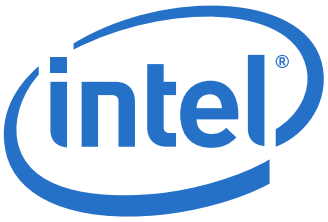

True Scale Fabric OFED+ Host Software

Release Notes

February 2014

Order Number: H27570002US

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

Any software source code reprinted in this document is furnished for informational purposes only and may only be used or copied and no license, express or implied, by estoppel or otherwise, to any of the reprinted source code is granted by this document.

Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as the property of others.

Copyright © 2014, Intel Corporation. All rights reserved.

Contents

1.0 Overview of the Release 5
1.1 Introduction 5
1.2 Audience 5
1.3 If You Need Help 5
1.4 New Features and Enhancements 5
1.4.1 Release 7.2 Features 5
1.4.2 Release 7.1.1 Enhancements 6
1.4.3 Release 7.1 Features 6
1.4.4 Release 7.0.1 Features 7
1.5 Operating Environments Supported 8
1.6 Qualified Parallel File Systems 9
1.7 Intel Interface for NVIDIA GPUs 9
1.8 Hardware Supported 10
1.9 Installation Requirements 10
1.9.1 Software and Firmware Requirements 10
1.9.2 Installation Instructions 10
1.10 Changes for this Release 11
1.10.1 Changes to Hardware Support 11
1.10.2 Changes to Operating System Support 12
1.10.3 Changes to Software Components 13
1.10.4 Changes to Industry Standards Compliance 13
1.11 Product Constraints 13
1.12 Product Limitations 14
1.13 Other Information 15
1.14 Documentation 17
2.0 System Issues for Release 7.2 19
2.1 Introduction 19
2.2 Resolved Issues in this Release 19
2.3 Known Issues 21
2.3.1 Severity 21
2.3.2 Open Issues Table 21
A Performance Gain Conditions Test 25

Tables

1 Operating Environments Supported 8
2 CPU Model of Linux Kernel 8
3 NVIDIA's CUDA Tested with OFED+ 9
4 Hardware Supported 10
5 Changes to Hardware Support 11
6 Changes to Operating System Support 12
7 Changes to Software Component Support 13
8 Changes to Industry Standards Compliance 13
9 Related Documentation for this Release 17
10 Resolved Issues 19
11 Open Issues 21

1.0 Overview of the Release

1.1 Introduction

These Release Notes provide a brief overview of the changes introduced into the Intel® True Scale Fabric OFED+ by this release. References to more detailed information are provided where necessary. The information contained in this document is intended for supplemental use only; it should be used in conjunction with the documentation provided for each component.

These Release Notes list the new features of the release, as well as the system issues that were closed in the development of Release 7.2.1.1.22.

1.2 Audience

The information provided in this document is intended for installers, software support engineers, and service personnel.

1.3 If You Need Help

If you need assistance while working with the OFED+ Host Software, contact your Intel® approved reseller or Intel® True Scale Technical Support:

- By E-mail:
  ibsupport@intel.com
- On the Support tab at web site:
  http://www.intel.com/infiniband

For OEM-specific server platforms supported by this release, contact your OEM.

1.4 New Features and Enhancements

The new features and enhancements added since Release 7.2 and the two previous major/minor releases for the OFED+ Host Software are listed below.

1.4.1 Release 7.2.1.1.22 Enhancements

- Added support for RedHat EL 6.4 and SLES 11 SP3
- Added support for servers with Ivy-Bridge CPUs

1.4.2 Release 7.2 Features

- PSM can support multiple rails connected to the same fabric or different fabrics (or planes). In addition PSM can now support striping across the rails for a single process or MPI rank allowing a single process to benefit from the aggregate bandwidth from two separate HCAs. The environment variables PSM_MULTIRAIL and PSM_MULTIRAIL_MAP enable this feature. More details can be found in the *Intel® True Scale Fabric OFED+ Host Software User Guide*.
- This release includes fixes in performance-related issues for IPoIB and verbs which provide significant improvement in IPoIB bandwidth and latency and improvement in RDMA/verbs bandwidth.
- Added support for RedHat EL 6.3 and RedHat EL 5.9

- qlgc_srp and qlgc_vnic packages are dropped from version 7.2 onwards. On new installation or upgrades of version 7.2, these packages will be removed from the host.
- Branding for Intel has been incorporated in this release. The following packages are included in the branding effort:
  - OFED+ IB-Basic
  - OFED_MPIs
  - OPENIB_ROLL

1.4.3 Release 7.1.1 Enhancements

- The "Optimal Assignment of PSM Processes to HCAs" enhancement added in this release automatically improves MPI/PSM performance to the greatest extent in systems where the following conditions are in effect:
  1. The OS and CPUs support NUMA (Non-Uniform Memory Architecture) and NUMA node to I/O device binding.
  2. Two HCAs connect to different PCIe root complexes which, in turn, connect to different NUMA nodes.

  Typical systems where these conditions hold are those with two or more Intel® Xeon® Processor E5 2600 series CPUs (known as Sandy Bridge) with dual HCAs in slots connected to the PCIe root complexes in different CPUs, running RHEL 6.1, or SLES 11 SP1 or newer operating systems.These enhancements will also offer some additional performance improvement on systems with one HCA.

  There are systems based on other CPUs that have not been tested to verify the performance gain, where conditions 1 and 2 hold. Systems where the OS does not support conditions 1 and 2 will still get a performance benefit from two HCAs versus one HCA.

  To determine if conditions 1 and 2 hold in a system refer to Appendix A

1.4.4 Release 7.1 Features

- `rnfs-utils` command was removed by Open Fabrics and is not included in OFED 1.5.4.1.
- iscsi and iSER Target packages were removed by Open Fabrics and are not included in OFED 1.5.4.1. When installing Intel® OFED+, any previous versions of these packages already on the system will not be affected. Intel® recommends to uninstall these packages using the `iba_config` script or ./INSTALL TUI from the previous version prior to installing the new version. One exception is `scsi-target-utils`, which will be removed if found on the system.
- MPI is no longer installed or included in the release. If found on the system when installing or un-installing the Intel® libraries, it will be removed. Several MPIs including Intel MPI, Open MPI, MVAPICH, and MVAPICH2 continue to be supported with this release. To aid in the transition please consult the examples in the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

- Shared Memory (SHMEM) is included in the ./INSTALL TUI menus as a selection. SHMEM will be installed by default and will also be installed if `mpi` or `psm_mpi` or `mpidev` is selected on the command line. To function, SHMEM requires that at least one MPI be installed on the system. This requirement of one MPI being installed is not enforced by ./INSTALL TUI.

  SHMEM is a user-level communications library for one-sided operations. It implements the SHMEM Application Programming Interface (API) and runs on the Intel® InfiniBand* (IB) stack. The SHMEM API provides global distributed shared memory across a network of hosts.

  SHMEM is quite distinct from local shared memory (often abbreviated as "shm" or even "shmem"). Local shared memory is the sharing of memory by processes on the same host running the same OS system image. SHMEM provides access to global shared memory distributed across a cluster. The SHMEM API is completely different from and unrelated to the standard System V Shared Memory API provided by UNIX operating systems.
- The questions for autostart enable of components have been replaced with an interactive menu showing all the autostart selections in both the ./INSTALL TUI and the `iba_config` script.
- The version required for the PGI compiler has been upgraded from "9.0.4 or later" to "10.5 or later." The upgrade to 10.5 is required for proper operation of MVAPICH2 version 1.7.
- NVIDIA CUDA 4.0 and 4.1 have been tested and are supported.

1.4.5 Release 7.0.1 Features

- The `iba_manage_switch` script, along with the xedge tools, is included as part of IB-Basic (OFED+) allowing customers not using the IFS software to manage externally managed switches. Unlike FastFabric, it is designed to operate on one switch at a time, taking a mandatory target GUID parameter.

  Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.
- Performance tuning parameters for the QLE7340 and QLE7342 drivers can now be set on a per port or per unit basis. If there are two HCAs (HCA) in a server, settings for one HCA can be optimized for storage traffic and settings for the other HCA can be optimized for MPI traffic. Please see the *Intel® True Scale Fabric OFED+ Host Software User Guide* for details.

1.5 Operating Environments Supported

The Release 7.2.1.1.22 version of OFED+ Host Software allows for the Operating Systems listed in Table 1.

Table 1. Operating Environments Supported

Operating System	Update/ SP	Version
Red Hat Enterprise Linux (RHEL) 5 X86_64 (AMD Opteron and Intel EM64T)	Update 8	2.6.18-308.el5.x86_64
	Update 9	2.6.18-348.el5.x86_64
RHEL 6 X86_64 (AMD Opteron and Intel EM64T)	Update 2	2.6.32-220.el6.x86_64
	Update 3	2.6.32-279.el6.x86_64
	Update 4	2.6.32-358.el6.x86_64
SLES 11 X86_64 (AMD Opteron and Intel EM64T)	SP2	3.0.13-0.27-default
	SP3	3.0.76-0.11-default
Community Enterprise Operating System (CentOS) X86_64 (AMD Opteron and Intel EM64T)	Update 5.8	2.6.18-308.el5.x86_64
	Update 5.9	2.6.18-348.el5.x86_64
Community Enterprise Operating System (CentOS) X86_64 (AMD Opteron and Intel EM64T)	Update 6.2	2.6.32-220.el6.x86_64
	Update 6.3	2.6.32-279.el6.x86_64
	Update 6.4	2.6.32-358.el6.x86_64
Scientific Linux X86_64	Update 5.8	2.6.18-308.1.1.el5.x86_64
	Update 5.9	2.6.18-348.el5.x86_64
Scientific Linux X86_64	Update 6.2	2.6.32-220.el6.x86_64
	Update 6.3	2.6.32-279.el6.x86_64
	Update 6.4	2.6.32-358.el6.x86_64
StackIQ Cluster Manager (Rocks+) HPC 6.1	RHEL 6.3	2.6.32-279.el6.x86_64
	CentOS 6.3	2.6.32-279.el6.x86_64
Platform HPC-4.1.1	RHEL 6.4	2.6.32-358.el6.x86_64

CPU model of Linux kernel can be identified by `uname -m` and `/proc/cpuinfo` shown in Table 2

Table 2. CPU Model of Linux Kernel

Model	uname	/proc/cpuinfo
EM64T	x86_64	Intel CPUs
Opteron*	x86_64	AMD CPUs

Note: Other combinations (such as i586 uname) are not currently supported.

1.6 Qualified Parallel File Systems

Lustre and IBM General Parallel File System (GPFS) listed below have been tested for use with this release of the Intel® OFED+ host software using the operating systems listed below:

- Lustre 2.3
  - RHEL 6.3
- Lustre 2.4.1
  - RHEL 6.4
- IBM GPFS 3.4
  - RHEL 6.1
  - SLES 11 SP1
- IBM GPFS 3.5
  - RHEL 6.3
  - SLES 11 SP2
- IBM GPFS 3.5.0.12
  - RHEL 6.4
  - SLES 11 SP3

Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for the latest configuration recommendations for optimizing Lustre and GPFS performance with Intel® True Scale Fabric.

1.7 Intel Interface for NVIDIA GPUs

NVIDIA's CUDA parallel computing platform and programing models have been tested for use with this release of the Intel® OFED+ host software using the operating systems listed in Table 3:

Table 3. NVIDIA's CUDA Tested with OFED+

Distributions	CUDA 5.5
RHEL 6.4	X
SLES 11 SP3	X

1.8 Hardware Supported

Table 4 list the hardware supported in this release.

Table 4. Hardware Supported

HCAs
QLE7340
QLE7342
QME7342
QME7362
QMH7342
MHQH29-*
MHQH19-*
MHQH19B-XTR
MHQH29B-XTR
MHQH29B-XSR
MCX354A-QCAT
MCX353A-QCAT
NC543i (HP SL390 G7 in-built InfiniBand Host Channel Adapter)
CX-3 LOM down QDR
46M2199
46M2203

1.9 Installation Requirements

1.9.1 Software and Firmware Requirements

All Intel IB software on a given node must be at a compatible release level. Each distribution of the Intel IB Installation Wrapper will have a qualified and compatible version of each Intel IB Software component. Prior to installing Intel IB Software, any versions of the Silverstorm IB stack (and any other vendor's IB stack) must be uninstalled.

1.9.2 Installation Instructions

Note: An upgrade from IntelIB-7.1 or later may be performed in which case the installation will detect the existing installation and properly upgrade existing installed components and shall be removing components which are no longer supported.

Note: Any versions of an older IB stack must be uninstalled first. If you are downgrading from a newer IntelIB release to an older IntelIB or InfiniServ release, the newer IntelIB release must be stopped and uninstalled, prior to installing the older release. The newer IntelIB must be uninstalled using the `iba_config` tool or the `./INSTALL` tool provided with the newer IntelIB release.

Note: The installation process attempts to uninstall any existing 3rd party versions of OFED, however some packagings of OFED may not be completely uninstalled. If using a 3rd party OFED installation, it is recommended to uninstall it prior to installing IntelIB.

Note: The installation process attempts to uninstall any existing distribution versions of OFED, however some rpms included in the distribution packaging of OFED may not be completely uninstalled. It is recommended to uninstall any OFED rpms which come with the distribution prior to installing IntelIB.

Note: FastFabric may be used to install the IntelIB-Basic package on all nodes in the cluster.

Refer to the Intel® True Scale Fabric Software Installation Guide for installation information and procedures.

1.10 Changes for this Release

The following sections describe the changes that have been made to the Intel® OFED+ Host software package between versions 7.2.0.0.42 and 7.2.1.1.22, including the following releases:

- 7.1.0.0.58
- 7.1.1.0.25
- 7.2.0.0.42

For detailed information about any of the previous releases listed, refer to the Release Notes for the specific version.

1.10.1 Changes to Hardware Support

Table 5 shows the new hardware supported for the releases listed.

Table 5. Changes to Hardware Support

Release	Supported Hardware Added
7.1.0.0.58	None
7.1.1.0.25	None
7.2.0.0.42	None
7.2.1.1.22	None

1.10.2 Changes to Operating System Support

Table 6 shows the new operating systems supported for the releases listed.

Table 6. Changes to Operating System Support

Release	Supported Operating System Added
7.1.0.0.58	RHEL 6 X86_64 (AMD Opteron and Intel EM64T): • (Update 2) 2.6.32-220.el6.x86_64 CentOS X86_64 (AMD Opteron and Intel EM64T): • (Update 5.7) 2.6.18-274.el5 • (Update 6.1) 2.6.32-131.0.15.el6.x86_64 • (Update 6.2) 2.6.32-220.el6.x86_64 Scientific Linux X86_64: • (Update 5.7) 2.6.18-274.el5 Rocks 6.0.1: • (RHEL 6.2) 2.6.32-220.el6.x86_64 • (CentOS 6.0) 2.6.32-71.el6.x86_64 Platform Cluster Manager 3.0.1 Standard Edition: • (RHEL 5.7) 2.6.32-131.0.15.el6.x86_64 Platform Cluster Manager 3.1 Standard Edition: • (RHEL 6.2) 2.6.32-220.el6.x86_64
7.1.1.0.25	RHEL 5 X86_64 (AMD Opteron and Intel EM64T): • (Update 8) 2.6.18-308.el5 SLES 11 X86_64 (AMD Opteron and Intel EM64T) • (SP2) 3.0.13-0.27-default Platform Cluster Manager 3.1 Dell Edition: • (RHEL 6.2) 2.6.32-220.el6.x86_64
7.2.0.0.42	RHEL 5 X86_64 (AMD Opteron and Intel EM64T): • (Update 9) 2.6.18-348.el5.x86_64 RHEL 6 X86_64 (AMD Opteron and Intel EM64T): • (Update 3) 2.6.32-279.el6.x86_64 CentOS X86_64 (AMD Opteron and Intel EM64T): • (Update 5.8) 2.6.18-308.el5.x86_64 • (Update 6.3) 2.6.32-279.el6.x86_64 Scientific Linux X86_64: • (Update 5.8) 2.6.18-308.1.1.el5.x86_64 • (Update 6.2) 2.6.32-220.el6.x86_64 Rocks+ 6.0.2: • (RHEL 6.2) 2.6.32-220.el6.x86_64 • (CentOS 6.2) 2.6.32-220.el6.x86_64 Rocks+ HPC 3.0: • (RHEL 6.3) 2.6.32-220.el6.x86_64 • (CentOS 6.3) 2.6.32-279.el6.x86_64
7.2.1.1.22	RHEL 6 X86_64 (AMD Opteron and Intel EM64T): • (Update 4) 2.6.32-358.el6.x86_64 SLES 11 X86_64 (AMD Opteron and Intel EM64T) • (SP3) 3.0.76-0.11-default

1.10.3 Changes to Software Components

Table 7 shows the new software components supported for the releases listed.

Table 7. Changes to Software Component Support

Release	Supported Software Added or Changed
7.1.0.0.58	Intel® OFED+ Software Intel® IB Tools Software
7.1.1.0.25	Intel® OFED+ Software Intel® IB Tools Software
7.2.0.0.42	Intel® OFED+ Software
7.2.1.1.22	Intel® OFED+ Software

1.10.4 Changes to Industry Standards Compliance

Table 8 shows each Basic OFED version that is supported and the Intel® OFED+ Releases that include each

Table 8. Changes to Industry Standards Compliance

Basic OFED Software Package Supported	Intel® OFED+ Host Software Package
Version 1.5.4.1	Version 7.1.0.0.58, 7.1.1.0.25, 7.2.0.0.42, and 7.2.1.1.22

1.11 Product Constraints

The following is a list of product constraints for this release:

- The libgoto BLAS library included in the MPI sample applications for use by HPL is no longer supported by the developer and does not support arbitrary combinations of OS and CPU types.
  - — Mixing AMD and Intel CPUs is not supported
  - — AMD "Bulldozer" CPUs are not supported
  - — Mixing different Linux distributions may not work reliably.

  An alternative to libgoto is the ATLAS library, which must be manually compiled for each desired CPU type and Linux distribution. A sample version of ATLAS has been provided in the same directory as the other sample MPI applications, and the latest version may be found online at http://www.netlib.org/atlas/

  Examples of using the ATLAS library can be found in the hpl and hpl-2.0 directories with the other sample applications."

- The version of Open MPI shipped with the Intel® True Scale Fabric Suite Software is incompatible with the Performance Application Programming Interface (“papi”) libraries optionally available in Red Hat Enterprise Linux version 6. If you have installed the optional papi RPMs and try to use FastFabric to recompile Open MPI on RHEL 6 you will first have to uninstall any installed version of papi-4.x. Older versions of papi (papi-3.x) are still compatible with the shipped version of Open MPI.
- All installation and uninstallation of Intel® OFED+ Host software package components must be performed using the `./INSTALL` or `iba_config` commands. If software is manually installed or uninstalled using other methods (RPM, other scripts, and so on), the installation on the system could become inconsistent and cause unreliable operation, in which case subsequent runs of `./INSTALL` or `iba_config` may make incorrect conclusions about the configuration of the system and consequently make incorrect recommendations. If the system becomes inconsistently configured, Intel recommends running the `./INSTALL` TUI and selecting `ReInstall` on all components. Once the re-installation has started, carefully review all prompts and choices.
- OFED SDP has not been qualified for this release. IPoIB is recommended for data transfers.

1.12 Product Limitations

The following is a list of product limitations for this release:

- Intel products will auto-negotiate with devices that utilize IBTA-compliant auto-negotiation. When attaching Intel products to a third-party device, the bit error rate is optimized if the third-party device utilizes attenuation-based tuning.

1.13 Other Information

The following is a list of "need to know" information for this release:

- The Dell PowerEdge M1000e Blade System has been updated with a new backplane (version 1.1), which requires different QME734x transmitter tuning settings. In order to facilitate proper transmitter settings, a new module parameter `qme_bp` has been added to the QIB driver. The `qme_bp` parameter should be set by the user to one of two values for the version of the installed backplane. A value of `0` means the backplane is version 1.0 and the default value of `1` means the backplane is version 1.1.

  The module parameter can be set by editing the `/etc/modprobe.d/ib_qib.conf` file for nodes installed with RHEL operational environment or `/etc/modprobe.conf.local` file for nodes installed with SLES operational environment The string "qme_bp=value" needs to be added to the "options ib_qib ..." option line. The value contained in this option, in combination with the chassis slot number, is used by the driver and support script to select the correct settings and program the QME734x transmitter.

  To determine the version of the Dell backplane use the following procedure:

  1. Login to the CMC of the Dell PowerEdge M1000e Blade System
  2. Type "getsysinfo"
  3. Under the Chassis Information, look for Chassis Midplane Version, refer to the following example:.

```
              Output from getsysinfo:


Chassis Information:

System Model              = PowerEdgeM1000

System AssetTag           = 123abc

Service Tag               =

Chassis Name              = dell-cmc

Chassis Location          = [UNDEFINED]

Chassis Midplane Version  = 1.1    ←

Power Status              = ON
```

- For information on Oracle* Remote Data Service (RDS) support, refer to RAC Technologies Matrix for Linux Platforms
- The OpenSHMEM effort (see http://www.openshmem.org) is defining a standardized API specification for SHMEM. Although it is premature to claim compliance, Intel® SHMEM aims to be compliant with the OpenSHMEM 1.0 specification. Intel provides a SHMEM API that is compatible with the OpenSHMEM 1.0 specification, other than any omissions or bugs documented in these release notes. If compliance with OpenSHMEM's passive progress statements are required, Intel® SHMEM's passive progress mechanism must be enabled.

- When using Mellanox HCAs, any changes to Virtual Fabrics (vFabrics) in the Fabric Manager, may require a reboot of the hosts with Mellanox HCAs. This limitation relates to the Mellanox HCAs not properly responding to changes to the Fabric Manager service level (SL). For some vFabric configuration changes, if the Fabric Manager SL changes or is mapped to a different Virtual Lane (VL) than previously, the Mellanox HCA can continue to use the previous VL. If that VL is presently disabled by the Fabric Manager, future uses of applications which use the Fabric Manager SL may hang or timeout because there are no VL Arbitration cycles for that VL. As a result, anytime vFabric configuration is changed, it is recommended to reboot all hosts with Mellanox HCAs so that the desired Quality of Service (QoS) configuration changes fully take effect. Any hosts with Intel® HCAs will not need to be rebooted.

  Due to Mellanox HCAs not correctly handling changes to the Fabric Manager SL, Intel recommends that all the hosts using Mellanox ConnectX or ConnectX-2 HCAs be rebooted when used in a virtual fabric configuration.
- When Dispersive Routing is enabled, it allows packets sent using an MPI program run over PSM to take any one of several routes through a fabric, thus often increasing performance. The number of routes is determined by the value of 2 to the power of the Lid Mask Control setting (LMC). Because LMC defaults to 0, the default number of routes through the fabric is 2^0 or 1. LMC can be set as high as 3, allowing a total number of 2^3 or 8 routes through the fabric. Providing these additional routes can reduce fabric congestion, and thus improve performance. Dispersive Routing is supported when the Fabric Manager is used in the fabric. Dispersive Routing is not supported when using OpenSM.
- When running MVAPICH2, Intel recommends turning off RDMA fast path. To turn off RDMA fast path, specify `MV2_USE_RDMA_FAST_PATH=0` in the `mpirun_rsh` command line or set this option in the parameter file for mvapich2.
- The `ib_send_bw` benchmark, when run in UC mode, is written such that it will hang if even one packet is dropped.

1.14 Documentation

Table 9 lists the Release 7.2 related documentation. All related documentation is available on the Intel® download site.

Documentation for Intel® Partners is available at the vendors web site.

Table 9. Related Documentation for this Release

Document Title	Document Number	Revision
Intel® Hardware Documents		
Intel® True Scale Fabric Switches 12000 Series Hardware Installation Guide	G91928	002US
Intel® True Scale Fabric Switches 12000 Series Users Guide	G91930	002US
Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide	G91931	002US
Intel® True Scale Fabric Adapter Hardware Installation Guide	G91929	002US
Intel® OFED+ Documents		
Intel® True Scale Fabric Software Installation Guide	G91921	002US
Intel® True Scale Fabric OFED+ Host Software User Guide	G91902	002US
Intel® *True Scale Fabric OFED+ Host Software Release Notes*	H27570	002US
Intel® IFS Documents		
Intel® True Scale Fabric Suite FastFabric User Guide	G91916	002US
Intel® True Scale Fabric Suite Fabric Manager User Guide	G91918	002US

Table 9. Related Documentation for this Release (Continued)

Document Title	Document Number	Revision
Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide	G91904	002US
Intel® True Scale Fabric Suite Software Release Notes	H27571	002US
Intel® Fabric Viewer Documents		
Intel® *True Scale Fabric Suite Fabric Viewer Online Help*	N/A	N/A
Intel® True Scale Fabric Suite Fabric Viewer Release Notes	G91934	002US

2.0 System Issues for Release 7.2

2.1 Introduction

This section provides a list of the resolved Issues in the OFED+ Host Software that were verified by this release. It also lists the open Issues with a description and workaround for each.

2.2 Resolved Issues in this Release

Table 10 is a list of issues that are resolved in this and the previous two releases.

Table 10. Resolved Issues

Product	Release	Description
IFS/ Install/Uninstall	7.1.0.0.58	When installing the Intel-Basic (Intel OFED+) or Intel-IFS SW on SLES10 SP3, there may be conflicts with software that is already installed on the system. SLES10 SP3 is no longer a supported distribution, therefore all SLES10 SP3 known issues are closed.
IB Third Party/ Platform HPC/LSF	7.1.0.0.58	LSF job submission now works correctly with Platform HPC
IB Third Party/ Platform HPC/LSF	7.1.0.0.58	The **Command Task** drop-down in the Platform HPC GUI now shows all of the options properly.
IB Third Party/ Platform HPC/LSF	7.1.0.0.58	The FastFabric and Fabric Viewer options are now shown on the Platform HPC GUI with RHEL 5.6 and SLES 11 SP1.
Intel IB HCA/ Driver	7.1.0.0.58	In a dual-HCA system, if cfgctxts is set in `modprobe.conf` only for HCA unit 1, and there is no default value explicitly entered, then HCA unit 0 will no longer have 0 available contexts and MPI can run.
Intel IB HCA/ Driver	7.1.0.0.58	RHEL6.1 distribution of IB software (included with the OS distribution) no longer has a problem causing page faults or kernel panics on greater than 16 core servers when a Intel IB HCA is installed with an IB cable connected to the fabric.
Intel IB HCA/ Driver	7.1.0.0.58	When the per-unit `singleport` setting in `modprobe.conf` is specified for only qib1, with no default explicitly stated, qib0 no longer acts as if `singleport` is set.
Intel IB HCA/ Driver	7.1.0.0.58	When the per-unit `pcie_coalesce` setting in `modprobe.conf` is specified for only qib1, with no default explicitly stated, then `qib0` no longer acts as if `pcie_coalesce` is set on it as well.
TrueScale/ Tools	7.2.0.0.42	On SLES11SP2 systems, the `ipathstats -c <interval>` command now shows valid data.
IFS/ IBAccess	7.2.0.0.42	For SLES10 and 11, the `--32bit` installation option is no longer used.
IFS/ MPI2	7.2.0.0.42	When uninstalling MVAPICH2 (for verbs or PSM), some files under the `/usr/mpi/*/mvapich2*/` directory tree that are created at runtime by MVAPICH2 may not be removed. The uninstall program is designed not to delete these files.

Table 10. Resolved Issues (Continued)

Product	Release	Description
IFS/ Other	7.2.0.0.42	After installing IFS on a Lustre1.8.5 patched kernel, there can be a lot of messages in `dmesg` from `ib_iser` complaining about Unknown symbol. Lustre 1.8.5 is no longer supported, and this issue is no longer seen with the newer Lustre versions.
IFS/ Open SM	7.2.0.0.42	When reinstalling Intel OFED+ and the IB Subnet Manager is not running a message shows as follows: `Stopping IB Subnet Manager [FAILED]` This failed to stop because it was not running. This is normal process for an installation to stop installed applications that are running.
IB Third Party/ OFED	7.2.0.0.42	`rdma_bw` requires the use of the -f option (don't fail even if `cpufreq_ondemand` module is loaded) when being used for testing.
IFS/ HCA	7.2.0.0.42	IPoIB connections no longer become hung (`ib0: transmit timed out`) under certain stressful conditions.
True Scale/ Tools	7.2.1.1.22	IB services no longer fail to start on a Dell M420 blade.
True Scale PSM	7.2.1.1.22	The PTRANS test of the HPC Challenge benchmark no longer hangs.
True Scale PSM	7.2.1.1.22	PSM now works properly for intra-node kcopys with message lengths greater than 2 GB.
True Scale Driver	7.2.1.1.22	The True Scale driver no longer causes a deadlock related to mmap_sem locks and a copy from userspace.

2.3 Known Issues

The subsections below catalog the known open issues for the release as well as a description and a workaround by component.

2.3.1 Severity

This document provides a level of severity for each issue listed The levels are:

- **Critical** – Could result in a service outage
- **Major** – Could degrade system performance
- **Minor** – Could cause minimal impact to ongoing operations
- **None** – No operational impact

2.3.2 Open Issues Table

Table 11 is the list of open issues for Release 7.2. The table is sorted by Severity then Product.

Table 11. Open Issues

Product/ Component	Severity	Description	Workaround

IFS/ Rolls/Kits	Minor	In Release 7.2.1 of IFS and OFED kit, Platform HPC 4.1.1 GUI shows Version as 1.5.4.1 and release version is no longer displayed.	kit name (.bz2) file shows right version after OFED version (1.5.4.1) as 7.2.1.1.19.
			
			`kit-intel_ifs-1.5.4.1-7.2.1.1.19-rhels-6-x86_64.tar.bz2`
			
			`kit-intel_ofed-1.5.4.1-7.2.1.1.19-rhels-6-x86_64.tar.bz2`

IFS/ Install/Uninstall	Major	The SRP Target contained in IFS for release 7.2 nodes cannot be installed on nodes running RHEL6.x.	Download and Install the required SRP drivers from http://sourceforge.net/projects/scst/files/srpt on designated SRP target node
IFS/ MPI	Major	If `LD_LIBRARY_PATH` is exported inconsistently with the version of Open MPI being used, applications may build or run incorrectly. This issue can impact FastFabric tools that use MPI, rebuilding of mpi apps, or rebuilding Open MPI itself using the `do_build` or `do_openmpi_build` tools.	When using Open MPI, make sure `PATH` and `LD_LIBRARY_PATH` are not exported specifying a different path than the Open MPI path that is being used. The mpi-selector can configure a `LD_LIBRARY_PATH` for subsequent logins. Open MPI does not require the `LD_LIBRARY_PATH` to be set.

Table 11. Open Issues (Continued)

Product/ Component	Severity	Description	Workaround

IFS/ MPI	Major	MVAPICH2 jobs run between PCIe HCAs and Third Party or Intel HCAs, may not complete successfully. The test may abort with an `ibv_post_recv` error.	Any MVAPICH2 job attempted on fabrics with PCIe HCAs and Third Party or Intel HCAs, must zero the MV2_USE_SRQ environment variable as show in the example of the NAS CG benchmark:
			
			`cd directory_containing_benchmark`
			
			`/usr/mpi/gcc/mvapich2-1.7/bin/mpirun_rsh -np 4 -hostfile mpi_hosts \`
			
			`MV2_USE_RDMA_FAST_PATH=0 MV2_USE_SRQ=0 ./cg.B.4`

IB Third Party/ Other	Major	Some applications using Platform MPI 8.1 or 8.2 over PSM exhibit poor scalability or worse performance at larger core counts.	Use the Platform MPI `mpirun` command with the `-intra=nic` option. This means that PSM's shared memory communications will be used instead of Platform MPI's native shared memory communications.

Intel IB HCA/ SHMEM	Minor	SHMEM collective calls using PE subsets can hang. This problem does not affect collective calls using the entire PE set, which is the more common case. A collective call specifies the entire PE set when the parameters are set as follows:
		
		`PE_start = 0`
		
		`logPE_stride = 0`
		
		`PE_size = num_pes()`
			None

Intel IB HCA/ SHMEM	Minor	The SHMEM API reductions on complex types are not implemented:
		
		`shmem_complexd_sum_to_all`
		
		`shmem_complexf_sum_to_all`
		
		`shmem_complexd_prod_to_all`
		
		`shmem_complexf_prod_to_all`
			None

Table 11. Open Issues (Continued)

Product/ Component	Severity	Description	Workaround

IFS/ Install/Uninstall	Minor	When installing the Intel-Basic (Intel OFED+) or Intel-IFS SW on SLES11SP1, there may be conflicts with software that is already installed on the system. The following message may appear:
		
		`error: %preun(ofed-1.4.2-0.9.6.x86_64) scriptlet failed, exit status 1`
		
		`error: %preun(opensm-3.2.6_20090317-0.1.42.x86_64)scriptlet failed, exit status 1`
		
		`Unable to uninstall previous OFED RPMs`
			Manually uninstall the old version OFED before you install the newer software.
			1. As root, run the following command:
			
			`rpm -e --noscripts ofed`
			
			2. Re-run the normal installation.

IFS/ IBAccess	Minor	When a port is down and does not have a LID assigned, `clear_p1stats` or `clear_p2stats` will fail against the given port	None
IFS/ IBAccess	Minor	When using vFabric, the OFED `saquery` command may use the wrong P-Key and timeout waiting for responses.	Intel recommends using the `iba_saquery` tool, which is included with IntelIB-Basic or IntelIB-IFS. `iba_saquery` will work properly when vFabric is configured.
IFS/ Open SM	Minor	When using opensm, after bouncing ports on a node, the port may not return to an active state for a period of time. As a result, commands that issue an SA query such as OFED's `saquery` command, or various FastFabric tools such as `iba_report` and `iba_saquery`, may hang waiting for the port to become active and the SA to respond.	Restart opensm. Intel recommends using the Intel® Fabric Manager, which has much greater resiliency and quicker handling of port state changes.
IFS/ IPoIB	Minor	When using vFabric to change an IPoIB application from Networking to Non-Networking, the IPoIB interface may remain in a running state.	After changing the application, restart the network services or bring the interface down/up to force IPoIB to re-query the SM and correct the situation.
IFS/ MPI2	Minor	When trying to rebuild mvapich2-1.6-qlc (PSM) with PGI 11.7, it fails with the following error message: `configure: error: cannot run C compiled programs`	Add the following line to the login scripts, in the environment such that is it available for both interactive and non-interactive logins. `export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$PGI/linux86-64/11.7/libso` After making the change to the login scripts, exit and log back into the server so its defined and then run the `do_*_build` script.
IFS/ Rolls/Kits	Minor	When using the Platform HPC and installing the Intel OFED+ or IFS kits, some messages appear in the system log file regarding unknown symbol and symbol version disagreement. These are due to the way and order in which the Intel OFED+ or IFS kits are installed and are resolved by the time OFED is started.	None

Table 11. Open Issues (Continued)

Product/ Component	Severity	Description	Workaround
IFS/ Rolls/Kits	Minor	When the IFS is uninstalled and then reinstalled, some messages appear in Platform HPC's CFM log about issues with dependency resolution. These issues do not affect the operation of any IFS utilities including FastFabric and Fabric Manager.	None

IB Third Party/ Other	Minor	When installing Moab, the following error is seen:
		`[nsgib103 .ssh (Thu May 12 05:43:36)]# ldconfig`
		`ldconfig: /usr/local/lib/libsqlite3.so.0 is not a symbolic link`
			Move/Delete `libsqlite3.so.0` files and execute `ldconfig` command.
			`ldconfig` can create symbolic link properly and the error message will not appear.

IFS/ Rolls/Kits	Minor	If the IFS kit is already installed and running the `updatenode` command on the Installer and/or compute node (`updatenode<headnode/computenode>` command) emits errors similar to the following will be displayed.
		`compute000: Error: Package: opensm-devel-3.3.13-1.x86_64 (installed)`
		`compute000:Requires: opensm-libs = 3.3.13-1`
		`compute000:Removing: opensm-libs-3.3.13-1.x86_64 (installed)`
		`compute000: opensm-libs = 3.3.13-1`
		`compute000: Updated By: opensm-libs-3.3.15-1.el6.x86_64 (xCAT-rhels6.4-path0)`
			These errors may be safely ignored.

IFS/ FastFabric	Minor	Result of iba_verifynodes for C-states can be misleading on SLES 11; as SLES11 does not relies on testing method used in said script. SLES11 uses sysfs interface as oppose to /proc interface. This causes misleading results for C-states. iba_verifynodes shows result as "SKIP" for C-states.	The user is advised to manually check C-States of the sytems by using either of the following workarounds: 1) BIOS settings 2) Check if `/sys/devices/system/cpu/cpu*/cpuidle` exists. If it does then C-states are "Enabled" If not C-states are disabled.

§ §

Appendix A Performance Gain Conditions Test

The following example shows how to determine if conditions 1 and 2, described in the first bullet of "Release 7.1.1 Enhancements" on page 6 hold:

```
$ numactl --hardware

available: 2 nodes (0-1)

node 0 cpus: 0 1 2 3 4 5 6 7

...

node 1 cpus: 8 9 10 11 12 13 14 15

...
```

If `numactl --hardware` shows more than 1 NUMA node, then your OS supports NUMA.

To see whether your system supports NUMA node to IO device binding and whether your HCAs connect to different NUMA nodes, look at the files in the `/sys` directories to see if the **numa_node** field is populated correctly. The following steps indicate how to do this.

1. Change directory to `/sys/class/infiniband`

```
$ cd /sys/class/infiniband
```

2. List all files in the `/infiniband` directory in long format:

```
$ ls -la
```

   This list the symbolic links to the HCA devices with the pci bus, slot and function number (in the following example 6:00.0 and 82:00.0):

```
lrwxrwxrwx  1 root root 0 Jul  9 11:24 qib0 -> ../../devices/pci0000:00/
0000:00:03.0/0000:06:00.0/infiniband/qib0/

lrwxrwxrwx  1 root root 0 Jul  9 11:24 qib1 -> ../../devices/pci0000:80/
0000:80:02.0/0000:82:00.0/infiniband/qib1/
```

3. Print the numa node id for the respective devices:

```
[infiniband]$ cat ../../devices/pci0000:00/0000:00:03.0/0000:06:00.0/numa_node

0

[infiniband]$ cat ../../devices/pci0000:80/0000:80:02.0/0000:82:00.0/numa_node

1
```

   The HCAs are bound to the two NUMA nodes first shown in this appendix with the 'numactl --hardware" command: 0 and 1.

§ §