Connect. Accelerate. Outperform.™

Mellanox OFED for Linux Release Notes

Rev 2.3-1.0.1

www.mellanox.com

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

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

Mellanox Technologies, Ltd.
Beit Mellanox
PO Box 586 Yokneam 20692
Israel
www.mellanox.com
Tel: +972 (0)74 723 7200
Fax: +972 (0)4 959 3245

Table of Contents

Table of Contents 3
List Of Tables 5
Release Update History 7
Chapter 1 Overview 8
1.1 Main Features in This Release 8
1.2 Content of Mellanox OFED for Linux 8
1.3 Supported Platforms and Operating Systems 9
1.3.1 Supported Hypervisors 10
1.3.2 Supported Non-Linux Virtual Machines 10
1.4 Hardware and Software Requirements 11
1.5 Supported HCAs Firmware Versions 11
1.6 Compatibility 12
1.7 RoCE Modes Matrix 12
Chapter 2 Changes in Rev 2.3-1.0.1 From Rev 2.2-1.0.1 13
2.1 API Changes in MLNX_OFED Rev 2.3-1.0.1 14
Chapter 3 Known Issues 15
3.1 IPoIB Known Issues 15
3.2 Ethernet Known Issues 17
3.3 General Known Issues 18
3.4 VGT+ Known Issues 18
3.5 eIPoIB Known Issues 18
3.6 XRC Known Issues 19
3.7 ABI Compatibility Known Issues 19
3.8 System Time Known Issues 20
3.9 ConnectX®-3 Adapter Cards Family Known Issues 20
3.10 Verbs Known Issues 20
3.11 Resiliency Known Issues 20
3.12 Driver Start Known Issues 21
3.13 Performance Tools Known Issues 21
3.14 Performance Known Issues 22
3.15 Connection Manager (CM) Known Issues 22
3.16 SR-IOV Known Isuues 23
3.17 Port Type Management Known Issues 24
3.18 Flow Steering Known Issues 25
3.19 Quality of Service Known Issues 25
3.20 Installation Known Issues 25
3.21 Driver Upload Known Issues 26
3.22 InfiniBand Counters Known Issues 26
3.23 UEFI Secure Boot Known Issues 26
3.24 Fork Support Known Issues 26

3.25 ISCSI over IPoIB Known Issues 27
3.26 MLNX_OFED Sources Known Issues 27
3.27 InfiniBand Utilities Known Issues. 27
3.28 mlx5 Driver Known Issues 27
3.29 Ethernet Performance Counters Known Issues 27
3.30 Uplinks Known Issues 28
3.31 Resources Limitation Known Issues 29
3.32 RoCE Known Issues 29
3.33 Storage Known Issues 31
3.34 SRP Known Issues. 31
3.35 SRP Interop Known Issues 31
3.36 DDN Storage Fusion 10000 Target Known Issues 31
3.37 Oracle Sun ZFS storage 7420 Known Issues. 31
3.38 iSER Known Issues 32
3.39 ZFS Appliance Known Issues 32

Chapter 4 Bug Fixes History. 33

Chapter 5 Change Log History. 35

Chapter 6 API Change Log History. 39

List Of Tables

Table 1: Release Update History 7
Table 2: Mellanox OFED for Linux Software Components 8
Table 3: Supported Platforms and Operating Systems 9
Table 4: Additional Software Packages 11
Table 5: Supported HCAs Firmware Versions 11
Table 6: MLNX_OFED Rev 2.3-1.0.1 Compatibility Matrix 12
Table 7: RoCE Modes Matrix 12
Table 8: Changes in v2.3-1.0.1 13
Table 9: API Change Log History 14
Table 10: IPoIB Known Issues 15
Table 11: Ethernet Known Issues 17
Table 12: General Known Issues 18
Table 13: VGT+ Known Issues 18
Table 14: eIPoIB Known Issues 18
Table 15: XRC Known Issues 19
Table 16: ABI Compatibility Known Issues 19
Table 17: System Time Known Issues 20
Table 18: ConnectX®-3 Adapter Cards Family Known Issues 20
Table 19: Verbs Known Issues 20
Table 20: Resiliency Known Issues 20
Table 21: Driver Start Known Issues 21
Table 22: Performance Tools Known Issues 21
Table 23: Performance Known Issues 22
Table 24: Connection Manager (CM) Known Issues 22
Table 25: SR-IOV Known Issues 23
Table 26: Port Type Management Known Issues 24
Table 27: Flow Steering Known Issues 25
Table 28: Quality of Service Known Issues 25
Table 29: Installation Known Issues 25
Table 30: Driver Upload Known Issues 26
Table 31: InfiniBand Counters Known Issues 26
Table 32: UEFI Secure Boot Known Issues 26
Table 33: Fork Support Known Issues 26
Table 34: ISCSI over IPoIB Known Issues 27
Table 35: MLNX_OFED Sources Known Issues 27

Table 36: InfiniBand Utilities Known Issues 27
Table 37: mlx5 Driver Known Issues 27
Table 38: Ethernet Performance Counters Known Issues 27
Table 39: Uplinks Known Issues 28
Table 40: Resources Limitation Known Issues 29
Table 41: RoCE Known Issues 29
Table 42: Storage Known Issues 31
Table 43: SRP Known Issues. 31
Table 44: SRP Interop Known Issues 31
Table 45: DDN Storage Fusion 10000 Target Known Issues 31
Table 46: Oracle Sun ZFS storage 7420 Known Issues 31
Table 47: iSER Known Issues 32
Table 48: ZFS Appliance Known Issues 32
Table 49: Fixed Bugs List 33
Table 50: Change Log History. 35
Table 51: API Change Log History 39

Release Update History

Table 1 - Release Update History

Release	Date	Description
2.3-1.0.1	19 October 2014	Added Section 1.7, “RoCE Moded Matrix”, on page 12
	23 September 2014	Updated the Known Issue section. Added SR-IOV and Resources Limitation Known Issues.
	September 2014	Initial version

1 Overview

These are the release notes of Mellanox OFED for Linux Driver, Rev 2.3-1.0.1. Mellanox OFED is a single Virtual Protocol Interconnect (VPI) software stack and operates across all Mellanox network adapter solutions supporting the following uplinks to servers:

- 10, 20, 40 and 56 Gb/s InfiniBand (IB)
- 10, 40 and 56[1] Gb/s Ethernet
- 2.5 or 5.0 GT/s PCI Express 2.0
- 8 GT/s PCI Express 3.0

1.1 Main Features in This Release

MLNX_OFED Rev 2.3-1.0.1 provides the following new features:

- Secure Host
- Virtual Guest Tagging (VGT+)
- User-Mode Memory Registration (UMR)
- Masked Atomics
- On-Demand-Paging (ODP)
- Reset Flow for ConnectX®-3 (+SR-IOV)
- RoCE v2
- Checksum offload for packets without L4 header
- Flow Steering: A0 simplified steering
- Memory re-registration
- Cable EEPROM reporting
- 128 Byte Completion Queue Entry (CQE)
- Explicit Congestion Notification (ECN)
- Disable/Enable ethernet RX VLAN tag striping offload via ethtool
- Windows Virtual Machine over Linux KVM Hypervisor (SR-IOV with InfiniBand only)

1.2 Content of Mellanox OFED for Linux

Mellanox OFED for Linux software contains the following components:

Table 2 - Mellanox OFED for Linux Software Components

Components	Description
OpenFabrics core and ULPs	• IB HCA drivers (mlx4, mlx5) • core • Upper Layer Protocols: IPoIB, SRP and iSER Initiator

1. 56 GbE is a Mellanox propriety link speed and can be achieved while connected to Mellanox SX10XX switch series

Table 2 - Mellanox OFED for Linux Software Components

Components	Description
OpenFabrics utilities	• OpenSM: IB Subnet Manager with Mellanox proprietary Adaptive Routing • Diagnostic tools • Performance tests
MPI	• OSU MPI (mvapich2-1.9-1) stack supporting the InfiniBand interface • Open MPI stack 1.6.5 and later supporting the InfiniBand interface • MPI benchmark tests (OSU benchmarks, Intel MPI benchmarks, Presta)
PGAS	• ScalableSHMEM v2.2 supporting InfiniBand, MXM and FCA • ScalableUPC v2.2 supporting InfiniBand, MXM and FCA
HPC Acceleration packages	• Mellanox MXM v3.0 (p2p transport library acceleration over Infiniband) • Mellanox FCA v2.5 (MPI/PGAS collective operations acceleration library over InfiniBand) • KNEM, Linux kernel module enabling high-performance intra-node MPI/PGAS communication for large messages
Extra packages	• ibutils2 • ibdump • MFT
Sources of all software modules (under conditions mentioned in the modules' LICENSE files) except for MFT, OpenSM plugins, ibutils2, and ibdump	
Documentation	

1.3 Supported Platforms and Operating Systems

The following are the supported OSs in MLNX_OFED Rev 2.3-1.0.1:

Table 3 - Supported Platforms and Operating Systems

Operating System	Platform
RHEL/CentOS 6.3	x86_64
RHEL/CentOS 6.4	x86_64/PPC
RHEL/CentOS 6.5	x86_64/PPC
RHEL/CentOS 7.0	x86_64/PPC
SLES11 SP2	x86_64/PPC
SLES11 SP3	x86_64/PPC
OEL 6.3	x86_64
OEL 6.4	x86_64
OEL 6.5	x86_64
Citrix XenServer Host 6.2	i686
Fedora 19	x86_64
Fedora 20	x86_64
Ubuntu 12.04.4	x86_64
Ubuntu 14.04	x86_64/PPC4le
Debian 7.2	x86_64

Table 3 - Supported Platforms and Operating Systems

Operating System	Platform
Debian 7.4	x86_64
Debian 7.5	x86_64
kernel 3.10.48[a]	
kernel 3.11.10[a]	
kernel 3.12.24[a]	
kernel 3.13.1[a]	
kernel 3.14.12[a]	
kernel 3.15.5[a]	

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

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

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

1.3.1 Supported Hypervisors

The following are the supported hypervisors in MLNX_OFED Rev 2.3-1.0.1:

- KVM
- Xen

1.3.2 Supported Non-Linux Virtual Machines

The following are the supported Non-Linux (InfiniBand only) Virtual Machines in MLNX_OFED Rev 2.3-1.0.1:

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

1.4 Hardware and Software Requirements

The following are the hardware and software requirements of MLNX_OFED Rev 2.3-1.0.1.

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

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

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

Table 4 - Additional Software Packages

Operating System	Required Packages Installation Command
RHEL/OEL/ Fedora	yum install perl pciutils python gcc-gfortran libxml2-python tcsh libnl.i686 libnl expat glib2 tcl libstdc++ bc tk gtk2 atk cairo numactl
XenServer	yum install perl pciutils python libxml2-python libnl expat glib2 tcl bc libstdc++ tk
SLES 11 SP2	zypper install perl pciutils python libnl-32bit libxml2-python tcsh libnl libstdc++46 expat glib2 tcl bc tklibcurl4 gtk2 atk cairo
SLES 11 SP3	zypper install perl pciutils python libnl-32bit libxml2-python tcsh libstdc++43 libnl expat glib2 tcl bc tk libcurl4 gtk2 atk cairo
Ubuntu/Debian	apt-get install perl dpkg autotools-dev autoconf libtool automake1.10 automake m4 dkms debhelper tcl tcl8.4 chrpath swig graphviz tcl-dev tcl8.4-dev tk-dev tk8.4-dev bison flex dpatch zlib1g-dev curl libcurl4-gnutls-dev python-libxml2 libvirt-bin libvirt0 libnl-dev libglib2.0-dev libgfortran3 automake m4

1.5 Supported HCAs Firmware Versions

MLNX_OFED Rev 2.3-1.0.1 supports the following Mellanox network adapter cards firmware versions:

Table 5 - Supported HCAs Firmware Versions

HCA	Recommended Firmware Rev.	Additional Firmware Rev. Supported
Connect-IB®	Rev 10.10.4020	Rev 10.10.3000
ConnectX®-3 Pro	Rev 2.32.5100	Rev 2.31.5050
ConnectX®-3	Rev 2.32.5100	Rev 2.31.5050
ConnectX®-2	Rev 2.9.1000	Rev 2.9.1000

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

1.6 Compatibility

MLNX_OFED Rev 2.3-1.0.1 is compatible with the following:

Table 6 - MLNX_OFED Rev 2.3-1.0.1 Compatibility Matrix

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

a. MLNX_OFED v2.3-1.0.1 was tested with this switch however, additional switches might be supported as well.

1.7 RoCE Modes Matrix

The following is RoCE modes matrix:

Table 7 - RoCE Modes Matrix

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

2 Changes in Rev 2.3-1.0.1 From Rev 2.2-1.0.1

Table 8 - Changes in v2.3-1.0.1

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 reregistartions 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 becames equal to the "master_sm_priority". So that additional SM handover is avoided. Default value of the master_sm_priority is 14. To disable this feature, set the "master_sm_priority" in opensm.conf to 0.
	Added credit-loop free unicast/multicast updn/ftree routing
	Added multithreaded Minhop/UPDN/DOR routing
RoCE	Added IP routable RoCE modes. For further information, please refer to the MLNX_OFED User Manual.
Installation	Added apt-get installation support.
Ethernet	Added support for arbitrary UDP port for VXLAN. From upstream 3.15-rc1 and onward, it is possible to use arbitrary UDP port for VXLAN. This feature requires firmware version 2.32.5100 or higher. Additionally, the following kernel configuration option CONFIG_MLX4_EN_VXLAN=y must be enabled.
	MLNX_OFED no longer changes the OS sysctl TCP parameters.
	Added Explicit Congestion Notification (ECN) support
	Added Flow Steering: A0 simplified steering support
	Added RoCE v2 support

Table 8 - Changes in v2.3-1.0.1

Category	Description
InfiniBand Network	Added Secure host to enable the device to protect itself and the subnet from malicious software.
	Added User-Mode Memory Registration (UMR) to enable the usage of RDMA operations and to scatter the data at the remote side through the definition of appropriate memory keys on the remote side.
	Added On-Demand-Paging (ODP), a technique to alleviate much of the shortcomings of memory registration.
	Added Masked Atomics operation support
	Added Checksum offload for packets without L4 header support
	Added Memory re-registration to allow the user to change attributes of the memory region.
Resiliency	Added Reset Flow for ConnectX®-3 (+SR-IOV) support
SR-IOV	Added Virtual Guest Tagging (VGT+), an advanced mode of Virtual Guest Tagging (VGT), in which a VF is allowed to tag its own packets as in VGT, but is still subject to an administrative VLAN trunk policy.
Ethtool	Added Cable EEPROM reporting support
	Disable/Enable ethernet RX VLAN tag striping offload via ethtool
	128 Byte Completion Queue Entry (CQE)
Non-Linux Virtual Machines	Added Windows Virtual Machine over Linux KVM Hypervisor (SR-IOV with InfiniBand only) support

2.1 API Changes in MLNX_OFED Rev 2.3-1.0.1

The following are the API changes in MLNX_OFED Rev 2.3-1.0.1:

Table 9 - API Change Log History

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

3 Known Issues

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

3.1 IPoIB Known Issues

Table 10 - IPoIB Known Issues

Index	Description	Workaround
1.	When user increases receive/send a buffer, it might consume all the memory when few child's interfaces are created.	-
2.	The size of send queue in Connect-IB® cards cannot exceed 1K.	-
3.	In 32 bit devices, the maximum number of child interfaces that can be created is 16. Creating more that, might cause out-of-memory issues.	-
4.	In RHEL7.0, the Network-Manager can detect when the carrier of one of the IPoIB interfaces is OFF and can decide to disable its IP address.	Set "`ignore-carrier`" for the corresponding device in NetworkManager.conf. For further information, please refer to *"man NetworkManager.conf"*
5.	IPoIB interface does not function properly if a third party application changes the PKey table. We recommend modifying PKey tables via OpenSM.	-
6.	Fallback to the primary slave of an IPoIB bond does not work with ARP monitoring. (https://bugs.openfabrics.org/show_bug.cgi?id=1990)	-
7.	Out-of memory issue might occur due to overload of interfaces created.	To calculate the allowed memory per each IPoIB interface check the following: • Num-rings = min(num-cores-on-that-device, 16) • Ring-size = 512 (by default, it is module parameter) • UD memory: 2 * num-rings * ring-size * 8K • CM memory: ring-size * 64k • Total memory = UD mem + CM mem
8.	Connect-IB does not reach the bidirectional line rate	Optimize the IPoIB performance in Connect-IB: `cat /sys/class/net/<interface>/device/local_cpus > /sys/class/net/<interface>/queues/rx-0/rps_cpus`
9.	If the `CONNECTED_MODE=no` parameter is set to "`no`" or missing from the ifcfg file for Connect-IB® IPoIB interface then the `"service network restart"` will hang.	Set the `CONNECTED_MODE=yes parameter` in the ifcfg file for Connect-IB® interface.
10.	Joining a multicast group in the SM using the RDMA_CM API requires IPoIB to first join the broadcast group.	-

Table 10 - IPoIB Known Issues (Continued)

Index	Description	Workaround
11.	Whenever the IOMMU parameter is enabled in the kernel it can decrease the number of child interfaces on the device according to resource limitation. The driver will stuck after unknown amount of child interfaces creation. For further information, please see: https://access.redhat.com/site/articles/66747 http://support.citrix.com/article/CTX136517 http://www.novell.com/support/kb/doc.php?id=7012337	To avoid such issue: • Decrease the amount of the RX receive buffers (module parameter, the default is 512) • Decrease the number of RX rings (sys/fs or ethtool in new kernels) • Avoid using IOMMU if not required For **KVM** users: Run: `echo 1 > /sys/module/kvm/parameters/allow_unsafe_assigned_interrupts` To make this change persist across reboots, add the following to the `/etc/modprobe.d/kvm.conf` file (or create this file, if it does not exist): `options kvm allow_unsafe_assigned_interrupts=1` `kernel parameters`
12.	System might crash in `skb_checksum_help()` while performing TCP retransmit involving packets with 64k packet size. A similar out to the below will be printed: `kernel BUG at net/core/dev.c:1707!` `invalid opcode: 0000 [#1] SMP` `RIP: 0010:[<ffffffff81448988>]` `skb_checksum_help+0x148/0x160` `Call Trace:` `<IRQ>` `[<ffffffff81448d83>]` `dev_hard_start_xmit+0x3e3/0x530` `[<ffffffff8144c805>]` `dev_queue_xmit+0x205/0x550` `[<ffffffff8145247d>]` `neigh_connected_output+0xbd/0x1` … …	Use UD mode in ipoib
13.	When InfiniBand ports are removed from the host (e.g when changing port type from IB to Eth or removing a card from the PCI bus) the remaining IPoIB interface might be renamed.	To avoid it and have persistent IPoIB network devices names for ConnectX ports, add to the `/etc/udev/rules.d/70-persistent-net.rules` file: `SUBSYSTEM=="net", ACTION=="add", DRIVERS=="?*", ATTR{address}=="*<Port GID>", NAME="ibN"` Where N is the IPoIB required interface index
14.	After releasing a bond interface that contains IPoIB slaves, a call trace might be printed into the dmesg.	-

Table 10 - IPoIB Known Issues (Continued)

Index	Description	Workaround
15.	The LRO feature cannot be disabled via ethtool on kernels> 3.9.	-

3.2 Ethernet Known Issues

Table 11 - Ethernet Known Issues

Index	Description	Workaround
1.	When creating more than 125 VLANs and SR-IOV mode is enabled, a kernel warning message will be printed indicating that the native VLAN is created but will not work with RoCE traffic. kernel warning: `mlx4_core 0000:07:00.0: vhcr command ALLOC_RES (0xf00) slave:0 in_param 0x7e in_mod=0x107, op_mod=0x1 failed with error:0, status -28`	-
2.	Kernel panic might occur during fio splice in kernels before 2.6.34-rc4.	Use kernel v2.6.34-rc4 which provides the following solution: `baff42a net: Fix oops from tcp_collapse() when using splice()`
3.	In PPC systems when QoS is enabled a harmless Kernel DMA mapping error messages might appear in kernel log (iommu related issue).	-
4.	Transmit timeout might occur on RH6.3 as a result of lost interrupt (OS issue). In this case, the following message will be shown in dmesg: `do_IRQ: 0.203 No irq handler for vector (irq -1)`	-
5.	The default priority to TC mapping assigns all priorities to TC0. This configuration achieves fairness in transmission between priorities but may cause undesirable PFC behavior where pause request for priority "n" affects all other priorities.	Run: `mlnx_qos –i <dev> -p 0,1,2,3,4,5,6,7 -s ets,ets,ets,ets,ets,ets,ets,ets -t 12,13,12,13,12,13,12,13` This needs to be applied every time after loading the `mlx4_en` driver.
6.	Mixing ETS and strict QoS policies for TCs in 40GbE ports may cause inaccurate results in bandwidth division among TCs.	-
7.	Creating a VLAN with user priority >= 4 on ConnectX® 2 HCA is not supported.	-
8.	Affinity hints are not supported in Xen Hypervisor (an irqblancer issue). This causes a non-optimal IRQ affinity.	To overcome this issues, run: `set_irq_affinity.sh eth<x>`

3.3 General Known Issues

Table 12 - General Known Issues

Index	Description	Workaround
1.	On ConnectX-2/ConnectX-3 Ethernet adapter cards, there is a mismatch between the GUID value returned by firmware management tools and that returned by fabric/driver utilities that read the GUID via device firmware (e.g., using ibstat). Mlxburn/ flint return 0xffff as GUID while the utilities return a value derived from the MAC address. For all driver/ firmware/software purposes, the latter value should be used.	N/A. Please use the GUID value returned by the fabric/driver utilities (not 0xffffff).

3.4 VGT+ Known Issues

Table 13 - VGT+ Known Issues

Index	Description	Workaround
1.	Before adding a VLAN on the VM, the parent interface should be brought up. Otherwise VLAN creation will fail and the following message will be presented: "Fail to register network rule."	-
2.	Bringing down and up the parent interface with VLANs configured over it, may result in traffic over VLANs being lost.	
3.	On some of the OSes the callback `ndo_vlan_rx_add/kill_vid` returns void, therefore ethX.Y is created. However, if VLAN Y is not listed in the set of the allowed VLANs, no traffic will pass.	
4.	When untagged traffic is not allowed, the below message will appear on Dom0 after driver restart on DomU: `mlx4_core 0000:16:00.0: vhcr command ALLOC_RES (0xf00) slave:1 in_param 0x0 in_mod=0x207, op_mod=0x1 failed with error:0, status -1`	-

3.5 eIPoIB Known Issues

Table 14 - eIPoIB Known Issues

Index	Description	Workaround
1.	On rare occasions, upon driver restart the following message is shown in the dmesg: `'cannot create duplicate filename '/ class/net/eth_ipoib_interfaces'`	-
2.	No indication is received when eIPoIB is non functional.	Run `'ps -ef | grep ipoibd'` to verify its functionality.
3.	eIPoIB requires libvirtd, python	-

Table 14 - eIPoIB Known Issues (Continued)

Index	Description	Workaround
4.	eIPoIB supports only active-backup mode for bonding.	-
5.	eIPoIB supports only VLAN Switch Tagging (VST) mode on guests.	-
6.	IPv6 is currently not supported in eIPoIB	-
7.	eIPoIB cannot run when Flow Steering is enabled	-

3.6 XRC Known Issues

Table 15 - XRC Known Issues

Index	Description	Workaround
1.	Legacy API is deprecated, thus when recompiling applications over MLNX_OFED v2.0-3.x.x, warnings such as the below are displayed. `rdma.c:1699: warning: 'ibv_open_xrc_domain' is deprecated (declared at /usr/include/infiniband/ofa_verbs.h:72)` `rdma.c:1706: warning: 'ibv_create_xrc_srq' is deprecated (declared at /usr/include/infiniband/ofa_verbs.h:89)` These warnings can be safely ignored.	-
2.	XRC is not functional in heterogeneous clusters containing non Mellanox HCAs.	-
3.	XRC options do not work when using qperf tool.	Use perftest instead
4.	Out-of memory issue might occur due to overload of XRC receive QP with non zero receive queue size created. XRC QPs do not have receive queues.	-

3.7 ABI Compatibility Known Issues

Table 16 - ABI Compatibility Known Issues

Index	Description	Workaround
1.	MLNX_OFED Rev 2.3-1.0.1 is not ABI compatible with previous MLNX_OFED/OFED versions.	Recompile the application over the new MLNX_OFED version

3.8 System Time Known Issues

Table 17 - System Time Known Issues

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

3.9 ConnectX®-3 Adapter Cards Family Known Issues

Table 18 - ConnectX®-3 Adapter Cards Family Known Issues

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

3.10 Verbs Known Issues

Table 19 - Verbs Known Issues

Index	Description	Workaround
1.	Using libnl1_1_3~26 or earlier, requires ibv_create_ah protection by a lock for multi-threaded applications.	-

3.11 Resiliency Known Issues

Table 20 - Resiliency Known Issues

Index	Description	Workaround
1.	Reset Flow can run on XenServer 6 only after the active user space applications running verbs are terminated.	-
2.	SR-IOV non persistent configuration (such as VGT, VST, Host assigned GUIDs, and QP0-enabled VFs) may be lost upon Reset Flow.	Reset Admin configuration post Reset Flow
3.	Upon Reset Flow or after running restart driver, Ethernet VLANs are lost.	Reset the VLANs using the `ifup` command.
4.	Restarting the driver or running `connectx_port_config` when Reset Flow is running might result in a kernel panic	-
5.	Networking configuration (e.g. VLANs, IPv6) should be statically defined in order to have them set after Reset Flow as of after restart driver.	-

3.12 Driver Start Known Issues

Table 21 - Driver Start Known Issues

Index	Description	Workaround
1.	"Out of memory" issues may rise during drivers load depending on the values of the driver module parameters set (e.g. log_num_cq).	-
2.	When reloading/starting the driver using the `/etc/init.d/openibd` the following messages are displayed if there is a third party RPM or driver installed: "Module mlx4_core does not belong to MLNX_OFED" or "Module mlx4_core belong to <rpm name> which is not a part of MLNX_OFED"	Remove the third party RPM/non MLNX_OFED drivers directory, run: `"depmod"` and then rerun `"/etc/init.d/openibd restart"`
3.	Occasionally, when trying to repetitively reload the nes hardware driver on SLES11 SP2, a soft lockups occurs that required reboot.	-
4.	In ConnectX-2, if the driver load succeeds, the informative message below is presented conveying the below limitations: • If port type is IB the number of maximum supported VLs is 4 • If port type is ETH then the maximum priority for VLAN tagged is 3 `"mlx4_core 0000:0d:00.0: command SET_PORT (0xc) failed: in_param=0x120064000, in_mod=0x2, op_mod=0x0, fw status = 0x40"`	-
5.	"openibd start" unloads kernel modules that were loaded from initrd/initramfs upon boot. This affects only kernel modules which come with MLNX_OFED and are included in initrd/initramfs.	-

3.13 Performance Tools Known Issues

Table 22 - Performance Tools Known Issues

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

3.14 Performance Known Issues

Table 23 - Performance Known Issues

Index	Description	Workaround
1.	On machines with irqbalancer daemon turned off, the default InfiniBand interrupts will be routed to a single core which may cause overload and software/ hardware lockups.	To avoid this issue copy the following script to `/etc/infiniband/post-start-hook.sh` and execute it as root: `#!/usr/bin/perl` `use strict;` `if ( $< != 0 ) {` `print "This script must be run as root\n";` `exit (0);` `}` `open(F, "/proc/interrupts") or die "$!";` `my $n = `cat /proc/cpuinfo \| grep processor \| wc -l`;` `chomp($n);` `print "Spreading over $n cpus\n";` `while(<F>) {` `#print $_;` `my ($irq,$chan,$dev);` `if (/(\d+):.*mlx5_comp(\d+)/) {` `($irq,$chan,$dev) = ($1,$2,"mlx5");` `} elsif (/(\d+):.*(mlx4-ib-\d)-(\d+)/) {` `($irq,$chan,$dev) = ($1,$3,$2);` `} else {` `next;` `}` `my $place = ($chan % $n);` `my $mask = 1 << $place;` `printf ("$dev irq=%d chan=%d bit=%d mask=%0x\n", $irq, $chan, $place, $mask);` `my $cmd = sprintf("echo %0x > /proc/irq/$irq/smp_affinity", $mask);` `print "\t$cmd\n";` `system($cmd);` `}`

3.15 Connection Manager (CM) Known Issues

Table 24 - Connection Manager (CM) Known Issues

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

3.16 SR-IOV Known Isuues

Table 25 - SR-IOV Known Issues

Index	Description	Workaround
1.	When using legacy VMs with MLNX_OFED 2.x hypervisor, you may need to set the `'enable_64b_cqe_eqe'` parameter to zero on the hypervisor. It should be set in the same way that other module parameters are set for mlx4_core at module load time. For example, add `"options mlx4_core enable_64b_cqe_eqe=0"` as a line in the file `/etc/modprobe.d/mlx4_core.conf`.	-
2.	InfiniBand counters are not available in the VM.	-
3.	`mlx4_port1_mtu` sysfs entry shows a wrong MTU number in the VM.	-
4.	When at least one port is configured as InfiniBand, and the `num_vfs` is provided but the `probe_vf` is not, HCA initialization fails.	Use both the `num_vfs` and the `probe_vf` in the modprobe line.
5.	When working with a bonding device to enslave the Ethernet devices in active-backup mode and failover MAC policy in a Virtual Machine (VM), establishment of RoCE connections may fail.	Unload the module mlx4_ib and reload it in the VM.
6.	Attaching or detaching a Virtual Function on SLES11 SP3 to a guest Virtual Machine while the `mlx4_core` driver is loaded in the Virtual Machine may cause a kernel panic in the hypervisor.	Unload the `mlx4_core` module in the hypervisor before attaching or detaching a function to or from the guest.
7.	When detaching a VF without shutting down the driver from a VM and reattaching it to another VM with the same IP address for the Mellanox NIC, RoCE connections will fail	Shut down the driver in the VM before detaching the VF.
8.	Enabling SR-IOV requires appending the `"intel_iommu=on"` option to the relevant OS in file `/boot/grub/grub.conf`. Without that SR-IOV cannot be loaded.	-
9.	On various combinations of Hypervisor/OSes and Guest/OSes, an issue might occur when attaching/detaching VFs to a guest while that guest is up and running.	Attach/detach VFs to/from a VM only while that VM is down.
10.	When working with SR-IOV in Xen-4.2 virtualization platform, only the built-in xen_pciback driver should be loaded. The xen_pciback module in dom0 should not be loaded, as loading them simultaneously may cause interrupts loss and cause the driver to enter the reset flow.	-

Table 25 - SR-IOV Known Issues (Continued)

Index	Description	Workaround
11.	The known PCI BDFs for all VFs in kernel command line should be specified by adding `xen-pciback.hide` For further information, please refer to http://wiki.xen.org/wiki/Xen_PCI_Passthrough	-
12.	The qemu version (2.0) provided inbox with Ubuntu 14.04 does not work properly when more than 2 VMs are run over an Ubuntu 14.04 Hypervisor.	-
13.	SR-IOV UD QPs are forced by the Hypervisor to use the base GID (i.e., the GID that the VF sees in its GID entry at its paravirtualized index 0). This is needed for security, since UD QPs use Address Vectors, and any GID index may be placed in such a vector, including indices not belonging to that VF.	-
14.	Attempting to attach a PF to a VM when SR-IOV is already enabled on that PF may result in a kernel panic.	-
15.	osmtest on the Hypervisor fails when SR-IOV is enabled. However, only the test fails, OpenSM will operate correctly with the host. The failure reason is that if an mcg is already joined by the host, a subsequent join request for that group succeeds automatically (even if the join parameters in the request are not correct). This success does no harm.	-

3.17 Port Type Management Known Issues

Table 26 - Port Type Management Known Issues

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

3.18 Flow Steering Known Issues

Table 27 - Flow Steering Known Issues

Index	Description	Workaround
1.	Flow Steering is disabled by default in firmware version < 2.32.5100.	To enable it, set the parameter below as follow: `log_num_mgm_entry_size should set to -1`
2.	IPv4 rule with source IP cannot be created in SLES 11.x OSes.	-
3.	RFS does not support UDP.	-

3.19 Quality of Service Known Issues

Table 28 - Quality of Service Known Issues

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

3.20 Installation Known Issues

Table 29 - Installation Known Issues

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

3.21 Driver Upload Known Issues

Table 30 - Driver Upload Known Issues

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

3.22 InfiniBand Counters Known Issues

Table 31 - InfiniBand Counters Known Issues

Index	Description	Workaround
1.	Occasionally, `port_rcv_data` and `port_xmit_data` counters may not function properly.	-

3.23 UEFI Secure Boot Known Issues

Table 32 - UEFI Secure Boot Known Issues

Index	Description	Workaround
1.	On RHEL7 and SLES12, the following error is displayed in dmesg if the Mellanox's x.509 Public Key is not added to the system: `[4671958.383506] Request for unknown module key 'Mellanox Technologies signing key: 61feb074fc7292f958419386ffdd9d5ca999e403' err -11` This error can be safely ignored as long as Secure Boot is disabled on the system.	For further information, please refer to the User Manual section "Enrolling Mellanox's x.509 Public Key On your Systems".
2	Ubuntu12 requires update of user space open-iscsi to v2.0.873	-
3	The initiator does not respect interface parameter while logging in.	Configure each interface on a different subnet.

3.24 Fork Support Known Issues

Table 33 - Fork Support Known Issues

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

3.25 ISCSI over IPoIB Known Issues

Table 34 - ISCSI over IPoIB Known Issues

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

3.26 MLNX_OFED Sources Known Issues

Table 35 - MLNX_OFED Sources Known Issues

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

3.27 InfiniBand Utilities Known Issues

Table 36 - InfiniBand Utilities Known Issues

Index	Description	Workaround
1.	When running the ibdiagnet check `nodes_info` on the fabric, a warning specifying that the card does not support general info capabilities for all the HCAs in the fabric will be displayed.	Run `ibdiagnet --skip nodes_info`

3.28 mlx5 Driver Known Issues

Table 37 - mlx5 Driver Known Issues

Index	Description	Workaround
1.	Atomic Operations in Connect-IB® are fully supported on big-endian machines (e.g. PPC). Their support is limited on little-endian machines (e.g. x86)	-

3.29 Ethernet Performance Counters Known Issues

Table 38 - Ethernet Performance Counters Known Issues

Index	Description	Workaround
1.	In a system with more than 61 VFs, the 62nd VF and onwards is assigned with the SINKQP counter, and as a result will have no statistics, and loopback prevention functionality for SINK counter.	-

Table 38 - Ethernet Performance Counters Known Issues (Continued)

Index	Description	Workaround
2.	Since each VF tries to allocate 2 more QP counter for its RoCE traffic statistics, in a system with less than 61 VFs, if there is free resources it receives new counter otherwise receives the default counter which is shared with Ethernet. In this case RoCE statistics is not available.	-
3.	In ConnectX®-3, when we enable function-based loopback prevention for Ethernet port by default (i.e., based on the QP counter index), the dropped self-loopback packets increase the IfRxErrorFrames/ Octets counters.	-

3.30 Uplinks Known Issues

Table 39 - Uplinks Known Issues

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

3.31 Resources Limitation Known Issues

Table 40 - Resources Limitation Known Issues

Index	Description	Workaround
1.	The device capabilities reported may not be reached as it depends on the system on which the device is installed and whether the resource is allocated in the kernel or the userspace.	-
2.	mlx4_core can allocate up to 64 MSI-X vectors, an MSI-X vector per CPU.	-
3.	Setting more IP addresses than the available GID entries in the table results in failure and the "update_gid_table error message is displayed: `GID table of port 1 is full. Can't add <address>" message.`	-
4.	Registering a large amount of Memory Regions (MR) may fail because of DMA mapping issues on RHEL 7.0.	-
5.	Occasionally, a user process might experience some memory shortage and not function properly due to Linux kernel occupation of the system's free memory for its internal cache.	To free memory to allow it to be allocated in a user process, run the `drop_caches` procedure below. Performing the following steps will cause the kernel to flush and free pages, dentries and inodes caches from memory, causing that memory to become free. **Note:** As this is a non-destructive operation and dirty objects are not freeable, run `` `sync' `` first. • To free the pagecache: `echo 1 > /proc/sys/vm/drop_caches` • To free dentries and inodes: `echo 2 > /proc/sys/vm/drop_caches` • To free pagecache, dentries and inodes: `echo 3 > /proc/sys/vm/drop_caches`

3.32 RoCE Known Issues

Table 41 - RoCE Known Issues

Index	Description	Workaround
1.	Not configuring the Ethernet devices or independent VMs with a unique IP address in the physical port, may result in RoCE GID table corruption.	Restart the driver
2.	If RDMA_CM is not used for connection management, then the source and destination GIDs used to modify a QP or create AH should be of the same type - IPv4 or IPv6.	-

Table 41 - RoCE Known Issues (Continued)

Index	Description	Workaround
3.	On rare occasions, the driver reports a wrong GID table (read from /sys/class/infiniband/mlx4_*/ports/*/gids/*). This may cause communication problems.	-
4.	MLNX_OFED v2.1-1.0.0 and onwards is not interoperable with older versions of MLNX_OFED.	-
5.	Since the number of GIDs per port is limited to 128, there cannot be more than the allowed IP addresses configured to Ethernet devices that are associated with the port. Allowed number is: • "127" for a single function machine • "15" for a hypervisor in a multifunction machine • "(127-15)/n" for a guest in a multifunction machine (where n is the number of virtual functions)	-
6.	A working IP connectivity between the RoCE devices is required when creating an address handle or modifying a QP with an address vector.	-
7.	IPv4 multicast over RoCE requires the MGID format to be as follow ::ffff:<Multicast IPv4 Address>	-
8.	IP routable RoCE does not support Multicast Listener Discovery (MLD) therefore, multicast traffic over IPv6 may not work as expected.	-
9.	Using GID index 0 (the default GID) is possible only if the matching IPv6 link local address is configured on the net device of the port. This behavior is possible even though the default GID is configured regardless the presence of the IPv6 address.	-
10.	Using IPv6 link local address (GID0) when VLANs are configured is not supported.	-

3.33 Storage Known Issues

Table 42 - Storage Known Issues

Index	Description	Workaround
1.	Older versions of `rescan_scsi_bus.sh` may not recognize some newly created LUNs.	If encountering such issues, it is recommended to use the '-c' flag.
2.	RHEL7.0: The `rescan-scsi-bus.sh` script does not rediscover provisioned LUNs both on iSER and SRP.	• Use older version of the script from RHEL6.4 • iSER: Use `"iscsiadm -m session --rescan"`

3.34 SRP Known Issues

Table 43 - SRP Known Issues

Index	Description	Workaround
1.	In a high stress IO with unstable links, SRP Initiator may generate a call trace that can be safely ignored.	-
2.	MLNX_OFED SRP installation breaks the ibmvstgt and ibmvscsi symbol resolution in RHEL7.0	-

3.35 SRP Interop Known Issues

Table 44 - SRP Interop Known Issues

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

3.36 DDN Storage Fusion 10000 Target Known Issues

Table 45 - DDN Storage Fusion 10000 Target Known Issues

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

3.37 Oracle Sun ZFS storage 7420 Known Issues

Table 46 - Oracle Sun ZFS storage 7420 Known Issues

Index	Description	Workaround
1.	Occasionally the first command to a LUN may not be serviced, aborted, and cause a successful re-connection to the target	-
2.	Ungraceful power cycle of an initiator connected with Targets DDN, Nimbus, NetApp may result in temporary "stale connection" messages when initiator reconnects.	-

3.38 iSER Known Issues

Table 47 - iSER Known Issues

Index	Description	Workaround
1.	SM LID reassignmet during traffic, on OEL6.4 uek kernel with a Virtual Function, generates soft lockup trace	-
2.	On SLES OSs, the `ib_iser` module does not load on boot	Add a dummy interface using iscsiadm: • `# iscsiadm -m iface -I ib_iser -o new` • `# iscsiadm -m iface -I ib_iser -o update -n iface.transport_name -v ib_iser`
3	DIF: When running IO over FS over DM during unstable ports, block layer BIOs merges may cause false DIF error.	-
4	Ubuntu12 requires update of user space open-iscsi to v2.0.873	-
5	The initiator does not respect interface parameter while logging in.	Configure each interface on a different subnet.
6	iSCSID v2.0.873 can enter an endless loop on bind error	-
7	DIX: Under heavy IO stress with large block size, the HCA card might generate error Completions in the log and traffic might be affected.	Disable block merges.
8	iSCSID may hang if target crashes during logout sequence (reproducible with TCP)	-

3.39 ZFS Appliance Known Issues

Table 48 - ZFS Appliance Known Issues

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

4 Bug Fixes History

Table 49 lists the bugs fixed in this release.

Table 49 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
1.	IPoIB	Changing the GUID of a specific SR-IOV guest after the driver has been started, causes the ping to fail. Hence, no traffic can go over that InfiniBand interface.	2.1-1.0.0	2.3-1.0.1
2.	Ethernet	Fixed kernel panic on Debian-6.0.7 which occurred when the number of TX channels was set above the default value.	2.1-1.0.0	2.2-1.0.1
3.		Fixed a crash incidence which occurred when enabling Ethernet Time-stamping and running VLAN traffic.	2.0-2.0.5	2.2-1.0.1
4.	XRC	XRC over ROCE in SR-IOV mode is not functional	2.0-3.1.0	2.2-1.0.1
5.	mlx4_en	Fixed wrong calculation of packet true-size reporting in LRO flow.	2.1-1.0.0	2.2-1.0.1
6.	IB Core	Fixed the QP attribute mask upon smac resolving	2.1-1.0.0	2.1-1.0.6
7.	mlx5_ib	Fixed a send WQE overhead issue	2.1-1.0.0	2.1-1.0.6
8.		Fixed a NULL pointer dereference on the debug print	2.1-1.0.0	2.1-1.0.6
9.		Fixed arguments to kzalloc	2.1-1.0.0	2.1-1.0.6
10.	mlx4_core	Fixed the locks around completion handler	2.1-1.0.0	2.1-1.0.6
11.	mlx4_core	Restored port types as they were when recovering from an internal error.	2.0-2.0.5	2.1-1.0.0
12.		Added an N/A port type to support port_type_array module param in an HCA with a single port	2.0-2.0.5	2.1-1.0.0
13.	SR-IOV	Fixed memory leak in SR-IOV flow.	2.0-2.0.5	2.0-3.0.0
14.		Fixed communication channel being stuck	2.0-2.0.5	2.0-3.0.0

Table 49 - Fixed Bugs List

#	Issue	Description	Discovered in Release	Fixed in Release
15.	mlx4_en	Fixed ALB bonding mode failure when enslaving Mellanox interfaces	2.0-3.0.0	2.1-1.0.0
16.		Fixed leak of mapped memory	2.0-3.0.0	2.1-1.0.0
17.		Fixed TX timeout in Ethernet driver.	2.0-2.0.5	2.0-3.0.0
18.		Fixed ethtool stats report for Virtual Functions.	2.0-2.0.5	2.0-3.0.0
19.		Fixed an issue of VLAN traffic over Virtual Machine in paravirtualized mode.	2.0-2.0.5	2.0-3.0.0
20.		Fixed ethtool operation crash while interface down.	2.0-2.0.5	2.0-3.0.0
21.	IPoIB	Fixed memory leak in Connected mode.	2.0-2.0.5	2.0-3.0.0
22.		Fixed an issue causing IPoIB to avoid pkey value 0 for child interfaces.	2.0-2.0.5	2.0-3.0.0

5 Change Log History

Table 50 - Change Log History

Release	Category	Description
Rev 2.2-1.0.1	mlnxofedinstall	32-bit libraries are no longer installed by default on 64-bit OS. To install 32-bit libraries use the `'--with-32bit'` installation parameter.
	openibd	Added pre/post start/stop scripts support. For further information, please refer to section *"openibd Script"* in the MLNX_OFED User Manual.
	Reset Flow	Reset Flow is not activated by default. It is controlled by the mlx4_core`'internal_err_reset'` module parameter.
	InfiniBand Core	Asymmetric MSI-X vectors allocation for the SR-IOV hypervisor and guest instead of allocating 4 default MSI-X vectors. The maximum number of MSI-X vectors is `num_cpu` for port ConnectX®-3 has 1024 MSI-X vectors, 28 MSI-X vectors are reserved. • Physical Function - gets the number of MSI-X vectors according to the `pf_msix_table_size` (multiple of 4 - 1) INI parameter • Virtual Functions – the remaining MSI-X vectors are spread equally between all VFs, according to the `num_vfs mlx4_core` module parameter
	Ethernet	Ethernet VXLAN support for kernels 3.12.10 or higher
		Power Management Quality of Service: when the traffic is active, the Power Management QoS is enabled by disabling the CPU states for maximum performance.
		Ethernet PTP Hardware Clock support on kernels/OSes that support it
	Verbs	Added additional experimental verbs interface. This interface exposes new features which are not integrated yet in to the upstream libibverbs. The Experimental API is an extended API therefor, it is backward compatible, meaning old application are not required to be recompiled to use MLNX-OFED v2.2-1.0.1.
	Performance	Out of the box performance improvements: • Use of affinity hints (based on NUMA node of the device) to indicate the IRQ balancer daemon on the optimal IRQ affinity • Improvement in buffers allocation schema (based on the hint above) • Improvement in the adaptive interrupt moderation algorithm

Table 50 - Change Log History

Release	Category	Description
Rev 2.1-1.0.6	IB Core	Added allocation success verification process to ib_alloc_device.
	dapl	dapl is recompiled with no FCA support.
	openibd	Added the ability to bring up child interfaces even if the parent's ifcfg file is not configured.
	libmlx4	Unmapped the hca_clock_page parameter from mlx4_uninit_context.
	scsi_transport_srp	scsi_transport_srp cannot be cleared up when rport reconnecting fails.
	mlnxofedinstall	Added support for the following parameters: • '--umad-dev-na' • '--without-<package>'
	Content Packages Updates	The following packages were updated: • bupc to v2.2-407 • mstflint to v3.5.0-1.1.g76e4acf • perftest to v2.0-0.76.gbf9a463 • hcoll to v2.0.472-1 • Openmpi to v1.6.5-440ad47 • dapl to v2.0.40
Rev 2.1-1.0.0	EoIB	EoIB is supported only in SLES11SP2 and RHEL6.4.
	eIPoIB	eIPoIB is currently at GA level.
	Connect-IB®	Added the ability to resize CQs.
	IPoIB	Reusing DMA mapped SKB buffers: Performance improvements when IOMMU is enabled.
	mlnx_en	Added reporting autonegotiation support.
		Added Transmit Packet Steering (XPS) support.
		Added reporting 56Gbit/s link speed support.
		Added Low Latency Socket (LLS) support.
		Added check for dma_mapping errors.
	eIPoIB	Added non-virtual environment support.

Table 50 - Change Log History

Release	Category	Description
Rev 2.0-3.0.0	Operating Systems	Additional OS support: • SLES11SP3 • Fedora16, Fedora17
	Drivers	Added Connect-IB™ support
	Installation	Added ability to install MLNX_OFED with SR-IOV support.
		Added Yum installation support
	EoIB	EoIB (at beta level) is supported only in SLES11SP2 and RHEL6.4
	mlx4_core	Modified module parameters to associate configuration values with specific PCI devices identified by their bus/device/function value format
	mlx4_en	Reusing DMA mapped buffers: major performance improvements when IOMMU is enabled
		Added Port level QoS support
	IPoIB	Reduced memory consumption
		Limited the number TX and RX queues to 16
		Default IPoIB mode is set to work in Datagram, except for Connect-IB™ adapter card which uses IPoIB with Connected mode as default.
	Storage	iSER (at GA level)
Rev 2.0-2.0.5[a]	Virtualization	SR-IOV for both Ethernet and InfiniBand (at Beta level)
	Ethernet Network	RoCE over SR-IOV (at Beta level)
		eIPoIB to enable IPoIB in a Para-Virtualized environment (at Alpha level)
		Ethernet Performance Enhancements (NUMA related and others) for 10G and 40G
		Ethernet Time Stamping (at Beta level)
		Flow Steering for Ethernet and InfiniBand. (at Beta level)
		Raw Eth QPs: • Checksum TX/RX • Flow Steering
	InfiniBand Network	Contiguous pages: • Internal memory allocation improvements • Register shared memory • Control objects (QPs, CQs)
	Installation	YUM update support
	VMA	OFED_VMA integration to a single branch

Table 50 - Change Log History

Release	Category	Description
	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. SR-IOV, Ethernet Time Stamping and Flow Steering are ConnectX®-3 HCA capability.

6 API Change Log History

Table 51 - API Change Log History

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

Table 51 - API Change Log History

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

Table 51 - API Change Log History

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