



# Mellanox OFED for Linux Release Notes

---

Rev 4.1-1.0.2.0

NOTE:

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



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

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

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

All other trademarks are property of their respective owners.

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

# Table of Contents

<b>Table of Contents</b> .....	<b>3</b>
<b>Chapter 1 Overview</b> .....	<b>5</b>
1.1 Content of Mellanox OFED for Linux. ....	5
1.2 Supported Platforms and Operating Systems .....	6
1.2.1 Supported Non-Linux Virtual Machines .....	9
1.2.2 Tested Hypervisors in Paravirtualized and SR-IOV Environments. ....	9
1.3 Hardware and Software Requirements .....	9
1.4 Supported HCAs Firmware Versions .....	11
1.5 Compatibility Matrix. ....	11
1.6 RDMA CM and RoCE Modes .....	12
1.6.1 RoCE Modes Matrix. ....	12
1.6.2 RDMA CM Default RoCE Mode .....	12
<b>Chapter 2 Changes and New Features in Rev 4.1-1.0.2.0</b> .....	<b>13</b>
2.1 API Changes in MLNX_OFED .....	15
2.2 Unsupported Functionalities/Features/HCAs .....	15
<b>Chapter 3 Known Issues</b> .....	<b>16</b>
<b>Chapter 4 Bug Fixes History</b> .....	<b>43</b>
<b>Chapter 5 Change Log History</b> .....	<b>50</b>

## Release Update History

Release	Date	Description
Rev 4.1-1.0.2.0	August 28, 2017	Added RHEL/CentOS 7.4 and SLES12 SP3 as supported operating systems to <a href="#">Table 1, “Supported Platforms and Operating Systems,”</a> on page 6.
		Updated the version numbers of the products listed in <a href="#">Table 4, “Compatibility Matrix,”</a> on page 11
	July 18, 2017	Added EulerOS 2.0 SP2 as a supported operating system to <a href="#">Table 1, “Supported Platforms and Operating Systems,”</a> on page 6.
	June 29, 2017	Initial release of this version.

# 1 Overview

These are the release notes of MLNX\_OFED for Linux Driver, Rev 4.1-1.0.2.0 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	<ul style="list-style-type: none"> <li>InfiniBand: SDR, QDR, FDR10, FDR</li> <li>Ethernet: 10GigE, 40GigE and 56GigE<sup>a</sup></li> </ul>
ConnectX®-4	mlx5	<ul style="list-style-type: none"> <li>InfiniBand: SDR, QDR, FDR, FDR10, EDR</li> <li>Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, 56GigE<sup>a</sup>, and 100GigE</li> </ul>
ConnectX®-4 Lx		<ul style="list-style-type: none"> <li>Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, and 50GigE</li> </ul>
ConnectX®-5		<ul style="list-style-type: none"> <li>InfiniBand: SDR, QDR, FDR, FDR10, EDR</li> <li>Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, and 100GigE</li> </ul>
ConnectX®-5 Ex		<ul style="list-style-type: none"> <li>InfiniBand: SDR, QDR, FDR, FDR10, EDR</li> <li>Ethernet: 1GigE, 10GigE, 25GigE, 40GigE, 50GigE, and 100GigE</li> </ul>
Connect-IB®		<ul style="list-style-type: none"> <li>InfiniBand: SDR, QDR, FDR10, FDR</li> </ul>

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	<ul style="list-style-type: none"> <li>InfiniBand and Ethernet HCA drivers (mlx4, mlx5)</li> <li>core</li> <li>Upper Layer Protocols: IPoIB, SRP initiator, iSER initiator and target, NVMeoF host and target</li> </ul>
OpenFabrics utilities	<ul style="list-style-type: none"> <li>OpenSM: IB Subnet Manager with Mellanox proprietary Adaptive Routing</li> <li>Diagnostic tools</li> <li>Performance tests</li> <li>SSA (SLES12): libopensmssa plugin for OpenSM, ibssa, ibacm</li> </ul>
MPI	<ul style="list-style-type: none"> <li>Open MPI stack 1.6.5 and later supporting the InfiniBand interface</li> <li>MPI benchmark tests (OSU benchmarks, Intel MPI benchmarks, Presta)</li> </ul>

Components	Description
PGAS	<ul style="list-style-type: none"> <li>HPC-X OpenSHMEM v2.2 supporting InfiniBand, MXM and FCA</li> <li>HPC-X UPC v2.2 supporting InfiniBand, MXM and FCA</li> </ul>
HPC Acceleration packages	<ul style="list-style-type: none"> <li>Mellanox MXM v3.0 (p2p transport library acceleration over Infiniband)</li> <li>Mellanox FCA v3.x (MPI/PGAS collective operations acceleration library over InfiniBand)</li> <li>KNEM, Linux kernel module enabling high-performance intra-node MPI/PGAS communication for large messages</li> </ul>
Extra packages	<ul style="list-style-type: none"> <li>ibutils2</li> <li>ibdump</li> <li>MFT</li> </ul>
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 4.1-1.0.2.0:

**Table 1 - Supported Platforms and Operating Systems**

Operating System	Platform
RHEL6.2/CentOS6.2	x86_64
RHEL6.3/CentOS6.3	x86_64
RHEL6.4/CentOS6.4	x86_64
RHEL6.5/CentOS6.5	x86_64
RHEL6.7/CentOS6.7	x86_64
RHEL6.8/CentOS6.8	x86_64
RHEL6.9/CentOS6.9	x86_64
RHEL7.0/CentOS7.0	x86_64
RHEL7.1/CentOS7.1	x86_64
RHEL7.2/CentOS7.2	x86_64/PPC64 (Power8)/PPC64LE (Power8)
CentOS7.2 with Kernel 4.9 for NVMeoF	ARMv8(AMD) for Softiron
RHEL7.3/CentOS7.3	x86_64/PPC64 (Power8)/PPC64LE (Power8)/ARMv8 (Qualcomm)
RHEL7.4/CentOS7.4	x86_64/PPC64/PPC64LE (Power8)

**Table 1 - Supported Platforms and Operating Systems**

Operating System	Platform
Debian 7.8 + Kernel 4.1.35	x86_64
Debian 7.11	x86_64
Debian 8.7	x86_64
Debian 8.7 Kernel 4.1	x86_64
Debian 8.7 Kernel 4.4	x86_64
Fedora 20	x86_64
Fedora 21	x86_64
Fedora 24	x86_64
OL 6.6	x86_64
OL 6.8	x86_64
OL 7.1	x86_64
SLES11 SP2	x86_64
SLES11 SP3	x86_64
SLES11 SP4	x86_64/PPC64 (Power 8)
SLES12 SP1	x86_64
SLES12 SP2	x86_64/PPC64LE (Power 8)
SLES12 SP3	x86_64/PPC64LE (Power8)
Ubuntu 14.04	x86_64
Ubuntu 16.04 with Kernel 4.9 - Bandera for ARM	ARMv8 (Qualcomm) <b>[beta]</b>
Ubuntu 16.04	x86_64/PPC64LE (Power8)
Ubuntu 16.10	x86_64/PPC64LE (Power 8)
Ubuntu 17.04	x86_64/PPC64LE (Power 8)
Kernels 4.10-4.11	x86_64
WindRiver 6.0	x86_64 SR-IOV + PV
EulerOS 2.0 SP2	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 4.1-1.0.2.0:

- 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	SLES11 SP3 KVM
		SLES11 SP4 KVM
		SLES12 SP2 KVM
		Ubuntu 14.04 KVM
		Ubuntu 16.04 KVM
		Ubuntu 16.10 KVM
		Ubuntu 16.10 KVM-PPC
		RHEL 6.9 KVM
		RHEL 7.3 KVM
	ConnectX-4	SLES11 SP3 KVM
		SLES11 SP4 KVM
		SLES12 SP2 KVM
		Ubuntu 14.04 KVM
		Ubuntu 16.04 KVM
		Ubuntu 16.10 KVM
		RHEL6.9 KVM
		RHEL7.3 KVM
	ConnectX-4 Lx	SLES11 SP4 KVM
		Ubuntu 16.04 KVM
		Ubuntu 14.04 KVM
ConnectX-5	Ubuntu16.10 KVM	
	RHEL6.9 KVM	
	RHEL7.3 KVM	
Paravirtualized	ConnectX-3/ ConnectX-3 Pro	SLES12 SP2 KVM
		Ubuntu 16.10 KVM

### 1.3 Hardware and Software Requirements

The following are the hardware and software requirements of MLNX\_OFED Rev 4.1-1.0.2.0.

- Linux operating system
- Administrator privileges on your machine(s)
- Disk Space: 1GB

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

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

Operating System	Required Packages Installation Command
RHEL/OL/Fedora	<code>yum install perl pciutils python gcc-gfortran libxml2-python tcsh libnl.i686 libnl expat glib2 tcl libstdc++ bc tk gtk2 atk cairo numactl pkgconfig ethtool</code>
SLES 11 SP3	<code>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</code>
SLES 12	<code>zypper install pkg-config expat libstdc++6 libglib-2_0-0 lib-gtk-2_0-0 tcl libcairo2 tcsh python bc pciutils libatk-1_0-0 tk python-libxml2 lsof libnl3-200 ethtool</code>
Ubuntu/Debian	<code>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</code>
Debian 8	<code>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</code>

## 1.4 Supported HCAs Firmware Versions

MLNX\_OFED Rev 4.1-1.0.2.0 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	2.40.7000	2.36.5150
ConnectX®-3 Pro	2.40.7000	2.36.5150
ConnectX®-4	12.20.1010	12.18.2000
ConnectX®-4 Lx	14.20.1010	14.18.2000
ConnectX®-5	16.20.1010	16.19.1200
ConnectX®-5 Ex	16.20.1010	16.19.1200
Connect-IB®	10.16.1020	N/A

For the official firmware versions, please see:

[http://www.mellanox.com/content/pages.php?pg=firmware\\_download](http://www.mellanox.com/content/pages.php?pg=firmware_download)

## 1.5 Compatibility Matrix

MLNX\_OFED Rev 4.1-1.0.2.0 is compatible with the following:

**Table 4 - Compatibility Matrix**

Mellanox Product	Description/Version
MLNX-OS®	MSX6036 w/w MLNX-OS® version 3.6.4006 <sup>a</sup>
Grid Director™	4036 w/w Grid Director™ version 3.9.1-985
Unified Fabric Manager (UFM®)	v5.9
MXM	v3.6.3103
HCOLL <sup>b</sup>	v3.8.1652
OpenMPI	v2.1.2a

- MLNX\_OFED Rev 4.1-1.0.2.0 was tested with this switch. However, additional switches may be supported as well.
- 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 <sup>a</sup>	ConnectX-3 Pro	ConnectX-4/ ConnectX-4 Lx/ ConnectX-5/ ConnectX-5 Ex <sup>a</sup>	ConnectX-3 Pro/ConnectX-4/ ConnectX-4 Lx/ConnectX-5/ ConnectX-5 Ex <sup>a</sup>
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 4.1-1.0.2.0

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 4.1-1.0.2.0**

HCA's	Feature/Change	Description
mlx5 Driver	RoCE Diagnostics and ECN Counters	Added support for additional RoCE diagnostics and ECN congestion counters under <code>/sys/class/infiniband/mlx5_0/ports/1/hw_counters/</code> directory. For further information, refer to the <a href="#">Understanding mlx5 Linux Counters and Status Parameters</a> 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 <a href="#">Understanding ethtool rx-fcs for mlx5 Drivers</a> 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 <a href="#">HowTo Configure Trust Mode on Mellanox Adapters</a> Community post.
	RoCE ECN Parameters	ECN parameters have been moved to the following directory: <code>/sys/kernel/debug/mlx5/&lt;PCI BUS&gt;/cc_params/</code> For more information, refer to the <a href="#">HowTo Configure DCQCN (RoCE CC) for ConnectX-4 (Linux)</a> Community post.
	Flow Steering Dump Tool	Added support for <code>mlx_fs_dump</code> , 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 <a href="#">PeerDirect Programming</a> Community post.

**Table 6 - Changes and New Features in Rev 4.1-1.0.2.0**

HCA's	Feature/Change	Description
	Probed VFs	Added the ability to disable probed VFs on the hypervisor. For further information, see <a href="#">HowTo Configure and Probe VFs on mlx5 Drivers</a> Community post.
	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 <b>alpha</b> 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 <a href="#">HowTo Test 1PPS on Mellanox Adapters</a> 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 <a href="#">Running Linux PTP with ConnectX-4</a> Community post.
	Fast Driver Unload	Added support for fast driver teardown in shutdown and kexec flows.
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 <a href="#">HowTo Configure NVMeoF Target Offload</a> .
	MPI Tag Matching	Added support for offloading MPI tag matching to HCA.
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 <a href="#">RDMA CM and RoCE Version Defaults</a> Community post.
	Lustre	Added support for Lustre file system open-source project.
	Bug Fixes	See <a href="#">Section 4, "Bug Fixes History", on page 43</a> .

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
		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](http://www.mellanox.com/pdf/prod_software/MLNX_OFED_Archived_Known_Issues.pdf)

**Table 8 - Known Issues**

Internal Reference Number	Issue
965591	<b>Description:</b> Lustre support is limited to versions 2.9 and 2.10.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Lustre
	<b>Discovered in Release:</b> 4.1-1.0.2.0
995665	<b>Description:</b> Connection between NVMeoF host and target cannot be established in a hyper-threaded system with more than 64 CPUs on the NVMeoF host side.
	<b>Workaround:</b> On the host side, connect to NVMeoF subsystem using <code>--nr-io-queues &lt;num_queues&gt;</code> flag. Note that <code>num_queues</code> must be lower or equal to <code>num_sockets</code> multiplied with <code>num_cores_per_socket</code> .
	<b>Keywords:</b> NVMeoF
1039346	<b>Description:</b> Enabling multiple namespaces per subsystem while using NVMeoF target offload is not supported.
	<b>Workaround:</b> To enable more than one namespace, create a subsystem for each one.
	<b>Keywords:</b> NVMeoF Target Offload, namespace
1072347	<b>Description:</b> <code>Ethtool -i &lt;ibx&gt;</code> displays incorrect driver name for devices with enhanced IPoIB support.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Enhanced IPoIB, Ethtool



**Table 8 - Known Issues**

Internal Reference Number	Issue
1071457	<p><b>Description:</b> PKEY-related limitations in enhanced IPoIB:</p> <ul style="list-style-type: none"> <li>• Since the parent interface ib&lt;x&gt; and the child interface ib&lt;x&gt;.yyyy share the same receive resources, the parent interface's MTU cannot be less than the child interface's MTU</li> <li>• Interface counters and Ethtool control are not supported on child interfaces</li> <li>• Parent interface should be in UP state to enable child interface to receive traffic</li> </ul>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> PKEY, Enhanced IPoIB, MTU, Ethtool, Interface Counters</p>
1059451	<p><b>Description:</b> When Enhanced IPoIB is enabled, the following module parameters will not be functional:</p> <ul style="list-style-type: none"> <li>• send_queue_size</li> <li>• rcv_queue_size</li> <li>• max_nonsrq_conn_qp</li> </ul>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> Enhance IPoIB</p>
1030301	<p><b>Description:</b> Creating virtual functions on a device that is in LAG mode will destroy the LAG configuration. The bonding device over the Ethernet NICs will continue to work as expected.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> LAG, SR-IOV</p>
1047616	<p><b>Description:</b> When node GUID of a device is set to zero (0000:0000:0000:0000), RDMA_CM user space application may crash.</p>
	<p><b>Workaround:</b> Set node GUID to a nonzero value.</p>
	<p><b>Keywords:</b> RDMA_CM</p>
1061298	<p><b>Description:</b> Since enhanced IPoIB does not support connected mode on RedHat operating systems, when using network manager 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.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> Enhanced IPoIB</p>
1068215	<p><b>Description:</b> When enhanced IPoIB mode is enabled, ring size limit is 8k. When it is disabled, ring size limit is decreased to 4k.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> Enhanced IPoIB</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
1051701	<b>Description:</b> New versions of iproute which support new kernel features may misbehave on old kernels that do not support these new features.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> iproute
1006032	<b>Description:</b> Currently, iSER initiator supports T10-PI offload mechanism only in the following OSs: <ul style="list-style-type: none"> <li>• Ubuntu 16.10, 16.04</li> <li>• RedHat 7.3, 7.2</li> <li>• SLES 12.02</li> </ul>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> iSER initiator
1007830	<b>Description:</b> When working on Xenserver hypervisor with SR-IOV enabled on it, make sure the following instructions are applied: <ol style="list-style-type: none"> <li>1. Right after enabling SR-IOV, unbind all driver instances of the virtual functions from their PCI slots.</li> <li>2. It is not allowed to unbind PF driver instance while having active VFs.</li> </ol>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
1008583	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> RDMA-CM
1007356	<b>Description:</b> Creating a PKEY interface using “ip link” is not supported.
	<b>Workaround:</b> Use sysfs to create a PKEY interface.
	<b>Keywords:</b> IPoIB, PKEY
1000197	<b>Description:</b> Displaying multicast groups using sysfs may not show all the entries on Fedora 23 OS.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
1010148	<b>Description:</b> Upgrading from MLNX_OFED v3.x to v4.x using yum and apt-get repositories fails.
	<b>Workaround:</b> Remove MLNX_OFED v3.x using the ofed_uninstall.sh script, and only then install MLNX_OFED v4.x as usual.
	<b>Keywords:</b> Installation

**Table 8 - Known Issues**

Internal Reference Number	Issue
1005786	<p><b>Description:</b> When using ConnectX-5 adapter cards, the following error might be printed to dmesg, indicating temporary lack of DMA pages:  <code>"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"</code></p> <p><b>Example:</b> This might happen when trying to open more than 64 VFs per port.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> mlx5_core, DMA</p>
1008066	<p><b>Description:</b> 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 <code>get_vf_stats</code> (via <code>iptool</code>) during reboot.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> mlx5_core, reboot</p>
1009488	<p><b>Description:</b> 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 <code>"/media/CDROM(vcd)"/</code>, installation will fail and the following error message will be displayed:  <pre># cd /media/CDROM(vcd\)/ # ./mlnxofedinstall sh: 1: Syntax error: "(" unexpected</pre></p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Installation</p>
982144	<p><b>Description:</b> When offload traffic sniffer is on, the bandwidth could decrease up to 50%.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Offload Traffic Sniffer</p>
981045	<p><b>Description:</b> On kernels below v4.2, when removing a bonding module with devices different from <code>ARPHRD_ETHER</code>, a call trace may be received.</p> <p><b>Workaround:</b> Remove the bond in the following order:  Remove the slaves, delete the bond, and only then remove the bonding module.</p> <p><b>Keywords:</b> Bonding</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
980066	<b>Description:</b> Soft RoCE does not support Extended Reliable Connection (XRC).
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Soft RoCE, XRC
982534	<b>Description:</b> 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: <ol style="list-style-type: none"> <li>1. mlx4_core driver does not load.</li> <li>2. The mlx4_core driver does load, but calls to <code>ibv_open_device</code> may return ENOMEM errors.</li> </ol>
	<b>Workaround:</b> <ol style="list-style-type: none"> <li>1. Add the following parameter in the firmware's ini file under [HCA] section: <code>log2_uar_bar_megabytes = 7</code></li> <li>2. Re-burn the firmware with the new ini file.</li> </ol>
	<b>Keywords:</b> PPC
981362	<b>Description:</b> On several OSs, setting a number of TC is not supported via the <code>tc</code> tool.
	<b>Workaround:</b> Set the number of TC via the <code>/sys/class/net/&lt;interface&gt;/qos/tc_num</code> sysfs file.
	<b>Keywords:</b> Ethernet, TC
980257	<b>Description:</b> 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.
	<b>Workaround:</b> Delete the bond before restarting the driver.
	<b>Keywords:</b> Bonding, IPoIB
980034	<b>Description:</b> Soft RoCE counters located under <code>/sys/class/infiniband/&lt;rx-inf&gt;/ports/1/counters/</code> directory are not supported.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Soft RoCE
979907	<b>Description:</b> Only the following two experimental verbs are supported for Soft RoCE: <ul style="list-style-type: none"> <li>• <code>ibv_exp_query_device</code></li> <li>• <code>ibv_exp_poll_cq</code>.</li> </ul>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Soft RoCE

**Table 8 - Known Issues**

Internal Reference Number	Issue
979457	<b>Description:</b> When setting IOMMU=ON, a severe performance degradation may occur due to a bug in IOMMU.
	<b>Workaround:</b> Make sure the following patches are found in your kernel: <ul style="list-style-type: none"> <li>• iommu/vt-d: Fix PASID table allocation</li> <li>• iommu/vt-d: Fix IOMMU lookup for SR-IOV Virtual Functions</li> </ul> <b>Note:</b> These patches are already available in Ubuntu 16.04.02 and 17.04 OSs.
	<b>Keywords:</b> Performance, IOMMU
977852	<b>Description:</b> <code>rdma_cm</code> running over IB ports does not support UD QPs on ConnectX-3 adapter cards.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV, RDMA CM
955113	<b>Description:</b> In RoCE LAG over ConnectX-4 adapter cards, the script <code>ibdev2netdev</code> may show a wrong port state for the bonded device. This means that although the IB device/port <code>mlx5_bond_0/1</code> is up (as seen in <code>ibstat</code> ), <code>ibdev2netdev</code> may report that it is down.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> RoCE, LAG, bonding
942161	<b>Description:</b> On some kernels, there might be an issue in csum calculations of tunneled packets when the driver sets <code>CHECKSUM_COMPLETE</code> for the packet. This might print csum error messages to the <code>dmesg</code> log file.
	<b>Workaround:</b> Make sure your kernel version includes <a href="#">this fix</a> .
	<b>Keywords:</b> Ethernet, checksum, tunneling
931574	<b>Description:</b> When using a kernel with Generic Receive Offload (GRO) support, UDP performance results will reveal degradation in comparison to the UDP performance results in <code>MLNX_OFED v1.5.x</code> .
	<b>Workaround:</b> Turn off the GRO feature to get better UDP performance. Run: <code>#ethtool -K &lt;interface&gt; gro off</code>
	<b>Keywords:</b> GRO, UDP, performance
920707	<b>Description:</b> In SLES12 SP2, you may get a memory low warning at the netlink layer when configuring a large number of VFs.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV, SLES12

**Table 8 - Known Issues**

Internal Reference Number	Issue
969467	<p><b>Description:</b> 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"</p> <p><b>Workaround:</b> Use the following command to remove the kernel-mft packages:  <code>rpm -e --noscripts \$(rpm -qa   grep kernel-mft-mlnx)</code></p> <p><b>Keywords:</b> Installation</p>
918880	<p><b>Description:</b> The driver version shown in modinfo and ethtool outputs is 3.4-1.0.6 instead of 3.4-2.0.0.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Installation</p>
679801	<p><b>Description:</b> Updating MLNX_OFED via Yum (e.g. running "yum update mlnx-ofed-all") can fail with the following error:            --&gt; 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)</p> <p><b>Workaround:</b> Remove the mpitests packages manually:  <code># rpm -e --allmatches \$(rpm -qa   grep mpitests_)</code></p> <p><b>Keywords:</b> Installation</p>
690799	<p><b>Description:</b> OpenSM package removal fails with the following error on Ubuntu12.04:            Removing opensm ...            /sbin/insserv: No such file or directory</p> <p><b>Workaround:</b></p> <ol style="list-style-type: none"> <li>1. Create the missing link by running this command:  <code># ln -s /usr/lib/insserv/insserv /sbin/insserv</code></li> <li>2. Remove the package.</li> </ol> <p><b>Keywords:</b> Installation</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
764204	<p><b>Description:</b> 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.</p> <p><b>Workaround:</b></p> <ul style="list-style-type: none"> <li>As of MLNX_OFED v3.3, use the <code>mlnx_add_kernel_support.sh</code> script, or simply provide the <code>--add-kernel-support</code> flag to <code>mlnxofedinstall</code> script.</li> <li>Update the <code>kmod</code> package using the following link: <a href="https://rhn.redhat.com/errata/RHBA-2016-1832.html">https://rhn.redhat.com/errata/RHBA-2016-1832.html</a></li> </ul> <p><b>Keywords:</b> Installation</p>
785119	<p><b>Description:</b> When upgrading ConnectX-4/ConnectX-4 Lx firmware version from v12/14.14.2036 to a newer one (for example:12/14.16.1xxx), power cycle is necessary to enable working in Pass-Through mode. Using <code>mlxfwreset</code> instead of power cycle will print messages similar to the following when Passing-Through the device to Virtual Machine:  <pre>"-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=hostdev0,bus=pci.0,addr=0x7: Device initialization failed."</pre></p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Installation</p>
773774	<p><b>Description:</b> When downgrading from MLNX_OFED 3.3-x.x.x, driver reload might fail with the following error in <code>dmeg</code>:  <pre>rmmmod: ERROR: Module mlx_compat is in use by: ib_netlink</pre></p> <p><b>Workaround:</b> The issues will be resolved automatically after system reboot or by invoking the following commands:  <pre>rmmmod ib_netlink depmod -a /etc/init.d/openibd restart</pre></p> <p><b>Keywords:</b> Driver Start</p>
677998	<p><b>Description:</b> False alarm errors may be printed to <code>dmesg</code>.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Driver Start</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
610395	<b>Description:</b> 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.
	<b>Workaround:</b> Use the <code>mlnx_add_kernel_support.sh</code> script to compile MLNX_OFED drivers against the new kernel.
	<b>Keywords:</b> Driver Start
967356	<b>Description:</b> [Ethernet] <ul style="list-style-type: none"> <li>Bare-metal ConnectX-4/ConnectX-4 Lx might suffer up to 15, degradation in some scenarios due to higher CPU utilization.</li> <li>PPC: ConnectX-4 might suffer up to 20, degradation in some scenarios.</li> </ul>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Performance
956071	<b>Description:</b> [mlx5] OOB TCP performance for small message sizes may suffer from lower BW than expected.
	<b>Workaround:</b> Disable adaptive-rx and set higher static moderation: <pre>ethtool -C &lt;interface&gt; adaptive-rx off rx-frames 128 rx-usecs 128</pre>
	<b>Keywords:</b> Performance
765777	<b>Description:</b> Low VxLAN throughput due to broken GRO offload in most kernels older than kernel v4.6.
	<b>Workaround:</b> Use kernel version 4.6 or above.
	<b>Keywords:</b> Performance
414827	<b>Description:</b> Out-of-the-box throughput performance in Ubuntu14.04 is not optimal and may achieve results below the line rate in 40GE link speed.
	<b>Workaround:</b> For additional performance tuning, please refer to Performance Tuning Guide.
	<b>Keywords:</b> Performance
417751	<b>Description:</b> Performance degradation might occur when bonding Ethernet interfaces.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Performance
656415	<b>Description:</b> In RHEL7.0, when the irqbalance service is started or restarted, it incorrectly re-balances the IRQs, including the banned ones.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Performance



**Table 8 - Known Issues**

Internal Reference Number	Issue
651322	<p><b>Description:</b> 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: <a href="http://comments.gmane.org/gmane.linux.network/344243">http://comments.gmane.org/gmane.linux.network/344243</a></p>
	<p><b>Workaround:</b> Use RH7.2 to avoid such performance issues.</p>
	<p><b>Keywords:</b> Performance</p>
754646	<p><b>Description:</b> The default RX coalescing values yield to high CPU utilization when using VXLAN on VMs over PV.</p>
	<p><b>Workaround:</b> Increase the RX microseconds and frames coalescing parameters for a better utilization using the ethtool -C command.</p>
	<p><b>Keywords:</b> Performance</p>
783496	<p><b>Description:</b> When using a VF over RH7.X KVM, low throughput is expected.</p>
	<p><b>Workaround:</b> Install the following packages using the link below:</p> <ul style="list-style-type: none"> <li>• qemu-img-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm</li> <li>• qemu-kvm-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm</li> <li>• qemu-kvm-common-1.5.3-105.el7_2.1.bz1299846.0.x86_64.rpm</li> </ul> <p><a href="http://people.redhat.com/~alwillia/bz1299846/">http://people.redhat.com/~alwillia/bz1299846/</a></p>
	<p><b>Keywords:</b> Performance</p>
780782	<p><b>Description:</b> CALC operation on PowerPC may report completion with error.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> mlx5 Driver</p>
921252	<p><b>Description:</b> 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.</p>
	<p><b>Workaround:</b> Lower the number of Work Requests during the QP creation.</p>
	<p><b>Keywords:</b> mlx5 Driver</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
860311	<p><b>Description:</b> An allocation of high-order page in <code>mlx5e_alloc_striding_rx_wqe</code> fails with a call-trace.</p> <p><b>Workaround:</b> No action is required on users end. A fragmented fallback flow will handle this failure.</p> <p><b>Keywords:</b> mlx5 Driver</p>
435583	<p><b>Description:</b> EEH events that arrive while the mlx5 driver is loading may cause the driver to hang.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> mlx5 Driver</p>
434570	<p><b>Description:</b> The mlx5 driver can handle up to 5 EEH events per hour.</p> <p><b>Workaround:</b> If more events are received, cold reboot the machine.</p> <p><b>Keywords:</b> mlx5 Driver</p>
554120	<p><b>Description:</b> When working with Connect-IB firmware v10.10.5054, the following message would appear in driver start.  <code>command failed, status bad system state(0x4), syndrome 0x408b33</code>            The message can be safely ignored.</p> <p><b>Workaround:</b> Upgrade Connect-IB firmware to the latest available version.</p> <p><b>Keywords:</b> mlx5 Driver</p>
538843	<p><b>Description:</b> Bonding active-backup mode does not function properly.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> mlx5 Driver</p>
-	<p><b>Description:</b> Rate, speed and width using IB sysfs/tools are available in RoCE mode in ConnectX-4 only after port physical speed configuration is done.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> mlx5 Driver</p>
598092	<p><b>Description:</b> Since MLNX_OFED's <code>openibd</code> does not unload modules while OpenSM is running, removing <code>mlx5_core</code> manually while OpenSM is running, may cause it to be out of sync when probed again.</p> <p><b>Workaround:</b> Restart OpenSM</p> <p><b>Keywords:</b> mlx5 Driver</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
563022	<b>Description:</b> ConnectX-4 port GIDs table shows a duplicated RoCE v2 default GID.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> mlx5 Driver
947542	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
964991	<b>Description:</b> TX queue rate limit may sometimes exceed the rate that was set by the user by up to 10,.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
911693	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
948312	<b>Description:</b> [ConnectX-3 Pro] To enable/disable <code>rx-vlan-stag-hw-parse</code> by <code>ethtool</code> , <code>rxvlan</code> should be enabled/disabled accordingly ( <code>ethtool -K rxvlan on/off</code> ).
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
894547	<b>Description:</b> On SLES12 SP1 and SLES12 SP2, invalid udev rules might cause Ethernet interfaces renaming to fail, leaving some interfaces with names such as <code>renameXY</code> .
	<b>Workaround:</b> Modify the udev rules inside the <code>/etc/udev/rules.d/70-persistent-net.rules</code> 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.
	<b>Keywords:</b> Ethernet

**Table 8 - Known Issues**

Internal Reference Number	Issue
754709	<p><b>Description:</b> mlx5 Ethernet auto-negotiation related issues:</p> <ol style="list-style-type: none"> <li>1. The command <code>ethtool -s eth4 speed 25000 autoneg on</code> is not a valid <code>ethtool</code> command. <code>speed 25000</code> should not be passed in when <code>autoneg</code> is on. Instead, use <code>advertise 0x100000</code>.</li> <li>2. <code>ethtool</code> version older than v4.6 does not report neither support nor advertise for new speeds, such as 25G, 100G.</li> <li>3. When setting auto negotiation with <code>ethtool</code> version older than v4.6, advertised speed will be ignored, and the device will try to reach the highest supported speed available end-to-end.</li> </ol> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Ethernet</p>
843306	<p><b>Description:</b> [ConnectX-4/ConnectX-4 Lx] When configuring ETS, bandwidth values are limited between 1-100, and 0 is an invalid value.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Ethernet</p>
704750	<p><b>Description:</b> [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.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Ethernet</p>
433366	<p><b>Description:</b> Reboot might hang in SR-IOV when using the <code>probe_vf</code> parameter with many Virtual Functions. The following message is logged in the kernel log:</p> <pre>"waiting for eth to become free. Usage count =1"</pre> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> Ethernet</p>
539117	<p><b>Description:</b> On SLES12, the bonding interface over Mellanox Ethernet slave interfaces does not get IP address after reboot.</p> <p><b>Workaround:</b></p> <ol style="list-style-type: none"> <li>1. Set "STARTMODE=hotplug" in the bonding slave's <code>ifcfg</code> files. More details can be found in the SUSE documentations page: <a href="https://www.suse.com/documentation/sles-12/book_sle_admin/?page=/documentation/sles-12/book_sle_admin/data/sec_bond.html">https://www.suse.com/documentation/sles-12/book_sle_admin/?page=/documentation/sles-12/book_sle_admin/data/sec_bond.html</a></li> <li>2. Enable the nanny service to support hot-plugging: Open the <code>/etc/wicked/common.xml</code> file. Change: <code>&lt;use-nanny&gt;false&lt;/use-nanny&gt;</code> to <code>&lt;use-nanny&gt;true&lt;/use-nanny&gt;</code></li> <li>3. Run: <code># systemctl restart wickedd.service wicked</code></li> </ol> <p><b>Keywords:</b> Ethernet</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
989042	<b>Description:</b> <code>ethtool -x</code> command will not function on relatively old kernels that do not support <code>get/set_rxfh*</code> callbacks.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
516136	<b>Description:</b> Ethertype proto 0x806 not supported by <code>ethtool</code>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
592229	<b>Description:</b> When NC-SI is ON, the ports MTU cannot be set to lower than 1500.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
600242	<b>Description:</b> GRO is not functional when using VXLAN in ConnectX-3 adapter cards.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
596075	<b>Description:</b> <code>ethtool -X</code> : The driver supports only the 'equal' mode and cannot be set by using weight flags.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
600752	<b>Description:</b> Q-in-Q infrastructure in the kernel is supported only in kernel version 3.10 and up.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
596537	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
560575	<b>Description:</b> When using a hardware that has Time Stamping enabled, the system time might be higher than the expected variance.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet

**Table 8 - Known Issues**

Internal Reference Number	Issue
597758	<b>Description:</b> In Q-in-Q, ping failed when sending traffic with package size > 1468
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
665131	<b>Description:</b> Call trace may occur when configuring VXLAN or under high traffic stress.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
685069/ 689607	<b>Description:</b> ethtool header does not currently support the link speeds of 25/50/100. Therefore, these speeds cannot be seen as advertised/supported.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet
835239	<b>Description:</b> While running Q-in-Q packets with stag offloading, tcpdump/wireshark on host may show svlan ethertype as 0x8100 instead of 0x88A8.
	<b>Workaround:</b> Check the wire or a switch between the hosts, the wireshark will show 0x88A8 ethertype as expected.
	<b>Keywords:</b> Ethernet
954924	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Flow Steering
516136	<b>Description:</b> Setting ARP flow rules through ethtool is not allowed.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Flow Steering
448981	<b>Description:</b> QoS default settings are not returned after configuring QoS.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Quality of Service
940345	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Ethernet Performance Counters

**Table 8 - Known Issues**

Internal Reference Number	Issue
891241	<b>Description:</b> Sysfs for displaying neighbor information is not supported.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
920440	<b>Description:</b> Adaptive RX moderation in not supported.
	<b>Workaround:</b> 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: <code>on ib0: ethtool -C ib0 rx-usecs 16 rx-frames 88</code>
	<b>Keywords:</b> IPoIB
965910	<b>Description:</b> On RHEL7.3, when creating a PKEY using the ifcfg file with ConnectX-3 and ConnectX-3 Pro adapter cards, <code>ifdown</code> and <code>ifup</code> commands must be run respectively.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
854235	<b>Description:</b> IPoIB bonding interface has to be restarted in order to work on some operating systems.
	<b>Workaround:</b> Toggle bonding interface to state down and then to state up.
	<b>Keywords:</b> IPoIB
552840	<b>Description:</b> <code>ifdown</code> command does not function in RH7.x
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
665143	<b>Description:</b> Kernel Oops may occur after reboot.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
555632	<b>Description:</b> Kernel panic may occur while re-assigning LIDs.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
556352	<b>Description:</b> ICMP traffic might be lost after Vnic restart
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB

**Table 8 - Known Issues**

Internal Reference Number	Issue
560575	<b>Description:</b> Spikes may occur while running PTP protocol over ConnectX-3/ConnectX-3 Pro.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
684720	<b>Description:</b> <code>ifdown</code> fails on SLES12SP0/SP1 with the following errors <pre># ifdown ib0 wicked: ifdown: no matching interfaces</pre> The error indicates that there are active interfaces using the interface you are trying to bring down, and you must <code>ifdown</code> all dependent interfaces.
	<b>Workaround:</b> To see the list of all dependent interfaces, run: <pre># wicked --debug all ifdown ib0 .. .. wicked: skipping ib0 interface: unable to ifdown due to lower- dev dependency to: ib0.8001 wicked: ifdown: no matching interfaces wicked: Exit with status: 0</pre>
	<b>Keywords:</b> IPoIB
766451	<b>Description:</b> Occasionally, in kernel 3.10, under heavy load, the kernel fails to get free page. For more details, please refer to: <a href="https://bugs.centos.org/view.php?id=10245">https://bugs.centos.org/view.php?id=10245</a>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> IPoIB
383034	<b>Description:</b> On rare occasions, upon driver restart the following message is shown in the <code>dmesg</code> : <pre>'cannot create duplicate filename '/class/net/eth_ipoib_inter- faces'</pre>
	<b>Workaround:</b> N/A
	<b>Keywords:</b> eIPoIB
855362	<b>Description:</b> A compilation error will occur in kernel space application when setting the <code>wr_id</code> field upon initializing any of the following structures: <code>ib_wc</code> , <code>ib_send_wr</code> , or <code>ib_recv_wr</code> . This is caused due to <code>wr_id</code> insertion into an anonymous union.
	<b>Workaround:</b> Assign the enum field explicitly. For example: <code>wr.wr_id = MY_WR_ID;</code>
	<b>Keywords:</b> Verbs



**Table 8 - Known Issues**

Internal Reference Number	Issue
835061	<p><b>Description:</b> 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.</p>
	<p><b>Workaround:</b> Run QP command query to verify the value.</p>
	<p><b>Keywords:</b> Verbs</p>
935250	<p><b>Description:</b> 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).</p>
	<p><b>Workaround:</b></p> <ol style="list-style-type: none"> <li>1. Verify the NUM_OF_VFS value by running: <code>mlxconfig -d /dev/mst/mt4115_pci-conf0 query  grep NUM_OF_VFS</code> Output: NUM_OF_VFS 32</li> <li>2. Change NUM_OF_VFS value by running: <code>mlxconfig -d /dev/mst/mt4115_pci-conf0 set NUM_OF_VFS=0</code></li> <li>3. Reboot the machine.</li> </ol>
	<p><b>Keywords:</b> RoCE</p>
935250	<p><b>Description:</b> RoCE LAG for mlx5 is supported in kernel v4.5 and above.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> RoCE</p>
869158	<p><b>Description:</b> Occasionally, UC UD traffic over default GIDs, with high iterations may get stuck.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> RoCE</p>
854517	<p><b>Description:</b> Driver restart while having RDMA-CM running applications may hang.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> RoCE</p>
866072	<p><b>Description:</b> [mlx5] RoCE v2 multicast traffic using RDMA-CM with IPv4 address will not be received.</p>
	<p><b>Workaround:</b> N/A</p>
	<p><b>Keywords:</b> RoCE</p>

**Table 8 - Known Issues**

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

**Table 8 - Known Issues**

Internal Reference Number	Issue
959452	<b>Description:</b> When adding an RXE device to one HCA port, <code>ibdev2netdev</code> will show that the RXE device has been added to both HCA ports.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Soft RoCE
963537	<b>Description:</b> 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 <code>librxe</code> and the HCA library (for example: <code>libmlx5</code> ).
	<b>Workaround:</b> Perform one of the following to avoid the issue: <ul style="list-style-type: none"> <li>• Uninstall the actual library (For example: <code>uninstall libmlx5</code>)</li> <li>• Remove the library's configuration file (Run: <code>rm -f /etc/libib-verbs.d/mlx5.driver</code>)</li> </ul>
	<b>Keywords:</b> Soft RoCE
664110	<b>Description:</b> SDP is currently not supported in <code>mlx5</code> driver (Connect-IB and Connect-X 4 adapter cards)
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SDP
683370	<b>Description:</b> 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.
	<b>Workaround:</b> Set module param <code>always_register=N</code> <pre>\$ modprobe ib_iser always_register=N</pre>
	<b>Keywords:</b> iSER Initiator
960642	<b>Description:</b> [ <code>mlx5</code> ] <code>min_tx_rate</code> and <code>max_tx_rate</code> limit per virtual function is not supported on ConnectX-5 and ConnectX-5 Ex adapter cards.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
896859	<b>Description:</b> ConnectX-3 virtual function that runs on a <code>MLNX_OFED</code> version that is older than v4.0 cannot communicate with ConnectX-3 virtual function that runs on <code>MLNX_OFED</code> v4.0.
	<b>Workaround:</b> Set the OpenSM default alias-guid hop limit to 2 in OpenSM configuration file: <pre>aguid_default_hop_limit 2</pre>
	<b>Keywords:</b> SR-IOV

**Table 8 - Known Issues**

Internal Reference Number	Issue
858628	<b>Description:</b> PCI error handling is not supported during driver reload. This might cause a kernel panic or calltrace.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
860385	<b>Description:</b> Creating 127 VFs may cause kernel panic in SLES11 SP4 KVM with Kernel 3.0.101-63 because of a IOMMU kernel bug.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
795697	<b>Description:</b> [mlx4] While spoof-check filters the incoming traffic to a VM, when this feature is disabled, traffic still does not reach the VM.
	<b>Workaround:</b> The driver must be restarted for the disablement of the feature to take effect and all traffic to be reached to the VM.
	<b>Keywords:</b> SR-IOV
835065	<b>Description:</b> [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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
784940	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
784954	<b>Description:</b> When SR-IOV is disabled, the VF driver receives <code>pci_err_detected</code> event and a teardown flow will be started. During the teardown flow, all firmware commands will fail because the function is already deleted.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV

**Table 8 - Known Issues**

Internal Reference Number	Issue
819595	<b>Description:</b> [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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
775944	<b>Description:</b> Bonding VFs on the same physical port using bonding mode 0 requires configuration of <code>fail_over_mac=1</code> .
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
381764	<b>Description:</b> <code>mlx4_port1_mtu</code> sysfs entry shows a wrong MTU number in the VM.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
426988	<b>Description:</b> When at least one port is configured as InfiniBand, and the <code>num_vfs</code> is provided but the <code>probe_vf</code> is not, HCA initialization fails.
	<b>Workaround:</b> Use both the <code>num_vfs</code> and the <code>probe_vf</code> in the <code>modprobe</code> line.
	<b>Keywords:</b> SR-IOV
385750/ 378528	<b>Description:</b> 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.
	<b>Workaround:</b> Unload the module <code>mlx4_ib</code> and reload it in the VM.
	<b>Keywords:</b> SR-IOV
392172	<b>Description:</b> 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
	<b>Workaround:</b> Shut down the driver in the VM before detaching the VF.
	<b>Keywords:</b> SR-IOV
506512	<b>Description:</b> Setting 1 Mbit/s rate limit on Virtual Functions (Qos Per VF feature) may cause TX queue transmit timeout.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV

**Table 8 - Known Issues**

Internal Reference Number	Issue
567908	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
601749	<b>Description:</b> 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.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
649366	<b>Description:</b> Restarting the PF (Hypervisor) driver while Virtual Functions are assigned is not allowed in RH7 and above due to a <code>vfio-pci</code> bug.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
639046	<b>Description:</b> Due to an issue with SR-IOV loopback, prevention "Duplicate IPv6 detected" are seen in the VF driver.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
655410	<b>Description:</b> [ConnectX-4/Connect-IB] Failed to enable SR-IOV due to errors in PCI or BIOS.
	<b>Workaround:</b> 1. Add <code>pci=realloc=on</code> to the grub command line. 2. Add more memory to the server. 3. Upgrade BIOS version.
	<b>Keywords:</b> SR-IOV
651119	<b>Description:</b> Kernel panic may occur while running IPv6 UDP on SR-IOV ConnectX-4 environment
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
669910	<b>Description:</b> Bind/Unbind over ConnectX-4 Hypervisor may cause system lockup.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV

**Table 8 - Known Issues**

Internal Reference Number	Issue
650458	<b>Description:</b> Occasionally, IPv6 might not function properly and cause lockup on SR-IOV ConnectX-4 environment.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
688551	<b>Description:</b> In ConnectX-3 adapter cards, the extended counter <code>port_rcv_data_64</code> on the VF may not be updated in some flows.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
690656/ 690674	<b>Description:</b> When the physical link is down, any traffic from the PF to any VF on the same port will be dropped.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
691661	<b>Description:</b> 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.
	<b>Workaround:</b> Probe one of the VFs in the hypervisor and use for RoCE.
	<b>Keywords:</b> SR-IOV
691661	<b>Description:</b> Ethernet SR-IOV in ConnectX-4 requires firmware version 12.14.1100 and higher
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
737434	<b>Description:</b> VF vport statistics are not cleared upon <code>ifconfig</code> up/down.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
738464	<b>Description:</b> In SLES11 SP4, user cannot open all VFs announced in <code>sriov_v_totalvfs</code> . However he can set the <code>num_vfs</code> up to maximum <code>sriov_totalvfs-1</code> vfs.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
784127	<b>Description:</b> While disabling SR-IOV, all firmware teardown flow commands are expected to fail and error messages will be reported in the <code>dmesg</code> .
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV

**Table 8 - Known Issues**

Internal Reference Number	Issue
784146	<b>Description:</b> Creating/destroying as many as 64 VFs may sometimes take longer time than usual on some setups.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> SR-IOV
766105	<b>Description:</b> 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".
	<b>Workaround:</b> Upgrade to the latest version of QEMU in the hypervisor.
	<b>Keywords:</b> SR-IOV
413372	<b>Description:</b> SR-IOV non persistent configuration (such as VGT, VST, Host assigned GUIDs, and QP0-enabled VFs) may be lost upon Reset Flow.
	<b>Workaround:</b> Reset Admin configuration post Reset Flow
	<b>Keywords:</b> Reset Flow
926137	<b>Description:</b> IPV6 ping does not use device specified with -I parameter in iputils version s20160308.
	<b>Workaround:</b> Use iputils version s20161105 or above.
	<b>Keywords:</b> General
936768	<b>Description:</b> When querying the HCA core clock, data will be presented in KHz for all cards.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> General



**Table 8 - Known Issues**

Internal Reference Number	Issue
856033	<p><b>Description:</b> 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</p> <p><b>Workaround:</b> Edit the kernel parameters (in grub) and add <code>qiommu.identity_map_qiommu=PCIE0_MMU,PCIE4_MMU</code> (The bus numbers depend on the ConnectX-4 slot.)  Reboot the server.</p> <p><b>Keywords:</b> General</p>
552870/ 548518	<p><b>Description:</b> On rare occasions, under extremely heavy MAD traffic, MAD (Management Datagram) storms might cause soft-lockups in the UMAD layer.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> General</p>
663434	<p><b>Description:</b> 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.</p> <p><b>Workaround:</b> Run <code>update-pciids</code></p> <p><b>Keywords:</b> General</p>
767016	<p><b>Description:</b> Resetting hardware counters after netdev goes up can break statistics scripts.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> General</p>
959842	<p><b>Description:</b> 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.</p> <p><b>Workaround:</b> N/A</p> <p><b>Keywords:</b> ABI Compatibility</p>

**Table 8 - Known Issues**

Internal Reference Number	Issue
919836	<b>Description:</b> ucamtose fails when using a local loopback IP.
	<b>Workaround:</b> Use the device's interface IP instead of loopback IP.
	<b>Keywords:</b> Connection Manager (CM)
781382	<b>Description:</b> The number of local ports that rdma_cm ID can bind to is limited. This limitation depends on the OS dynamics.
	<b>Workaround:</b> Modify the range of available ports for binding, run: <code>sysctl net.ipv4.ip_local_port_range="MIN MAX"</code> The MIN and MAX values can range from 0 to 65535. <b>Note:</b> 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.
	<b>Keywords:</b> Connection Manager (CM)
387061	<b>Description:</b> <code>mlx4_core</code> can allocate up to 64 MSI-X vectors, an MSI-X vector per CPU.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Resources Limitation
553657	<b>Description:</b> Registering a large amount of Memory Regions (MR) may fail because of DMA mapping issues on RHEL 7.0.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> Resources Limitation
736136	<b>Description:</b> The maximum number of HCAs shown by <code>ibstat</code> is 32 HCAs.
	<b>Workaround:</b> N/A
	<b>Keywords:</b> 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](http://www.mellanox.com/pdf/prod_software/MLNX_OFED_Archived_Bug_Fixes.pdf)

**Table 9 - Bug Fixes History**

Internal Ref	Issue
1038933	<b>Description:</b> Fixed a backport issue where IPv6 procedures were called while they were not supported in the underlying kernel.
	<b>Keywords:</b> iw_cm
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
1064722	<b>Description:</b> Added log debug prints when changing HW configuration via DCB. To enable log debug prints, run: <code>ethtool -s &lt;devname&gt; msglvl hw on/off</code>
	<b>Keywords:</b> DCB, msglvl
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
1013076	<b>Description:</b> 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.
	<b>Keywords:</b> IPoIB, Packet Fragmentation
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
1022251	<b>Description:</b> 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.
	<b>Keywords:</b> Soft RoCE
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
1044546	<b>Description:</b> Fixed the issue where a kernel crash used to occur when Rxe device was coupled with a virtual (dummy) device.
	<b>Keywords:</b> Soft RoCE
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0

**Table 9 - Bug Fixes History**

Internal Ref	Issue
1047617	<b>Description:</b> Fixed the issue where a race condition in the RoCE GID cache used to cause for the loss of IP-based GIDs.
	<b>Keywords:</b> RoCE, GID
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
1006768	<b>Description:</b> 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.
	<b>Keywords:</b> RDMACM
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
801807	<b>Description:</b> Fixed an issue where RDMACM connection used to fail upon high connection rate accompanied with the error message: RDMA_CM_EVENT_UNREACHABLE.
	<b>Keywords:</b> RDMACM
	<b>Discovered in Release:</b> 3.0-2.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
869768	<b>Description:</b> Fixed the issue where SR-IOV was not supported in systems with a page size greater than 16KB.
	<b>Keywords:</b> SR-IOV, mlx5, PPC
	<b>Discovered in Release:</b> 4.0-2.0.0.1
	<b>Fixed in Release:</b> 4.1-1.0.2.0
919545	<b>Description:</b> Fixed the issue of when the Kernel becomes out of memory upon driver start, it could crash on SLES 12 SP2.
	<b>Keywords:</b> mlx_5 Eth Driver
	<b>Discovered in Release:</b> 3.4-2.0.0.0
	<b>Fixed in Release:</b> 4.0-2.0.0.1
970668	<b>Description:</b> Fixed the issue where very high stress on DC QP transport might have triggered NMI messages on specific servers.
	<b>Keywords:</b> mlx5 Driver
	<b>Discovered in Release:</b> 4.0-1.0.1.0
	<b>Fixed in Release:</b> 4.0-2.0.0.1
966134	<b>Description:</b> Allowed Ethernet VFs to open Raw Ethernet QPs even if RoCE is not supported for the VF.
	<b>Keywords:</b> mlx4_ib
	<b>Discovered in Release:</b> 3.0-1.0.1
	<b>Fixed in Release:</b> 4.0-2.0.0.1

**Table 9 - Bug Fixes History**

Internal Ref	Issue
864063	<b>Description:</b> Fixed the issue of when Spoof-check may have been turned on for MAC address 00:00:00:00:00:00
	<b>Keywords:</b> mlx4
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-2.0.0.1
869209	<b>Description:</b> Fixed an issue that caused TCP packets to be received in an out of order manner when Large Receive Offload (LRO) is on.
	<b>Keywords:</b> mlx5_en
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 4.0-2.0.0.1
913319	<b>Description:</b> Fixed the issue of low performance when creating many address handles.
	<b>Keywords:</b> libibverbs
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0
912897	<b>Description:</b> Added debug prints to <code>ib_umem_get</code> function to fix lack of error indication when this function fails.
	<b>Keywords:</b> InfiniBand
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0
945887	<b>Description:</b> [ConnectX-3] Fixed the issue where multicast traffic over Raw Ethernet QP on virtual functions were received on the same QP (loopback).
	<b>Keywords:</b> SR-IOV
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0
920292	<b>Description:</b> Fixed three issues in libmlx5 that were found by NVIDIA in the patches that are part of MLNX_OFED v3.4: 1. <code>mlx5_exp_peer_commit_qp</code> 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 <code>mlx5_exp_peer_peek_cq</code> . There is now fallthrough in the <code>IBV_EXP_PEER_PEEK_ABSOLUTE</code> case.
	<b>Keywords:</b> libmlx5
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0

**Table 9 - Bug Fixes History**

Internal Ref	Issue
890285	<b>Description:</b> Fixed the issue where memory allocation for CQ buffers used to fail when increasing the RX ring size.
	<b>Keywords:</b> mlx5_core
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0
867094	<b>Description:</b> Fixed the issue where MLNX_OFED used to fail to load on 4K page ARM architecture.
	<b>Keywords:</b> ARM
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 4.0-1.0.1.0
873538	<b>Description:</b> 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.
	<b>Keywords:</b> biosdavename
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
876329	<b>Description:</b> Fixed the issue of when the error flow was re-factored, the reading of the device caps was excluded from the error recovery flow.
	<b>Keywords:</b> mlx5 Driver
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
876419	<b>Description:</b> Fixed the issue where kernel panic was observed on openibd stop as a result of querying non-existent bond slave.
	<b>Keywords:</b> mlx4_en
	<b>Discovered in Release:</b> 3.3-2.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
868665	<b>Description:</b> Fixed the issue where kernel panic in mlx4_en_get_phys_port_id may occur during server reboot.
	<b>Keywords:</b> mlx4_en
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
882227	<b>Description:</b> 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.
	<b>Keywords:</b> mlx4_en
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0

**Table 9 - Bug Fixes History**

Internal Ref	Issue
887348	<b>Description:</b> Fixed the issue of when <code>prof_sel</code> was invalid, <code>mlx5_core</code> failed upon debug print.
	<b>Keywords:</b> <code>mlx5_core</code>
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
898161	<b>Description:</b> Fixed the issue where a compilation error in kernels of v4.6 or above used to occur due to a large stack size in the <code>get_numa_phys_mask</code> function.
	<b>Keywords:</b> <code>mlx5_core</code>
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
880269	<b>Description:</b> 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.
	<b>Keywords:</b> <code>ib_core</code>
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
887245	<b>Description:</b> Fixed the issue where the system used to pick the dummy <code>ib_isert</code> module instead of the real module on RHEL with errata kernel.
	<b>Keywords:</b> <code>ib_isert</code>
	<b>Discovered in Release:</b> 3.4-1.0.0.0
	<b>Fixed in Release:</b> 3.4-2.0.0.0
854344	<b>Description:</b> Fixed the issue where <code>mlnx_affinity</code> script on RHEL/CentOS7.x host did not disable or enable <code>irqbalancer</code> .
	<b>Keywords:</b> <code>irqbalancer</code>
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
824736	<b>Description:</b> Fixed wrong <code>skprio2UP</code> mapping by removing it and its scripts, such as <code>tc_wrap</code> , from the driver. This mapping should now be done using the kernel's <code>set_egress_map</code> commands. Note: Only for RDMACM over old kernels, the original <code>skprio2UP</code> mapping in <code>tc_wrap</code> remains valid as these kernels do not support <code>set_egress_map</code> .
	<b>Keywords:</b> QoS
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0

**Table 9 - Bug Fixes History**

Internal Ref	Issue
824775	<b>Description:</b> Fixed the issue where starting <code>ibacm</code> daemon failed on Debian based distributions with the following message: <code>/etc/init.d/ibacm: line 37: /sbin/start_daemon: No such file or directory.</code>
	<b>Keywords:</b> <code>ibacm</code>
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
799004	<b>Description:</b> 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.
	<b>Keywords:</b> IPoIB
	<b>Discovered in Release:</b> 3.0-2.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
777733/778099	<b>Description:</b> Fixed the issue where in ARM architecture, multiple kernel panics of <code>mlx4</code> and <code>mlx5</code> drivers were observed as a result of undefined behavior of <code>vmap(virt_to_page(dma_alloc_coherent))</code> call sequence on driver load, by allocating contiguous memory instead of vmapping it.
	<b>Keywords:</b> ARM
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
826686	<b>Description:</b> Fixed the issue where server reboot could get stuck because of kernel panic in <code>mlx4_en_get_drvinfo()</code> that is called from asynchronous event handler.
	<b>Keywords:</b> <code>mlx4_en</code>
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
824130	<b>Description:</b> Fixed the issue where <code>ethtool</code> self test used to fail on interrupt test after timeout if <code>mlx4_ib</code> module was not loaded.
	<b>Keywords:</b> <code>mlx4_en</code>
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
855311	<b>Description:</b> 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.
	<b>Keywords:</b> <code>mlx4</code> driver
	<b>Discovered in Release:</b> 3.0-1.0.1
	<b>Fixed in Release:</b> 3.4-1.0.0.0



**Table 9 - Bug Fixes History**

Internal Ref	Issue
786720	<b>Description:</b> Fixed a crash that used to occur when trying to bring the interface up in a kernel that did not support accelerated RFS (aRFS).
	<b>Keywords:</b> mlx5 driver
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
781747	<b>Description:</b> 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.
	<b>Keywords:</b> SR-IOV
	<b>Discovered in Release:</b> 3.3-1.0.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0
568602	<b>Description:</b> 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.
	<b>Keywords:</b> SR-IOV
	<b>Discovered in Release:</b> 3.0-2.0.0
	<b>Fixed in Release:</b> 3.4-1.0.0.0

## 5 Change Log History

**Table 10 - Change Log History**

Category	Description
<b>4.0-2.0.2.0</b>	
Operating Systems	Added support for Ubuntu v17.04.
<b>4.0-2.0.0.1</b>	
PCIe Error Counting	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] Added the ability to expose physical layer statistical counters to ethtool.
Multiprotocol Label Switching (MPLS) Tagged Packets Classification	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] 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	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] Added the ability to enable/disable RoCE on VFs.
RoCE LAG	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] Added support for RoCE over LAG interface.
Standard ethtool	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] Added support for flow steering and rx-all mode.
SR-IOV Bandwidth Share for Ethernet/RoCE ( <b>beta</b> )	[ <b>ConnectX-4/ConnectX-4 Lx</b> ] 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 ( <b>beta</b> )	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-nvme" 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.
<b>3.4-2.0.0.0</b>	
NVMeoF	Added support for NVMeoF in host/target systems over RDMA.
<b>3.4-1.0.0.0</b>	
VST Q-in-Q	[ <b>ConnectX®-3/ConnectX®-3 Pro</b> ] Added support for Q-in-Q encapsulation per VF in Linux (VST) for ConnectX-3 Pro adapter cards.

**Table 10 - Change Log History**

Category	Description
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).
Raw Ethernet Programming	[ConnectX®-4/ConnectX®-4 Lx] Added new APIs for enhanced raw Ethernet programming: <ul style="list-style-type: none"> <li>• Packet Pacing</li> <li>• TCP Segmentation Offload (TSO)</li> <li>• ToS based steering</li> <li>• Flow ID based steering (beta)</li> <li>• VxLAN based steering (beta)</li> </ul> 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: <ul style="list-style-type: none"> <li>• 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.</li> <li>• When the driver goes down, VF PCI function will not be removed.</li> <li>• Ethernet interface attributes (MTU, state, ring size, etc...) will be recovered after the error recovery stage is completed.</li> <li>• The net device kernel layer will not be aware of any ongoing PCI error recovery process.</li> </ul>
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.

**Table 10 - Change Log History**

Category	Description
Mellanox PeerDirect Async (beta level)	<p>[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:</p> <ul style="list-style-type: none"> <li>• These changes are not backward compatible.</li> <li>• The code adds CQ polling support for peer devices that do not support the NOR operation, replacing it with a GEQ operation.</li> </ul> <p>To see the API changes, refer to the man page.</p>
ABI Incompatibility	<p>[ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4/ConnectX®-4 Lx] Added the ability to fix the issue of preventing the load of MLNX_EN modules when a new kernel is not compatible with these modules.</p>
Mellanox Scalable Hierarchical Aggregation Protocol (SHARP™)	<p>[Connect-IB/ConnectX®-3/ConnectX®-3 Pro/ConnectX®-4]  <b>IB only:</b> 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 <a href="#">SHARP Deployment Guide</a>.</p>
<b>3.3-1.0.0.0</b>	
VF MAC Address Anti-Spoofing	<p>[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.</p>
VF All-multi Mode	<p>[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.  <b>Note:</b> Only privileged/trusted VFs can enter the all-multi RX mode.</p>
VF Promiscuous Mode	<p>[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.  <b>Note:</b> Only privileged/trusted VFs can enter the promiscuous RX mode.</p>
Privileged VF	<p>[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.</p>
DCBX	<p>[ConnectX-4/ConnectX-4 Lx] Added support for standard DCBX CEE API.</p>
Per Priority Counters	<p>[ConnectX-4/ConnectX-4 Lx] Exposed performance counters per priority.</p>
IB Error Counters	<p>[ConnectX-4/ConnectX-4 Lx] Exposed IB sysfs error counters for mlx5 driver.</p>
Accelerated Receive Flow Steering (aRFS)	<p>[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.</p>
Packet Pacing for UDP/TCP	<p>[ConnectX-4/ConnectX-4 Lx] Performs rate limit per UDP/TCP connection.</p>

**Table 10 - Change Log History**

Category	Description
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.
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.
<b>3.2-2.0.0.0</b>	
API Changes	<ul style="list-style-type: none"> <li>• Support FCS scattering for Raw Packet QPs and WQs.</li> <li>• Indication of L4 packet type on the receive side completions</li> <li>• Support CVLAN insertion for WQs</li> </ul>
IPoIB	<ul style="list-style-type: none"> <li>• Added support for the following IPoIB UD QP offloads:               <ul style="list-style-type: none"> <li>• RX check summing (AKA RX csu)</li> <li>• Large Send Offloads (AKA LSO)</li> </ul> </li> </ul> To see the new IPoIB UD mode, run: "ethtool -k <interface>"
<b>3.2-1.0.1.1</b>	
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.

**Table 10 - Change Log History**

Category	Description
Strided WQE User Space	[ConnectX-4 / ConnectX-4 Lx] Striding RQ is a receive queue comprised by work queue elements (i.e. WQEs), where multiple packets of LRO segments (i.e. message) are written to the same WQE.
VLAN Stripping in Linux Verbs	[ConnectX-4 / ConnectX-4 Lx] Adds access to the device's ability to offload the Customer VLAN (cVLAN) header stripping from an incoming packet.
iSER: Remote invalidation support (target and initiator)	[ConnectX-4 / ConnectX-4 Lx] Improves performance by enabling the hardware to perform implicit memory region invalidation.
iSER: Zero-Copy ImmediateData	[ConnectX-4 / ConnectX-4 Lx] Reduces the latency of small writes by avoiding an extra memory copy in the iSER target stack.
iSER: Indirect Memory Registration	[ConnectX-4 / ConnectX-4 Lx] Uses ConnectX®-4 adapter card's Indirect Memory Registration capabilities to avoid bounce buffer strategy implementation and to reduce the latency of highly unaligned vectored IO operations, and also in cases of BIO merging.
Vector Calculation/Erasure coding off-load	[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.
<b>3.1-1.0.3</b>	
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 <a href="#">Section 4, "Bug Fixes History"</a> , on page 43
<b>3.1-1.0.0</b>	
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.

**Table 10 - Change Log History**

Category	Description
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
<b>3.0-2.0.1</b>	
Virtualization	Added support for SR-IOV for ConnectX-4/Connect-IB adapter cards.
<b>3.0-1.0.1</b>	
HCAs	Added support for ConnectX@-4 Single/Dual-Port Adapter supporting up to 100Gb/s.
RoCE per GID	RoCE per GID provides the ability to use different RoCE versions/modes simultaneously.
RoCE Link Aggregation (RoCE LAG): ConnectX-3/ConnectX-3 Pro only	RoCE Link Aggregation (available in kernel 4.0 only) provides failover and link aggregation capabilities for mlx4 device physical ports. In this mode, only one IB port that represents the two physical ports, is exposed to the application layer.
Resource Domain Experimental Verbs	Resource domain is a verb object which may be associated with QP and/or CQ objects on creation to enhance data-path performance.
Alias GUID Support in InfiniBand	Enables the <code>query_gid</code> verb to return the admin desired value instead of the value that was approved by the SM, to prevent a case where the SM is unreachable or a response is delayed, or if the VF is probed into a VM before their GUID is registered with the SM.
Denial Of Service (DOS) MAD Prevention	Denial Of Service MAD prevention is achieved by assigning a threshold for each agent's RX. Agent's RX threshold provides a protection mechanism to the host memory by limiting the agents' RX with a threshold.
QoS per VF (Rate Limit per VF)	Virtualized QoS per VF, (supported in ConnectX-3/ConnectX-3 Pro adapter cards only with firmware v2.33.5100 and above), limits the chosen VFs' throughput rate limitations (Maximum throughput). The granularity of the rate limitation is 1Mbits.
Ignore Frame Check Sequence (FCS) Errors	Upon receiving packets, the packets go through a checksum validation process for the FCS field. If the validation fails, the received packets are dropped. Using this feature, enables you to choose whether or not to drop the frames in case the FCS is wrong and use the FCS field for other info.
Sockets Direct Protocol (SDP)	Sockets Direct Protocol (SDP) is a byte-stream transport protocol that provides TCP stream semantics. and utilizes InfiniBand's advanced protocol offload capabilities.
Scalable Subnet Administration (SSA)	The Scalable Subnet Administration (SSA) solves Subnet Administrator (SA) scalability problems for Infiniband clusters. It distributes the needed data to perform the path-record-calculation needed for a node to connect to another node, and caches these locally in the compute (client) nodes. SSA <sup>a</sup> requires AF_IB address family support (3.12.28-4 kernel and later).

**Table 10 - Change Log History**

Category	Description
SR-IOV in ConnectX-3 cards	Changed the Alias GUID support behavior in InfiniBand.
LLR max retransmission rate	Added LLR max retransmission rate as specified in Vendor Specific MAD V1.1, Table 110 - PortLLRStatistics MAD Description ibdiagnet presents the LLR max_retransmission_rate counter as part of the PM_INFO in db_csv file.
Experimental Verbs	Added the following verbs: <ul style="list-style-type: none"> <li>• <code>ibv_exp_create_res_domain</code></li> <li>• <code>ibv_exp_destroy_res_domain</code></li> <li>• <code>ibv_exp_query_intf</code></li> <li>• <code>ibv_exp_release_intf</code></li> </ul> Added the following interface families: <ul style="list-style-type: none"> <li>• <code>ibv_exp_qp_burst_family</code></li> <li>• <code>ibv_exp_cq_family</code></li> </ul>
<b>2.4-1.0.4</b>	
Bug Fixes	See “Bug Fixes History” on page 43.
<b>2.4-1.0.0</b>	
mlx4_en net-device Ethtool	Added support for Ethtool speed control and advertised link mode.
	Added ethtool txvlan control for setting ON/OFF hardware TX VLAN insertion: <code>ethtool -k txvlan [on/off]</code>
	Ethtool report on port parameters improvements.
	Ethernet TX packet rate improvements.
RoCE	RoCE uses now all available EQs and not only the 3 legacy EQs.
InfiniBand	IRQ affinity hints are now set when working in InfiniBand mode.
Virtualization	VXLAN fixes and performance improvements.
libmlx4 & libmlx5	Improved message rate of short messages.
libmlx5	Added ConnectX@-4 device (4114) to the list of supported devices ( <code>hca_table</code> ),
Storage	Added iSER Target driver.
Ethernet net-device	New adaptive interrupt moderation scheme to improve CPU utilization.
	RSS support of fragmented IP datagram.
Connect-IB Virtual Function	Added Connect-IB Virtual Function to the list of supported devices.
<b>2.3-2.0.5</b>	



**Table 10 - Change Log History**

Category	Description
mlx5_core	<p>Added the following files under <code>/sys/class/infiniband/mlx5_0/mr_cache/</code>:</p> <ul style="list-style-type: none"> <li><code>rel_timeout</code>: Defines the minimum allowed time between the last MR creation to the first MR released from the cache. When <code>rel_timeout = -1</code>, MRs are not released from the cache</li> <li><code>rel_imm</code>: Triggers the immediate release of excess MRs from the cache when set to 1. When all excess MRs are released from the cache, <code>rel_imm</code> is reset back to 0.</li> </ul>
Bug Fixes	See "Bug Fixes History" on page 43.
<b>2.3-2.0.1</b>	
Bug Fixes	See "Bug Fixes History" on page 43.
<b>2.3-2.0.0</b>	
Connect-IB	Added Suspend to RAM (S3).
Reset Flow	Added Enhanced Error Handling for PCI (EEH), a recovery strategy for I/O errors that occur on the PCI bus.
Register Contiguous Pages	Added the option to ask for a specific address when the register memory is using contiguous page.
mlx5_core	Moved the <code>mr_cache</code> subtree from <code>debugfs</code> to <code>mlx5_ib</code> while preserving all its semantics.
InfiniBand Utilities	Updated the <code>ibutils</code> package. Added to the <code>ibdiagnet</code> tool the " <code>ibdiagnet2.mlnx_cntrs</code> " option to enable reading of Mellanox diagnostic counters.
Bug Fixes	See "Bug Fixes History" on page 43.
<b>2.3-1.0.1</b>	
Ethernet	<p>Added support for arbitrary UDP port for VXLAN. From upstream 3.15-rc1 and onward, it is possible to use arbitrary UDP port for VXLAN. This feature requires firmware version 2.32.5100 or higher. Additionally, the following kernel configuration option <code>CONFIG_MLX4_EN_VXLAN=y</code> must be enabled.</p> <p>MLNX_OFED no longer changes the OS <code>sysctl</code> TCP parameters.</p> <p>Added Explicit Congestion Notification (ECN) support</p> <p>Added Flow Steering: A0 simplified steering support</p> <p>Added RoCE v2 support</p>

**Table 10 - Change Log History**

Category	Description
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.
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
<b>2.2-1.0.1</b>	
Reset Flow	Reset Flow is not activated by default. It is controlled by the mlx4_core' internal_err_reset' module parameter.

**Table 10 - Change Log History**

Category	Description
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 “ <i>openibd Script</i> ” in the MLNX-OFED User Manual.
InfiniBand Core	Asymmetric MSI-X vectors allocation for the SR-IOV hypervisor and guest instead of allocating 4 default MSI-X vectors. The maximum number of MSI-X vectors is <code>num_cpu</code> for port ConnectX®-3 has 1024 MSI-X vectors, 28 MSI-X vectors are reserved. <ul style="list-style-type: none"> <li>Physical Function - gets the number of MSI-X vectors according to the <code>pf_msix_table_size</code> (multiple of 4 - 1) INI parameter</li> <li>Virtual Functions – the remaining MSI-X vectors are spread equally between all VFs, according to the <code>num_vfs mlx4_core</code> module parameter</li> </ul>
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: <ul style="list-style-type: none"> <li>Use of affinity hints (based on NUMA node of the device) to indicate the IRQ balancer daemon on the optimal IRQ affinity</li> <li>Improvement in buffers allocation schema (based on the hint above)</li> <li>Improvement in the adaptive interrupt moderation algorithm</li> </ul>
<b>2.1-1.0.6</b>	
IB Core	Added allocation success verification process to <code>ib_alloc_device</code> .
dapl	dapl is recompiled with no FCA support.
openibd	Added the ability to bring up child interfaces even if the parent's <code>ifcfg</code> file is not configured.
libmlx4	Unmapped the <code>hca_clock_page</code> parameter from <code>mlx4_uninit_context</code> .
scsi_transport_srp	<code>scsi_transport_srp</code> cannot be cleared up when <code>rport</code> reconnecting fails.
mlnxofedinstall	Added support for the following parameters: <ul style="list-style-type: none"> <li>'--umad-dev-na'</li> <li>'--without-&lt;package&gt;'</li> </ul>

**Table 10 - Change Log History**

Category	Description
Content Packages Updates	The following packages were updated: <ul style="list-style-type: none"> <li>• bupc to v2.2-407</li> <li>• mstflint to v3.5.0-1.1.g76e4acf</li> <li>• perfest to v2.0-0.76.gbf9a463</li> <li>• hcoll to v2.0.472-1</li> <li>• Openmpi to v1.6.5-440ad47</li> <li>• dapl to v2.0.40</li> </ul>
<b>2.1-1.0.0</b>	
EoIB	EoIB is supported only in SLES11SP2 and RHEL6.4.
eIPoIB	eIPoIB is currently at GA level.
Connect-IB®	Added the ability to resize CQs.
IPoIB	Reusing DMA mapped SKB buffers: Performance improvements when IOMMU is enabled.
mlnx_en	Added reporting autonegotiation support.
	Added Transmit Packet Steering (XPS) support.
	Added reporting 56Gbit/s link speed support.
	Added Low Latency Socket (LLS) support.
	Added check for dma_mapping errors.
eIPoIB	Added non-virtual environment support.
<b>2.0-3.0.0</b>	
Operating Systems	Additional OS support: <ul style="list-style-type: none"> <li>• SLES11SP3</li> <li>• Fedora16, Fedora17</li> </ul>
Hardware	Added Connect-IB™ support
Installation	Added ability to install MLNX_OFED with SR-IOV support.
	Added Yum installation support
EoIB	EoIB (at beta level) is supported only in SLES11SP2 and RHEL6.4
mlx4_core	Modified module parameters to associate configuration values with specific PCI devices identified by their bus/device/function value format
mlx4_en	Reusing DMA mapped buffers: major performance improvements when IOMMU is enabled
	Added Port level QoS support
IPoIB	Reduced memory consumption
	Limited the number TX and RX queues to 16
	Default IPoIB mode is set to work in Datagram, except for Connect-IB™ adapter card which uses IPoIB with Connected mode as default.
Storage	iSER (at GA level)

**Table 10 - Change Log History**

Category	Description
<b>2.0-2.0.5</b>	
Virtualization	SR-IOV for both Ethernet and InfiniBand (at Beta level)
Ethernet Network	RoCE over SR-IOV (at Beta level)
	eIPoIB to enable IPoIB in a Para-Virtualized environment (at Alpha level)
	Ethernet Performance Enhancements (NUMA related and others) for 10G and 40G
	Ethernet Time Stamping (at Beta level)
	Flow Steering for Ethernet and InfiniBand. (at Beta level)
	Raw Eth QPs: <ul style="list-style-type: none"> <li>• Checksum TX/RX</li> <li>• Flow Steering</li> </ul>
InfiniBand Network	Contiguous pages: <ul style="list-style-type: none"> <li>• Internal memory allocation improvements</li> <li>• Register shared memory</li> <li>• Control objects (QPs, CQs)</li> </ul>
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).