: • As of MLNX_OFED v3.3, use the mlnx_add_kernel_support.sh script, or simply provide the --add-kernel-support flag to mlnxofedinstall script. • Update the kmod package using the following link: https://rhn.redhat.com/errata/RHBA-2016-1832.html
	Keywords: Installation

Table 8 - Known Issues

Internal Reference Number	Issue
785119	**Description:** 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 mlxfwreset 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."`
	Workaround: N/A
	Keywords: Installation
773774	**Description:** 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`

	Workaround: 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`

	Keywords: Driver Start
677998	**Description:** False alarm errors may be printed to dmesg.
	Workaround: N/A
	Keywords: Driver Start
610395	**Description:** 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.
	Workaround: Use the `mlnx_add_kernel_support.sh` script to compile MLNX_OFED drivers against the new kernel.
	Keywords: Driver Start
967356	**Description:** [**Ethernet**] • Bare-metal ConnectX-4/ConnectX-4 Lx might suffer up to 15, degradation in some scenarios due to higher CPU utilization. • PPC: ConnectX-4 might suffer up to 20, degradation in some scenarios.
	Workaround: N/A
	Keywords: Performance

Table 8 - Known Issues

Internal Reference Number	Issue
956071	**Description:** [**mlx5**] OOB TCP performance for small message sizes may suffer from lower BW than expected.
	Workaround: Disable adaptive-rx and set higher static moderation: `ethtool -C <interface> adaptive-rx off rx-frames 128 rx-usecs 128`
	Keywords: Performance
765777	**Description:** Low VxLAN throughput due to broken GRO offload in most kernels older than kernel v4.6.
	Workaround: Use kernel version 4.6 or above.
	Keywords: Performance
414827	**Description:** Out-of-the-box throughput performance in Ubuntu14.04 is not optimal and may achieve results below the line rate in 40GE link speed.
	Workaround: For additional performance tuning, please refer to Performance Tuning Guide.
	Keywords: Performance
417751	**Description:** Performance degradation might occur when bonding Ethernet interfaces.
	Workaround: N/A
	Keywords: Performance
656415	**Description:** In RHEL7.0, when the irqbalance service is started or restarted, it incorrectly re-balances the IRQs, including the banned ones.
	Workaround: N/A
	Keywords: Performance
651322	**Description:** 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
	Workaround: Use RH7.2 to avoid such performance issues.
	Keywords: Performance
754646	**Description:** The default RX coalescing values yield to high CPU utilization when using VXLAN on VMs over PV.
	Workaround: Increase the RX microseconds and frames coalescing parameters for a better utilization using the ethtool -C command.
	Keywords: Performance

Table 8 - Known Issues

Internal Reference Number	Issue
783496	**Description:** When using a VF over RH7.X KVM, low throughput is expected.
	Workaround: Install the following packages using the link below: • 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/
	Keywords: Performance
780782/870171	**Description:** CALC operation on PowerPC may report completion with error.
	Workaround: N/A
	Keywords: mlx5 Driver
921252	**Description:** Support for Memory Window impacts the maximum number of Work Requests (WRs) for connected QPs (for example: RC QP). Applications that create a QP with the maximal size might experience out of memory errors.
	Workaround: Lower the number of Work Requests during the QP creation.
	Keywords: mlx5 Driver
860311	**Description:** An allocation of high-order page in mlx5e_alloc_striding_rx_wqe fails with a call-trace.
	Workaround: No action is required on users end. A fragmented fallback flow will handle this failure.
	Keywords: mlx5 Driver
435583	**Description:** EEH events that arrive while the mlx5 driver is loading may cause the driver to hang.
	Workaround: N/A
	Keywords: mlx5 Driver
434570	**Description:** The mlx5 driver can handle up to 5 EEH events per hour.
	Workaround: If more events are received, cold reboot the machine.
	Keywords: mlx5 Driver
554120	**Description:** When working with Connect-IB firmware v10.10.5054, the following message would appear in driver start. `command failed, status bad system state(0x4), syndrome 0x408b33` The message can be safely ignored.
	Workaround: Upgrade Connect-IB firmware to the latest available version.
	Keywords: mlx5 Driver

Table 8 - Known Issues

Internal Reference Number	Issue
538843	**Description:** Bonding active-backup mode does not function properly.
	Workaround: N/A
	Keywords: mlx5 Driver
-	**Description:** Rate, speed and width using IB sysfs/tools are available in RoCE mode in ConnectX-4 only after port physical speed configuration is done.
	Workaround: N/A
	Keywords: mlx5 Driver
598092	**Description:** Since MLNX_OFED's openibd does not unload modules while OpenSM is running, removing mlx5_core manually while OpenSM is running, may cause it to be out of sync when probed again.
	Workaround: Restart OpenSM
	Keywords: mlx5 Driver
563022	**Description:** ConnectX-4 port GIDs table shows a duplicated RoCE v2 default GID.
	Workaround: N/A
	Keywords: mlx5 Driver
947542	**Description:** mlx5 hardware offload is supported when setting up to 4 VxLAN ports (one of these ports must be 4789). When attempting to set more VxLAN ports, these ports will still be supported, but a failure message will appear in the dmesg.
	Workaround: N/A
	Keywords: Ethernet
964991	**Description:** TX queue rate limit may sometimes exceed the rate that was set by the user by up to 10,.
	Workaround: N/A
	Keywords: Ethernet
911693	**Description:** In ConnectX-4 Lx and above, the minimal RX ring size is changed to 512, as a result of fundamental changes in receive flow structures.
	Workaround: N/A
	Keywords: Ethernet
948312	**Description:** [**ConnectX-3 Pro**] To enable/disable `rx-vlan-stag-hw-parse` by ethtool, rxvlan should be enabled/disabled accordingly (ethtool -K rxvlan on/off).
	Workaround: N/A
	Keywords: Ethernet

Table 8 - Known Issues

Internal Reference Number	Issue
894547	**Description:** On SLES12 SP1 and SLES12 SP2, invalid udev rules might cause Ethernet interfaces renaming to fail, leaving some interfaces with names such as renameXY.
	Workaround: Modify the udev rules inside the /etc/udev/rules.d/70-persistent-net.rules file, such that every rule is unique to the target interface. For further details, refer to the Ethernet Related Issues table under the Troubleshooting section in MLNX_OFED User Manual.
	Keywords: Ethernet
754709	**Description:** mlx5 Ethernet auto-negotiation related issues: 1. The command ethtool -s eth4 speed 25000 autoneg on is not a valid ethtool command. Speed 25000 should not be passed in when autoneg is on. Instead, use advertised 0x100000000. 2. ethtool version older than v4.6 does not report neither supported or advertised new speeds, such as 25G, 100G. 3. When setting auto negotiation with an ethtool version older than v4.6, or a kernel version with no set_link_ksettings/get_link_ksettings API, advertised speed will be ignored, and the device will try to reach the highest supported speed available end-to-end. 4. When using an ethtool version of 4.6 or newer, and a kernel version with set_link_ksettings/get_link_ksettings API, users can set an advertised speed only for ones which are part of include/uapi/linux/ethtool.h file. 5. New speeds (25G, 50G, 100G) will be shown in Supported/Advertised fields of ethtool, only if known to this ethtool version. Auto negotiation for those speeds might work, depending on the kernel's advertised information. 6. Upon power cycle, all supported speeds will be copied to Advertised field.
	Workaround: N/A
	Keywords: Ethernet
843306	**Description:** [**ConnectX-4/ConnectX-4 Lx**] When configuring ETS, bandwidth values are limited between 1-100, and 0 is an invalid value.
	Workaround: N/A
	Keywords: Ethernet
704750	**Description:** [**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.
	Workaround: N/A
	Keywords: Ethernet

Table 8 - Known Issues

Internal Reference Number	Issue
433366	**Description:** 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"`
	Workaround: N/A
	Keywords: Ethernet
539117	**Description:** On SLES12, the bonding interface over Mellanox Ethernet slave interfaces does not get IP address after reboot.
	Workaround: 1. Set "STARTMODE=hotplug" in the bonding slave's ifcfg files.More details can be found in the SUSE documentations page: https://www.suse.com/documentation/sles-12/book_sle_admin/?page=/documentation/sles-12/book_sle_admin/data/sec_bond.html 2. Enable the nanny service to support hot-plugging: Open the "`/etc/wicked/common.xml`" file. Change: "`<use-nanny>false</use-nanny>`" to "`<use-nanny>true</use-nanny>`" 3. Run: `# systemctl restart wickedd.service wicked`
	Keywords: Ethernet
989042	**Description:** `ethtool -x` command will not function on relatively old kernels that do not support get/set_rxfh* callbacks.
	Workaround: N/A
	Keywords: Ethernet
516136	**Description:** Ethertype proto 0x806 not supported by ethtool
	Workaround: N/A
	Keywords: Ethernet
592229	**Description:** When NC-SI is ON, the ports MTU cannot be set to lower than 1500.
	Workaround: N/A
	Keywords: Ethernet
600242	**Description:** GRO is not functional when using VXLAN in ConnectX-3 adapter cards.
	Workaround: N/A
	Keywords: Ethernet
596075	**Description:** ethtool -X: The driver supports only the 'equal' mode and cannot be set by using weight flags.
	Workaround: N/A
	Keywords: Ethernet

Table 8 - Known Issues

Internal Reference Number	Issue
600752	**Description:** Q-in-Q infrastructure in the kernel is supported only in kernel version 3.10 and up.
	Workaround: N/A
	Keywords: Ethernet
596537	**Description:** 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.
	Workaround: N/A
	Keywords: Ethernet
560575	**Description:** When using a hardware that has Time Stamping enabled, the system time might be higher than the expected variance.
	Workaround: N/A
	Keywords: Ethernet
597758	**Description:** In Q-in-Q, ping failed when sending traffic with package size > 1468
	Workaround: N/A
	Keywords: Ethernet
665131	**Description:** Call trace may occur when configuring VXLAN or under high traffic stress.
	Workaround: N/A
	Keywords: Ethernet
685069/689607	**Description:** ethtool header does not currently support the link speeds of 25/50/100. Therefore, these speeds cannot be seen as advertised/supported.
	Workaround: N/A
	Keywords: Ethernet
835239	**Description:** While running Q-in-Q packets with stag offloading, tcpsump/wireshark on host may show svlan ethertype as 0x8100 instead of 0x88A8.
	Workaround: Check the wire or a switch between the hosts, the wireshark will show 0x88A8 ethertype as expected.
	Keywords: Ethernet
954924/954994	**Description:** Accelerated Receive Flow Steering (aRFS) does not work properly with more than 50 streams. Thus, packets are not forwarded based on the location of the application consuming the packet.
	Workaround: N/A
	Keywords: Flow Steering

Table 8 - Known Issues

Internal Reference Number	Issue
516136	**Description:** Setting ARP flow rules through ethtool is not allowed.
	Workaround: N/A
	Keywords: Flow Steering
448981	**Description:** QoS default settings are not returned after configuring QoS.
	Workaround: N/A
	Keywords: Quality of Service
940345	**Description:** In ConnectX-3, when the virtual function (VF) runs on a MLNX-_OFED version that is below v4.0, and the physical function runs on MLNX_OFED v4.0 and higher, hardware counters in the VF will be set to zero and will not progress.
	Workaround: N/A
	Keywords: Ethernet Performance Counters
891241/967659	**Description:** Sysfs for displaying neighbor information is not supported.
	Workaround: N/A
	Keywords: IPoIB
920440	**Description:** Adaptive RX moderation in not supported.
	Workaround: To improve RX performance, manually configure RX moderation using ethtool. It is recommended to use rx-usecs 16 and rx-frames 88 for datagram mode: Example: `on ib0: ethtool -C ib0 rx-usecs 16 rx-frames 88`
	Keywords: IPoIB
965910	**Description:** On RHEL7.3, when creating a PKEY using the ifcfg file with ConnectX-3 and ConnectX-3 Pro adapter cards, `ifdown` and `ifup` commands must be run respectively.
	Workaround: N/A
	Keywords: IPoIB
854235	**Description:** IPoIB bonding interface has to be restarted in order to work on some operating systems.
	Workaround: Toggle bonding interface to state down and then to state up.
	Keywords: IPoIB
552840	**Description:** `ifdown` command does not function in RH7.x
	Workaround: N/A
	Keywords: IPoIB

Table 8 - Known Issues

Internal Reference Number	Issue
665143	**Description:** Kernel Oops may occur after reboot.
	Workaround: N/A
	Keywords: IPoIB
555632	**Description:** Kernel panic may occur while re-assigning LIDs.
	Workaround: N/A
	Keywords: IPoIB
556352	**Description:** ICMP traffic might be lost after Vnic restart
	Workaround: N/A
	Keywords: IPoIB
560575	**Description:** Spikes may occur while running PTP protocol over ConnectX-3/ConnectX-3 Pro.
	Workaround: N/A
	Keywords: IPoIB

684720	**Description:** `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.

	Workaround: 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`

	Keywords: IPoIB
766451	**Description:** 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
	Workaround: N/A
	Keywords: IPoIB

Table 8 - Known Issues

Internal Reference Number	Issue
383034	**Description:** On rare occasions, upon driver restart the following message is shown in the dmesg: `'cannot create duplicate filename '/class/net/eth_ipoib_interfaces'`
	Workaround: N/A
	Keywords: eIPoIB
855362	**Description:** A compilation error will occur in kernel space application when setting the wr_id field upon initializing any of the following structures: `ib_wc`, `ib_send_wr`, or `ib_recv_wr`. This is caused due to `wr_id` insertion into an anonymous union.
	Workaround: Assign the enum field explicitly. For example: `wr.wr_id = MY_WR_ID;`
	Keywords: Verbs
835061	**Description:** According to the verbs header (/usr/include/infiniband/verbs.h), 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.
	Workaround: Run QP command query to verify the value.
	Keywords: Verbs
935250	**Description:** When NUM_OF_VFS in firmware capabilities is 32 or higher, ConnectX-4 RoCE LAG will not be supported. This is true even if driver does not have SR-IOV enabled (no VFs are present).
	Workaround: 1. Verify the `NUM_OF_VFS` value by running: `mlxconfig -d /dev/mst/mt4115_pciconf0 query  |grep NUM_OF_VFS` Output: `NUM_OF_VFS          32` 2. Change NUM_OF_VFS value by running: `mlxconfig -d /dev/mst/mt4115_pciconf0 set NUM_OF_VFS=0` 3. Reboot the machine.
	Keywords: RoCE
935250	**Description:** RoCE LAG for mlx5 is supported in kernel v4.5 and above.
	Workaround: N/A
	Keywords: RoCE
869158	**Description:** Occasionally, UC|UD traffic over default GIDs, with high iterations may get stuck.
	Workaround: N/A
	Keywords: RoCE

Table 8 - Known Issues

Internal Reference Number	Issue
854517	**Description:** Driver restart while having RDMA-CM running applications may hang.
	Workaround: N/A
	Keywords: RoCE
392592	**Description:** On rare occasions, the driver reports a wrong GID table (read from /sys/class/infiniband/mlx4_*/ports/*/gids/*). This may cause communication problems.
	Workaround: N/A
	Keywords: RoCE
559276/591244	**Description:** Dynamically Connected (DC) in RoCE in ConnectX-4 is currently not supported.
	Workaround: N/A
	Keywords: RoCE
517825	**Description:** `ibv_create_ah_from_wc` is not supported for multicast messages.
	Workaround: N/A
	Keywords: RoCE
609950/649407	**Description:** Occasionally, when the Bonding Mode is set to other than active/backup mode (mode 1), the GID table is not populated correctly.
	Workaround: Add slave devices to the master before giving it an IP address.
	Keywords: RoCE
667399	**Description:** In ConnectX-4 adapter cards, when the port speed is lower than 10Gbps, the IB tools will present a higher rate.
	Workaround: N/A
	Keywords: RoCE
778492	**Description:** 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.
	Workaround: Work in fail_over_mac mode (bonding).
	Keywords: RoCE
781383	**Description:** Creating Address Handler (AH) may run slow or may hang under a heavy load on all nodes cores (for example: MPI All2All cases).
	Workaround: N/A
	Keywords: RoCE

Table 8 - Known Issues

Internal Reference Number	Issue
959452/961438	**Description:** When adding an RXE device to one HCA port, `ibdev2netdev` will show that the RXE device has been added to both HCA ports.
	Workaround: N/A
	Keywords: Soft RoCE
963537/963546	**Description:** When adding an RXE device to an HCA that supports RoCE, this might cause segmentation fault when running verbs applications due to a conflict between the librxe and the HCA library (for example: libmlx5).
	Workaround: Perform one of the following to avoid the issue: • Uninstall the actual library (For example: uninstall libmlx5) • Remove the library's configuration file (Run: `rm -f /etc/libibverbs.d/mlx5.driver`)
	Keywords: Soft RoCE
664110	**Description:** SDP is currently not supported in mlx5 driver (Connect-IB and Connect-X 4 adapter cards)
	Workaround: N/A
	Keywords: SDP
683370	**Description:** 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.
	Workaround: Set module `param always_register=N` `$ modprobe ib_iser always_register=N`
	Keywords: iSER Initiator
896859	**Description:** ConnectX-3 virtual function that runs on a MLNX_OFED version that is older than v4.0 cannot communicate with ConnectX-3 virtual function that runs on MLNX_OFED v4.0.
	Workaround: Set the OpenSM default alias-guid hop limit to 2 in OpenSM configuration file: aguid_default_hop_limit 2
	Keywords: SR-IOV
858628	**Description:** PCI error handling is not supported during driver reload. This might cause a kernel panic or calltrace.
	Workaround: N/A
	Keywords: SR-IOV

Table 8 - Known Issues

Internal Reference Number	Issue
860385	**Description:** Creating 127 VFs may cause kernel panic in SLES11 SP4 KVM with Kernel 3.0.101-63 because of a IOMMU kernel bug.
	Workaround: N/A
	Keywords: SR-IOV
795697	**Description:** [**mlx4**] While spoof-check filters the incoming traffic to a VM, when this feature is disabled, traffic still does not reach the VM.
	Workaround: The driver must be restarted for the disablement of the feature to take effect and all traffic to be reached to the VM.
	Keywords: SR-IOV
835065	**Description:** [**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.
	Workaround: N/A
	Keywords: SR-IOV
784940	**Description:** 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.
	Workaround: N/A
	Keywords: SR-IOV
784954	**Description:** When SR-IOV is disabled, the VF driver receives `pci_err_detected` event and a teardown flow will be started. During the teardown flow, all firmware commands will fail because the function is already deleted.
	Workaround: N/A
	Keywords: SR-IOV
819595	**Description:** [**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.
	Workaround: N/A
	Keywords: SR-IOV
775944	**Description:** Bonding VFs on the same physical port using bonding mode 0 requires configuration of `fail_over_mac=1`.
	Workaround: N/A
	Keywords: SR-IOV
381764	**Description:** `mlx4_port1_mtu` sysfs entry shows a wrong MTU number in the VM.
	Workaround: N/A
	Keywords: SR-IOV

Table 8 - Known Issues

Internal Reference Number	Issue
426988	**Description:** When at least one port is configured as InfiniBand, and the `num_vfs` is provided but the `probe_vf` is not, HCA initialization fails.
	Workaround: Use both the `num_vfs` and the `probe_vf` in the modprobe line.
	Keywords: SR-IOV
385750/378528	**Description:** 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.
	Workaround: Unload the module mlx4_ib and reload it in the VM.
	Keywords: SR-IOV
392172	**Description:** 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
	Workaround: Shut down the driver in the VM before detaching the VF.
	Keywords: SR-IOV
506512	**Description:** Setting 1 Mbit/s rate limit on Virtual Functions (Qos Per VF feature) may cause TX queue transmit timeout.
	Workaround: N/A
	Keywords: SR-IOV
567908	**Description:** 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.
	Workaround: N/A
	Keywords: SR-IOV
601749	**Description:** 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.
	Workaround: N/A
	Keywords: SR-IOV
649366	**Description:** Restarting the PF (Hypervisor) driver while Virtual Functions are assigned is not allowed in RH7 and above due to a `vfio-pci` bug.
	Workaround: N/A
	Keywords: SR-IOV
639046	**Description:** Due to an issue with SR-IOV loopback, prevention "`Duplicate IPv6 detected`" are seen in the VF driver.
	Workaround: N/A
	Keywords: SR-IOV

Table 8 - Known Issues

Internal Reference Number	Issue
655410	**Description:** [**ConnectX-4/Connect-IB**] Failed to enable SR-IOV due to errors in PCI or BIOS.
	Workaround: 1. Add `pci=realloc=on` to the grub command line. 2. Add more memory to the server. 3. Upgrade BIOS version.
	Keywords: SR-IOV
651119	**Description:** Kernel panic may occur while running IPv6 UDP on SR-IOV ConnectX-4 environment
	Workaround: N/A
	Keywords: SR-IOV
669910	**Description:** Bind/Unbind over ConnectX-4 Hypervisor may cause system lockup.
	Workaround: N/A
	Keywords: SR-IOV
650458	**Description:** Occasionally, IPv6 might not function properly and cause lockup on SR-IOV ConnectX-4 environment.
	Workaround: N/A
	Keywords: SR-IOV
688551	**Description:** In ConnectX-3 adapter cards, the extended counter `port_rcv_-data_64` on the VF may not be updated in some flows.
	Workaround: N/A
	Keywords: SR-IOV
690656/690674	**Description:** When the physical link is down, any traffic from the PF to any VF on the same port will be dropped.
	Workaround: N/A
	Keywords: SR-IOV
691661	**Description:** 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.
	Workaround: Probe one of the VFs in the hypervisor and use for RoCE.
	Keywords: SR-IOV
691661	**Description:** Ethernet SR-IOV in ConnectX-4 requires firmware version 12.14.1100 and higher
	Workaround: N/A
	Keywords: SR-IOV

Table 8 - Known Issues

Internal Reference Number	Issue
737434	**Description:** VF vport statistics are not cleared upon ifconfig up/down.
	Workaround: N/A
	Keywords: SR-IOV
738464	**Description:** In SLES11 SP4, user cannot open all VFs announced in `sriov_totalvfs`. However he can set the num_vfs up to maximum sriov_totalvfs-1 vfs.
	Workaround: N/A
	Keywords: SR-IOV
784127	**Description:** While disabling SR-IOV, all firmware teardown flow commands are expected to fail and error messages will be reported in the dmesg.
	Workaround: N/A
	Keywords: SR-IOV
784146	**Description:** Creating/destroying as many as 64 VFs may sometimes take longer time than usual on some setups.
	Workaround: N/A
	Keywords: SR-IOV
766105	**Description:** 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`".
	Workaround: Upgrade to the latest version of QEMU in the hypervisor.
	Keywords: SR-IOV
413372	**Description:** SR-IOV non persistent configuration (such as VGT, VST, Host assigned GUIDs, and QP0-enabled VFs) may be lost upon Reset Flow.
	Workaround: Reset Admin configuration post Reset Flow
	Keywords: Reset Flow
926137	**Description:** IPV6 ping does not use device specified with -I parameter in iputils version s20160308.
	Workaround: Use iputils version s20161105 or above.
	Keywords: General
936768	**Description:** When querying the HCA core clock, data will be presented in KHz for all cards.
	Workaround: N/A
	Keywords: General

Table 8 - Known Issues

Internal Reference Number	Issue
856033	**Description:** The following PCIe bus error on Qualcomm Arm processor might appear when mapping a large number of DMA addresses: 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
	Workaround: Edit the kernel parameters (in grub) and add qiommu.identity_map_qiommus=PCIE0_MMU,PCIE4_MMU (The bus numbers depend on the ConnectX-4 slot.) Reboot the server.
	Keywords: General
552870/548518	**Description:** On rare occasions, under extremely heavy MAD traffic, MAD (Management Datagram) storms might cause soft-lockups in the UMAD layer.
	Workaround: N/A
	Keywords: General
663434	**Description:** On ConnectX-4/ConnectX-4 Lx, when running `"lspci"` in RH7.0/7.1, the device information is displayed incorrect or the device is unnamed.
	Workaround: Run `update-pciids`
	Keywords: General
767016	**Description:** Resetting hardware counters after netdev goes up can break statistics scripts.
	Workaround: N/A
	Keywords: General
959842	**Description:** User space libraries (for example: libibverbs, libmlx4/5) provided by MLNX_OFED v4.0 cannot work with kernel modules provided by an older MLNX_OFED version.
	Workaround: N/A
	Keywords: ABI Compatibility
919836/946847	**Description:** ucamtose fails when using a local loopback IP.
	Workaround: Use the device's interface IP instead of loopback IP.
	Keywords: Connection Manager (CM)

Table 8 - Known Issues

Internal Reference Number	Issue
781382	**Description:** The number of local ports that rdma_cm ID can bind to is limited. This limitation depends on the OS dynamics.
	Workaround: Modify the range of available ports for binding, run: `sysctl net.ipv4.ip_local_port_range="MIN MAX"` The MIN and MAX values can range from 0 to 65535. **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.
	Keywords: Connection Manager (CM)
387061	**Description:** `mlx4_core` can allocate up to 64 MSI-X vectors, an MSI-X vector per CPU.
	Workaround: N/A
	Keywords: Resources Limitation
553657	**Description:** Registering a large amount of Memory Regions (MR) may fail because of DMA mapping issues on RHEL 7.0.
	Workaround: N/A
	Keywords: Resources Limitation
736136	**Description:** The maximum number of HCAs shown by ibstat is 32 HCAs.
	Workaround: N/A
	Keywords: Diagnostic Utilities

4 Bug Fixes History

This table lists the bugs fixed in this release.

For the list of old bug fixes, please refer to Mellanox OFED Archived Bug Fixes file at: http://www.mellanox.com/pdf/prod_software/MLNX_OFED_Archived_Bug_Fixes.pdf

Table 9 - Bug Fixes History

Internal Ref	Issue
1084791	**Description**: Fixed the issue where occasionally, after reboot, rpm commands used to fail and create a core file, with messages such as “Bus error (core dumped)”, causing the openibd service to fail to start.
	Keywords: rpm, openibd
	Discovered in Release: 3.4-2.0.0.0
	Fixed in Release: 4.2-1.0.0.0
960642/960653	**Description**: Added support for `min_tx_rate` and `max_tx_rate` limit per virtual function ConnectX-5 and ConnectX-5 Ex adapter cards.
	Keywords: SR-IOV, mlx5
	Discovered in Release: 4.0-1.0.1.0
	Fixed in Release: 4.2-1.0.0.0
866072/869183	**Description**: Fixed the issue where RoCE v2 multicast traffic using RDMA-CM with IPv4 address was not received.
	Keywords: RoCE
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.2-1.0.0.0
1163835	**Description**: Fixed an issue where `ethtool -P` output was 00:00:00:00:00:00 when using old kernels.
	Keywords: ethtool, Permanent MAC address, mlx4, mlx5
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.2-1.0.0.0
1067158	**Description**: Replaced a few “GPL only” legacy libibverbs functions with upstream implementation that conforms with libibverbs GPL/BSD dual license model.
	Keywords: libibverbs, license
	Discovered in Release: 4.1-1.0.2.0
	Fixed in Release: 4.2-1.0.0.0

Table 9 - Bug Fixes History

Internal Ref	Issue
1119377	**Description:** Fixed an issue where ACCESS_REG command failure used to appear upon RoCE Multihost driver restart in dmesg. Such an error message looked as follows: `mlx5_core 0000:01:00.0: mlx5_cmd_check:705:(pid 20037): ACCESS_REG(0x805) op_mod(0x0) failed, status bad parameter(0x3), syndrome (0x15c356)`
	Keywords: RoCE, multihost, mlx5
	Discovered in Release: 4.1-1.0.2.0
	Fixed in Release: 4.2-1.0.0.0
1122937	**Description:** Fixed an issue where concurrent client requests got corrupted when working in persistent server mode due to a race condition on the server side.
	Keywords: librdmacm, rping
	Discovered in Release: 4.1-1.0.2.0
	Fixed in Release: 4.2-1.0.0.0
1102158	**Description:** Fixed an issue where client side did not exit gracefully in RTT mode when the server side was not reachable.
	Keywords: librdmacm, rping
	Discovered in Release: 4.1-1.0.2.0
	Fixed in Release: 4.2-1.0.0.0
1038933	**Description:** Fixed a backport issue where IPv6 procedures were called while they were not supported in the underlying kernel.
	Keywords: iw_cm
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
1064722	**Description:** Added log debug prints when changing HW configuration via DCB. To enable log debug prints, run: `ethtool -s <devname> msglvl hw on/off`
	Keywords: DCB, msglvl
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
1013076	**Description:** Fixed the issue where reassembly of packets larger than 64k might have failed when ipfrag threshold was low. This issue was present only on RHEL 6.3, 6.4, 6.5, and Ubuntu 12.04. This packet drop could be seen from the netstat tool, indicated by the “packet reassembles failed” counter.
	Keywords: IPoIB, Packet Fragmentation
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0

Table 9 - Bug Fixes History

Internal Ref	Issue
1022251	**Description:** Fixed SKB memory leak issue that was introduced in kernel 4.11, and added warning messages to the Soft RoCE driver for easy detection of future SKB leaks.
	Keywords: Soft RoCE
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
1044546	**Description:** Fixed the issue where a kernel crash used to occur when RXe device was coupled with a virtual (dummy) device.
	Keywords: Soft RoCE
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
1047617	**Description:** Fixed the issue where a race condition in the RoCE GID cache used to cause for the loss of IP-based GIDs.
	Keywords: RoCE, GID
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
1006768	**Description:** Fixed the issue where an rdma_cm connection between a client and a server that were on the same host was not possible when working over VLAN interfaces.
	Keywords: RDMACM
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
801807	**Description:** Fixed an issue where RDMACM connection used to fail upon high connection rate accompanied with the error message: `RDMA_CM_EVENT_UNREACHABLE`.
	Keywords: RDMACM
	Discovered in Release: 3.0-2.0.1
	Fixed in Release: 4.1-1.0.2.0
869768	**Description:** Fixed the issue where SR-IOV was not supported in systems with a page size greater than 16KB.
	Keywords: SR-IOV, mlx5, PPC
	Discovered in Release: 4.0-2.0.0.1
	Fixed in Release: 4.1-1.0.2.0
919545	**Description:** Fixed the issue of when the Kernel becomes out of memory upon driver start, it could crash on SLES 12 SP2.
	Keywords: mlx_5 Eth Driver
	Discovered in Release: 3.4-2.0.0.0
	Fixed in Release: 4.0-2.0.0.1

Table 9 - Bug Fixes History

Internal Ref	Issue
970668	**Description:** Fixed the issue where very high stress on DC QP transport might have triggered NMI messages on specific servers.
	Keywords: mlx5 Driver
	Discovered in Release: 4.0-1.0.1.0
	Fixed in Release: 4.0-2.0.0.1
966134	**Description:** Allowed Ethernet VFs to open Raw Ethernet QPs even if RoCE is not supported for the VF.
	Keywords: mlx4_ib
	Discovered in Release: 3.0-1.0.1
	Fixed in Release: 4.0-2.0.0.1
864063	**Description:** Fixed the issue of when Spoof-check may have been turned on for MAC address 00:00:00:00:00:00.
	Keywords: mlx4
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-2.0.0.1
869209	**Description:** Fixed an issue that caused TCP packets to be received in an out of order manner when Large Receive Offload (LRO) is on.
	Keywords: mlx5_en
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 4.0-2.0.0.1
913319	**Description:** Fixed the issue of low performance when creating many address handles.
	Keywords: libibverbs
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0
912897	**Description:** Added debug prints to `ib_umem_get` function to fix lack of error indication when this function fails.
	Keywords: InfiniBand
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0
945887	**Description:** [**ConnectX-3**] Fixed the issue where multicast traffic over Raw Ethernet QP on virtual functions were received on the same QP (loopback).
	Keywords: SR-IOV
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0

Table 9 - Bug Fixes History

Internal Ref	Issue
920292	**Description:** Fixed three issues in libmlx5 that were found by NVIDIA in the patches that are part of MLNX_OFED v3.4: 1. mlx5_exp_peer_commit_qp returns number of entries = 4 instead of 3. 2. Peer capability check is wrong - should fail the check when there is neither NOR nor GEQ support. 3. Missing break in mlx5_exp_peer_peek_cq. There is now fallthrough in the IBV_EXP_PEER_PEEK_ABSOLUTE case.
	Keywords: libmlx5
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0
890285	**Description:** Fixed the issue where memory allocation for CQ buffers used to fail when increasing the RX ring size.
	Keywords: mlx5_core
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0
867094	**Description:** Fixed the issue where MLNX_OFED used to fail to load on 4K page Arm architecture.
	Keywords: Arm
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 4.0-1.0.1.0
873538	**Description:** Fixed the issue where `biosdavename` running on Redhat 6.x with MLNX_OFED may show the same name to ConnectX-3 Eth port 1 and ConnectX-3 Eth port 2.
	Keywords: biosdavename
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
876329	**Description:** Fixed the issue of when the error flow was re-factored, the reading of the device caps was excluded from the error recovery flow.
	Keywords: mlx5 Driver
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
876419	**Description:** Fixed the issue where kernel panic was observed on openibd stop as a result of querying non-existent bond slave.
	Keywords: mlx4_en
	Discovered in Release: 3.3-2.0.0.0
	Fixed in Release: 3.4-2.0.0.0

Table 9 - Bug Fixes History

Internal Ref	Issue
868665	**Description:** Fixed the issue where kernel panic in `mlx4_en_get_phys_port_id` may occur during server reboot.
	Keywords: mlx4_en
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
882227	**Description:** Fixed the issue of when EEH was injected and the mlx4 tear down code was called, the eqs were not released, causing a page fault.
	Keywords: mlx4_en
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
887348	**Description:** Fixed the issue of when `prof_sel` was invalid, `mlx5_core` failed upon debug print.
	Keywords: mlx5_core
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
898161	**Description:** 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 `get_numa_phys_mask` function.
	Keywords: mlx5_core
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
880269	**Description:** 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.
	Keywords: ib_core
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
887245	**Description:** Fixed the issue where the system used to pick the dummy `ib_isert` module instead of the real module on RHEL with errata kernel.
	Keywords: ib_isert
	Discovered in Release: 3.4-1.0.0.0
	Fixed in Release: 3.4-2.0.0.0
854344	**Description:** Fixed the issue where `mlnx_affinity` script on RHEL/CentOS7.x host did not disable or enable `irqbalancer`.
	Keywords: irqbalancer
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0

Table 9 - Bug Fixes History

Internal Ref	Issue
824736	**Description:** Fixed wrong skprio2UP mapping by removing it and its scripts, such as tc_wrap, from the driver. This mapping should now be done using the kernel's set_egress_map commands. Note: Only for RDMACM over old kernels, the original skprio2UP mapping in tc_wrap remains valid as these kernels do not support set_egress_map.
	Keywords: QoS
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
824775	**Description:** Fixed the issue where starting `ibacm` daemon failed on Debian based distributions with the following message: /etc/init.d/ibacm: line 37: /sbin/start_daemon: No such file or directory.
	Keywords: ibacm
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
799004	**Description:** 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.
	Keywords: IPoIB
	Discovered in Release: 3.0-2.0.0.0
	Fixed in Release: 3.4-1.0.0.0
777733/778099	**Description:** 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.
	Keywords: Arm
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
826686	**Description:** 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.
	Keywords: mlx4_en
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
824130	**Description:** Fixed the issue where ethtool self test used to fail on interrupt test after timeout if mlx4_ib module was not loaded.
	Keywords: mlx4_en
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0

Table 9 - Bug Fixes History

Internal Ref	Issue
855311	**Description:** 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.
	Keywords: mlx4 driver
	Discovered in Release: 3.0-1.0.1
	Fixed in Release: 3.4-1.0.0.0
786720	**Description:** Fixed a crash that used to occur when trying to bring the interface up in a kernel that did not support accelerated RFS (aRFS).
	Keywords: mlx5 driver
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
781747	**Description:** Fixed the issue of when attempting to disable SR-IOV while there are any VF netdevs open, the operation would fail and the driver would hang.
	Keywords: SR-IOV
	Discovered in Release: 3.3-1.0.0.0
	Fixed in Release: 3.4-1.0.0.0
568602	**Description:** 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.
	Keywords: SR-IOV
	Discovered in Release: 3.0-2.0.0
	Fixed in Release: 3.4-1.0.0.0

5 Change Log History

Table 10 - Change Log History

Category	Description
4.1-1.0.2.0	
HCAs: mlx5 Driver	
RoCE Diagnostics and ECN Counters	Added support for additional RoCE diagnostics and ECN congestion counters under /sys/class/infiniband/mlx5_0/ports/1/hw_counters/ directory. For further information, refer to the Understanding mlx5 Linux Counters and Status Parameters Community post.
rx-fcs Offload (ethtool)	Added support for rx-fcs ethtool offload configuration. Normally, the FCS of the packet will be truncated by the ASIC hardware before sending it to the application socket buffer (skb). Ethtool allows to set the rx-fcs not to be truncated, but to pass it to the application for analysis. For more information and usage, refer to Understanding ethtool rx-fcs for mlx5 Drivers Community post.
DSCP Trust Mode	Added the option to enable PFC based on the DSCP value. Using this solution, VLAN headers will no longer be mandatory for use. For further information, refer to the HowTo Configure Trust Mode on Mellanox Adapters Community post.
RoCE ECN Parameters	ECN parameters have been moved to the following directory: /sys/kernel/debug/mlx5/<PCI BUS>/cc_params/ For more information, refer to the HowTo Configure DCQCN (RoCE CC) for ConnectX-4 (Linux) Community post.
Flow Steering Dump Tool	Added support for mlx_fs_dump, which is a python tool that prints the steering rules in a readable manner.
Secure Firmware Updates	Firmware binaries embedded in MLNX_OFED package now support Secure Firmware Updates. This feature provides devices with the ability to verify digital signatures of new firmware binaries, in order to ensure that only officially approved versions are installed on the devices. For further information on this feature, refer to Mellanox Firmware Tools (MFT) User Manual.
Enhanced IPoIB	Added support for Enhanced IPoIB feature, which enables better utilization of features supported in ConnectX-4 adapter cards, by optimizing IPoIB data path and thus, reaching peak performance in both bandwidth and latency. Enhanced IPoIB is enabled by default.
PeerDirect	Added the ability to open a device and create a context while giving PCI peer attributes such as name and ID. For further details, refer to the PeerDirect Programming Community post.
Probed VFs	Added the ability to disable probed VFs on the hypervisor. For further information, see HowTo Configure and Probe VFs on mlx5 Drivers Community post.

Table 10 - Change Log History

Category	Description
Local Loopback	Improved performance by rendering Local loopback (unicast and multicast) disabled by mlx5 driver by default while local loopback is not in use. The mlx5 driver keeps track of the number of transport domains that are opened by user-space applications. If there is more than one user-space transport domain open, local loopback will automatically be enabled.
1PPS Time Synchronization (at **alpha** level)	Added support for One Pulse Per Second (1PPS), which is a time synchronization feature that allows the adapter to send or receive 1 pulse per second on a dedicated pin on the adapter card. For further information on this feature, refer to the HowTo Test 1PPS on Mellanox Adapters Community post.
Precision Time Protocol (PTP)	Added support for PTP feature in IPoIB offloaded devices. This feature allows for accurate synchronization between the distributed entities over the network. The synchronization is based on symmetric Round Trip Time (RTT) between the master and slave devices. The feature is enabled by default. For further information, refer to Running Linux PTP with ConnectX-4 Community post.
Fast Driver Unload	Added support for fast driver teardown in shutdown and kexec flows.
HCAs: ConnectX-5/ConnectX-5 Ex	
NVMEoF Target Offload	Added support for NVMe over fabrics (NVMEoF) offload, an implementation of the new NVMEoF standard target (server) side in hardware. For further information on NVMEoF Target Offload, refer to HowTo Configure NVMEoF Target Offload.
MPI Tag Matching	Added support for offloading MPI tag matching to HCA.
HCAs: All	
RDMA CM	Changed the default RoCE mode on which RDMA CM runs to RoCEv2 instead of RoCEv1. RDMA_CM session requires both the client and server sides to support the same RoCE mode. Otherwise, the client will fail to connect to the server. For further information, refer to RDMA CM and RoCE Version Defaults Community post.
Lustre	Added support for Lustre file system open-source project.
4.0-2.0.2.0	
Operating Systems	Added support for Ubuntu v17.04.
4.0-2.0.0.1	
PCIe Error Counting	[**ConnectX-4/ConnectX-4 Lx**] Added the ability to expose physical layer statistical counters to ethtool.

Table 10 - Change Log History

Category	Description
Multiprotocol Label Switching (MPLS) Tagged Packets Classification	[**ConnectX-4/ConnectX-4 Lx**] Enabled packet flow steering rules with IPv4/IPv6 classification (for raw packet QP (DPDK) only) to work on IPv4/IPv6 over MPLS (Ethertype 0x8847 and 0x8848) encapsulated packets.
RoCE VFs	[**ConnectX-4/ConnectX-4 Lx**] Added the ability to enable/disable RoCE on VFs.
RoCE LAG	[**ConnectX-4/ConnectX-4 Lx** Added support for RoCE over LAG interface.
Standard ethtool	[**ConnectX-4/ConnectX-4 Lx**] Added support for flow steering and rx-all mode.
SR-IOV Bandwidth Share for Ethernet/RoCE (**beta**)	[**ConnectX-4/ConnectX-4 Lx]** Added the ability to guarantee the minimum rate of a certain VF in SR-IOV mode.
Adapter Cards	Added support for ConnectX-5 and ConnectX-5 Ex HCAs.
DSCP ConfigFS Control for RDMA-CM QPs	Added the ability to configure ToS/DSCP for RDMA-CM QPs only.
Soft RoCE (**beta**)	Add software implementation of RoCE that allows RoCE to run on any Ethernet network adapter whether it offers hardware acceleration or not.
NVMe over Fabrics (NVMEoF)	NVMEoF related module installation has been disabled by default. In order to enable it, add the “`--with-nvmf`” installation option to the “mlnxofedinstall” script.
NFS over RDMA (NFSoRDMA)	Removed support for NFSoRDMA drivers. These drivers are no longer provided along with the MLNX_OFED package.
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 “mstconfig” 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 10 - Change Log History

Category	Description
Raw Ethernet Programming	**[ConnectX®-4/ConnectX®-4 Lx]** Added new APIs for enhanced raw Ethernet programming: • Packet Pacing • TCP Segmentation Offload (TSO) • ToS based steering • Flow ID based steering (beta) • VxLAN based steering (beta) For further information, refer to the “Programming” section in OFED User Manual.
Enhanced PCIe Error Recovery	**[ConnectX®-4/ConnectX®-4 Lx]** Enhanced PCIe error recovery by adding the following behaviors to the flow: • 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)	**[ConnectX®-4/ConnectX®-4 Lx]** Added the ability to rate-limit traffic per Virtual Function in SR-IOV mode.
Dynamically tuned Interrupt Moderation (DIM)	**[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.
Dump Configuration	**[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.
Ethernet Counters	**[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.
Mellanox PeerDirect Async (beta level)	**[ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4/ConnectX®-4 Lx]** The experimental PeerDirect Async APIs have been changed to make the implementation of peer clients simpler. Note the following: • 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. To see the API changes, refer to the man page.
ABI Incompatibility	**[ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4/ConnectX®-4 Lx]** 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.

Table 10 - Change Log History

Category	Description
Mellanox Scalable Hierarchical Aggregation Protocol (SHARP™)	**[Connect-IB/ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4]** **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. For further information on SHARP and its configuration, see SHARP Deployment Guide.
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).
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 subnets.

Table 10 - Change Log History

Category	Description
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	• 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	• Added support for the following IPoIB UD QP offloads: • RX check summing (AKA RX csu) • Large Send Offloads (AKA LSO) To see the new IPoIB UD mode, run: "`ethtool -k <interface>`"
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.
Striding WQE User Space	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.

Table 10 - Change Log History

Category	Description
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.
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 46
3.1-1.0.0	
Wake-on-LAN (WOL)	Wake-on-LAN (WOL) is a technology that allows a network professional to remotely power on a computer or to wake it up from sleep mode.
Hardware Accelerated 802.1ad VLAN (Q-in-Q Tunneling)	Q-in-Q tunneling allows the user to create a Layer 2 Ethernet connection between two servers. The user can segregate a different VLAN traffic on a link or bundle different VLANs into a single VLAN.
ConnectX-4 ECN	ECN in ConnectX-4 enables end-to-end congestions notifications between two end-points when a congestion occurs, and works over Layer 3.
RSS Verbs Support for ConnectX-4 HCAs	Receive Side Scaling (RSS) technology allows spreading incoming traffic between different receive descriptor queues. Assigning each queue to different CPU cores allows better load balancing of the incoming traffic and improve performance.
Minimal Bandwidth Guarantee (ETS)	The amount of bandwidth (BW) left on the wire may be split among other TCs according to a minimal guarantee policy.
SR-IOV Ethernet	SR-IOV Ethernet at Beta level
3.0-2.0.1	
Virtualization	Added support for SR-IOV for ConnectX-4/Connect-IB adapter cards.
3.0-1.0.1	
HCAs	Added support for ConnectX®-4 Single/Dual-Port Adapter supporting up to 100Gb/s.

Table 10 - Change Log History

Category	Description
RoCE per GID	RoCE per GID provides the ability to use different RoCE versions/modes simultaneously.
RoCE Link Aggregation (RoCE LAG): ConnectX-3/ConnectX-3 Pro only	RoCE Link Aggregation (available in kernel 4.0 only) provides failover and link aggregation capabilities for mlx4 device physical ports. In this mode, only one IB port that represents the two physical ports, is exposed to the application layer.
Resource Domain Experimental Verbs	Resource domain is a verb object which may be associated with QP and/or CQ objects on creation to enhance data-path performance.
Alias GUID Support in InfiniBand	Enables the `query_gid` verb to return the admin desired value instead of the value that was approved by the SM, to prevent a case where the SM is unreachable or a response is delayed, or if the VF is probed into a VM before their GUID is registered with the SM.
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.

Table 10 - Change Log History

Category	Description
Experimental Verbs	Added the following verbs: • `ibv_exp_create_res_domain` • `ibv_exp_destroy_res_domain` • `ibv_exp_query_intf` • `ibv_exp_release_intf` Added the following interface families: • `ibv_exp_qp_burst_family` • `ibv_exp_cq_family`
2.4-1.0.4	
Bug Fixes	See “Bug Fixes History” on page 46.
2.4-1.0.0	
mlx4_en net-device Ethtool	Added support for Ethtool speed control and advertised link mode.
	Added ethtool txvlan control for setting ON/OFF hardware TX VLAN insertion: `ethtool -k txvlan [on/off]`
	Ethtool report on port parameters improvements.
	Ethernet TX packet rate improvements.
RoCE	RoCE uses now all available EQs and not only the 3 legacy EQs.
InfiniBand	IRQ affinity hints are now set when working in InfiniBand mode.
Virtualization	VXLAN fixes and performance improvements.
libmlx4 & libmlx5	Improved message rate of short massages.
libmlx5	Added ConnectX®-4 device (4114) to the list of supported devices (hca_table),
Storage	Added iSER Target driver.
Ethernet net-device	New adaptive interrupt moderation scheme to improve CPU utilization.
	RSS support of fragmented IP datagram.
Connect-IB Virtual Function	Added Connect-IB Virtual Function to the list of supported devices.
2.3-2.0.5	
mlx5_core	Added the following files under `/sys/class/infiniband/mlx5_0/mr_cache/`: • `rel_timeout`: Defines the minimum allowed time between the last MR creation to the first MR released from the cache. When `rel_timeout = -1`, MRs are not released from the cache • `rel_imm`: Triggers the immediate release of excess MRs from the cache when set to 1. When all excess MRs are released from the cache, `rel_imm` is reset back to 0.
Bug Fixes	See “Bug Fixes History” on page 46.
2.3-2.0.1	
Bug Fixes	See “Bug Fixes History” on page 46.
2.3-2.0.0	

Table 10 - Change Log History

Category	Description
Connect-IB	Added Suspend to RAM (S3).
Reset Flow	Added Enhanced Error Handling for PCI (EEH), a recovery strategy for I/O errors that occur on the PCI bus.
Register Contiguous Pages	Added the option to ask for a specific address when the register memory is using contiguous page.
mlx5_core	Moved the `mr_cache` subtree from `debugfs` to `mlx5_ib` while preserving all its semantics.
InfiniBand Utilities	Updated the ibutils package. Added to the ibdiagnet tool the "ibdiagnet2.mlnx_cntrs" option to enable reading of Mellanox diagnostic counters.
Bug Fixes	See “Bug Fixes History” on page 46.
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 `CONFIG_MLX4_EN_VXLAN=y` must be enabled.
	MLNX_OFED no longer changes the OS sysctl TCP parameters.
	Added Explicit Congestion Notification (ECN) support
	Added Flow Steering: A0 simplified steering support
	Added RoCE v2 support
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 10 - Change Log History

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
2.2-1.0.1	
Reset Flow	Reset Flow is not activated by default. It is controlled by the mlx4_core`'internal_err_reset'` module parameter.
mlnxofedinstall	32-bit libraries are no longer installed by default on 64-bit OS. To install 32-bit libraries use the `'--with-32bit'` installation parameter.
openibd	Added pre/post start/stop scripts support. For further information, please refer to section *“openibd Script”* in the MLNX-_OFED User Manual.
InfiniBand Core	Asymmetric MSI-X vectors allocation for the SR-IOV hypervisor and guest instead of allocating 4 default MSI-X vectors. The maximum number of MSI-X vectors is `num_cpu` for port ConnectX®-3 has 1024 MSI-X vectors, 28 MSI-X vectors are reserved. • Physical Function - gets the number of MSI-X vectors according to the `pf_msix_table_size` (multiple of 4 - 1) INI parameter • Virtual Functions – the remaining MSI-X vectors are spread equally between all VFs, according to the `num_vfs mlx4_core` module parameter

Table 10 - Change Log History

Category	Description
Ethernet	Ethernet VXLAN support for kernels 3.12.10 or higher
	Power Management Quality of Service: when the traffic is active, the Power Management QoS is enabled by disabling the CPU states for maximum performance.
	Ethernet PTP Hardware Clock support on kernels/OSes that support it
Verbs	Added additional experimental verbs interface. This interface exposes new features which are not integrated yet in to the upstream libibverbs. The Experimental API is an extended API therefor, it is backward compatible, meaning old application are not required to be recompiled to use MLNX-OFED v2.2-1.0.1.
Performance	Out of the box performance improvements: • Use of affinity hints (based on NUMA node of the device) to indicate the IRQ balancer daemon on the optimal IRQ affinity • Improvement in buffers allocation schema (based on the hint above) • Improvement in the adaptive interrupt moderation algorithm
2.1-1.0.6	
IB Core	Added allocation success verification process to ib_alloc_device.
dapl	dapl is recompiled with no FCA support.
openibd	Added the ability to bring up child interfaces even if the parent's ifcfg file is not configured.
libmlx4	Unmapped the hca_clock_page parameter from mlx4_uninit_context.
scsi_transport_srp	scsi_transport_srp cannot be cleared up when rport reconnecting fails.
mlnxofedinstall	Added support for the following parameters: • '--umad-dev-na' • '--without-<package>'
Content Packages Updates	The following packages were updated: • bupc to v2.2-407 • mstflint to v3.5.0-1.1.g76e4acf • perftest to v2.0-0.76.gbf9a463 • hcoll to v2.0.472-1 • Openmpi to v1.6.5-440ad47 • dapl to v2.0.40
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.

Table 10 - Change Log History

Category	Description
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.
2.0-3.0.0	
Operating Systems	Additional OS support: • 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)
2.0-2.0.5	
Virtualization	SR-IOV for both Ethernet and InfiniBand (at Beta level)
Ethernet Network	RoCE over SR-IOV (at Beta level)
	eIPoIB to enable IPoIB in a Para-Virtualized environment (at Alpha level)
	Ethernet Performance Enhancements (NUMA related and others) for 10G and 40G
	Ethernet Time Stamping (at Beta level)
	Flow Steering for Ethernet and InfiniBand. (at Beta level)
	Raw Eth QPs: • Checksum TX/RX • Flow Steering

Table 10 - Change Log History

Category	Description
InfiniBand Network	Contiguous pages: • Internal memory allocation improvements • Register shared memory • Control objects (QPs, CQs)
Installation	YUM update support
VMA	OFED_VMA integration to a single branch
Storage	iSER (at Beta level) and SRP
Operating Systems	Errata Kernel upgrade support
API	VERSION query API: library and headers
Counters	64bit wide counters (port xmit/recv data/packets unicast/mcast)

[a] SSA is tested on SLES 12 only (x86-64 architecture).