Intel® True Scale Fabric OFED+ Host Software

Release Notes

February 2018

Contents

Tables

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 7.4.3.0.6.

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:
  fabricsupport@intel.com
- On the Support tab at website:
  Fabric Products

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

1.4 New Features and Enhancements

This section lists the new features and enhancements for the this release, as well as the
previous release.

1.4.1 Release 7.4.3.0.6 Enhancements

- Added support for
  — RHEL 6.9
  — CentOS 6.9
  — Scientific Linux 6.9
  — OFED 3.18-3

1.4.2 Release 7.4.2.0.6 Enhancements

- Added support for
  - RHEL 6.8
  - CentOS 6.8
  - Scientific Linux 6.8
  - OFED-3.18-2

1.5 Operating Environments Supported

The following Operating Systems are supported in this release:

Table 1-1.   Operating Environments Supported

Operating System	Update/SP	Version
RHEL 6 X86_64 (AMD Opteron and Intel EM64T)	Update 8	2.6.32-642.el6.x86_64
	Update 9	2.6.32-696.el6.x86_64
RHEL 7 X86_64 (AMD Opteron and Intel EM64T)	Update 1	3.10.0-229.el7.x86_64
	Update 2	3.10.0-327.el7.x86_64
SLES 11 X86_64 (AMD Opteron and Intel EM64T)	SP3	3.0.76-0.11-default
	SP4	3.0.101-0.63.1-default
SLES 12 X86_64 (AMD Opteron and Intel EM64T)		3.12.28-4.x86_64
	Update 1	3.12.49.11.1.x86_64
Community Enterprise Operating System (CentOS) X86_64 (AMD Opteron and Intel EM64T) (6.x)	Update 8	2.6.32-642.el6.x86_64
	Update 9	2.6.32-696.el6.x86_64
Scientific Linux X86_64 (6.x)	Update 8	2.6.32-573.el6.x86_64
	Update 9	2.6.32-696.el6.x86_64

CPU model of Linux kernel can be identified by `uname -m` and `/proc/cpuinfo` shown in Table 1-2.

Table 1-2.   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.
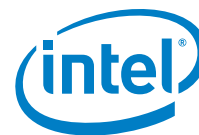
1.6 Qualified Parallel File Systems

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* 2.8
    - RHEL 6.7 (both server and client)
    - RHEL 7.2 (both server and client)
    - SLES11 SP4 (client)
- IBM GPFS 4.2.3
    - RHEL 6.8
    - SLES 12.1

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 1-3:

Table 1-3. NVIDIA's CUDA Tested with IFS

Distributions	CUDA 7
RHEL 6.8	X
RHEL 6.9	X
RHEL 7.1, 7.2	X
SLES 11.4	X

1.8 Compilers Supported

1.8.1 MPI

This release supports the following MPI implementations:

Table 1-4. MPI Compilers

MPI Implementation	Runs Over	Compiled With
Open MPI 1.10.7	PSM Verbs	GCC, Intel, PGI
MVAPICH2 version 2.1.1	PSM Verbs	GCC, Intel, PGI
Platform MPI 9.1	PSM Verbs	GCC (default)
Intel MPI version 2017.4.196	TMI/PSM, uDAPL	icc 17.0.4.196

1.8.2 MVAPICH2 and Open MPI

MVAPICH2 and Open MPI have been have been compiled for PSM to support the following versions of the compilers:

Table 1-5. MVAPICH2 and Open MPI

Compiler name	Distro	Compiler Version
(GNU) gcc	RHEL6	gcc (GCC) 4.4.7 20120313 (Red Hat 4.4.7-17)
(GNU) gcc	RHEL7	gcc (GCC) 4.8.5 20150623 (Red Hat 4.8.5-4)
(GNU) gcc	SLES11	gcc (SUSE Linux) 4.3.4 [gcc-4_3-branch revision 152973]
(PGI) pgcc	RHEL6	pgcc 2017-174 64-bit target on x86-64 Linux -tp nehalem-64
(PGI) pgcc	RHEL7	pgcc 2017-174 64-bit target on x86-64 Linux -tp sandybridge
(PGI) pgcc	SLES12	pgcc 2017-174 64-bit target on x86-64 Linux -tp nehalem-64
(Intel) icc	RHEL6	icc 17.0.4.196
(Intel) icc	RHEL7	icc 17.0.4.196
(Intel) icc	SLES11.4	icc 17.0.4.196
(Intel) icc	SLES12.1	icc 17.0.4.196

1.9 Hardware Supported

Table 1-6 list the hardware supported in this release.

Table 1-6. 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

1.10 Installation Requirements

1.10.1 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.

1.10.2 Installation Instructions

Note: An upgrade from Intel® True Scale Fabric OFED+ Host Software 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 remove 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 third-party versions of OFED, however some packagings of OFED may not be completely uninstalled. If using a

third-party OFED installation, Intel recommends uninstalling 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. Intel recommends uninstalling 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.

1.11 Changes for this Release

The following sections describe the changes that have been made to the Intel® True Scale Fabric Suite Software package for the current release.

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

1.11.1 Changes to Industry Standards Compliance

Table 1-7 shows the Basic OFED versions and the Intel® OFED+ Releases that support each one.

Table 1-7. Changes to Industry Standards Compliance

Basic OFED Software Package Supported	Intel® OFED+ Host Software Package
Version 1.5.4.1	Versions 7.2.1.1.22, 7.2.2.0.8
Version 3.5-2	Version 7.3.0.0.26, 7.3.1.0.12
Version 3-12.1	Version 7.4.0.0.21
Version 3-18.1	Version 7.4.1.0.24
Version 3-18.2	Version 7.4.2.0.6
Version 3.18-3	Version 7.4.3.0.6

1.12 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://math-atlas.sourceforge.net/

  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")
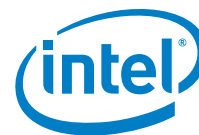
libraries optionally available in Red Hat Enterprise Linux version 6.x. If you have installed the optional papi RPMs and try to use FastFabric to recompile Open MPI on RHEL 6.x, you will first have to uninstall any installed version of papi if it is greater than version 3.x. Older versions of papi (for example, papi-3.x) are still compatible with the shipped version of OpenMPI

- 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.

1.13 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.

1.14 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. Log in 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.

1.14.1 Verbs Performance

Whenever possible, Intel recommends installing at least one True Scale HCA on a PCIe bus connected to the CPU 1 Socket for optimal verbs performance. For details on finding the correct PCIe slot, consult the documentation for your server chassis or motherboard.

1.15 Documentation

Table 1-8 lists the end-user documentation for the current release. All related documentation is available on the Intel download site.

Documentation for Intel® Partners is available at the vendors website.

Table 1-8.   Related Documentation for this Release

Document Title
Intel® Hardware Documents
Intel® True Scale Fabric Switches 12000 Series Hardware Installation Guide
Intel® True Scale Fabric Switches 12000 Series Users Guide
Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide
Intel® True Scale Fabric Adapter Hardware Installation Guide
Intel® OFED+ Documents
Intel® True Scale Fabric Software Installation Guide
Intel® True Scale Fabric OFED+ Host Software User Guide
Intel® True Scale Fabric OFED+ Host Software Release Notes
Intel® IFS Documents
Intel® True Scale Fabric Suite FastFabric User Guide
Intel® True Scale Fabric Suite Fabric Manager User Guide
Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide
Intel® True Scale Fabric Suite Software Release Notes
Intel® Fabric Viewer Documents
Intel® True Scale Fabric Suite Fabric Viewer Online Help
Intel® True Scale Fabric Suite Fabric Viewer Release Notes
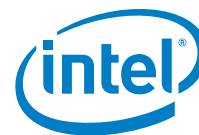
2.0 System Issues for Release 7.4.3.0.6

2.1 Introduction

This section provides a list of the resolved and open issues in the OFED+ Host Software that were verified by this release. The list of open issues includes a description and workaround.

2.2 Resolved Issues in this Release

Table 2-1 is a list of issues that are resolved in this release.

Table 2-1. Resolved Issues

Product/ Component/ PR Number	Release	Description
IFS/MPI 131369	7.4.3	For dual-rail systems, by default HCA2, port 1 is chosen for running traffic. If this port is disabled, mvapich2-2.1 fails to find any other active port. This causes mpirun to fail with error(s).
IFS 133007	7.4.3	Running the installation script with the parameter `--with-xeon-phi` provides XEON PHI support within the IFS installation process. The parameter `--with-xeon-phi=<path>` accepts the path of where to look for `mpss-psm*` and `mpss-psm-dev*` packages which should be installed. If the parameter will be added without any path, the `– installer` script looks for `mpss-psm` packages within the default directories such as **/root**, **/tmp**, **/opt** or **/home** and automatically installs it.
True Scale/PSM 137021	7.4.3	When an app is linked to **libpsm_infinipath.so**, signals like **SIGSEGV** are not passed to the parent.
True Scale/IFS 134031	7.4.2	In order to realize good intra-node performance, the kcopy module must be installed. SLES blocks the module from being installed.

2.3 Known Issues

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

2.3.1 Open Issues Table

Table 2-2 lists the open issues for this release.

Table 2-2. Open Issues

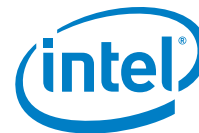| Product/Component/PR Number	Description	Workaround
IFS/Fabric Manager 111338	When the **LogFile** parameter is in use, the Fabric Manager outputs to the named file instead of **syslog**. If a high LogLevel is selected, the log file can grow quickly and consume too much disk space.	Limit use of **LogFile** to short duration debug type operations and use **syslog** for normal Fabric Manager operations.
IFS/Rolls/Kits 123572	If the IFS kit is already installed, then running the **updatenode** command on the Installer and/or compute node (**updatenode <headnode/compuetnode>** command) returns errors similar to the following: **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 124980	If the **libedit** rpm is not present on a host, it should be installed prior to installing IFS. If the **libedit** libraries are not installed, the following error is displayed. Additionally, commands such as **iba_manage_switch** and **iba_switch_admin** would also fail with an error message similar to: **#/opt/iba/ib_tools/L8simlnx** **/opt/iba/ib_tools/L8simlnx: error while loading shared libraries: libedit.so.0: cannot open shared object file: No such file or directory** If the **libedit** rpm is not installed prior to the IFS installation, the capture operation would fail. However, installing the **libedit** rpm would fix the issue. A reinstall of IFS is not needed.	If not already present on the host, install the **libedit** rpm.
IFS/Install/Uninstall 131005	SLES 12 systems have problems with modules not loading at boot.	On SLES12.x, set the **allow_unsupported_modul es** parameter to **1** in the file **/etc/modprobe.d/10-unsupp ortedmodules. conf**.

Table 2-2. Open Issues (Continued)

Product/ Component/PR Number	Description	Workaround
True Scale/Driver 131486	**rdma-ndd** is a background service; it updates the node description as **hostname devicename**. However, the service **openibd** already sets the appropriate node description as **Hostname HCAn**.	Since **rdma-ndd** overwrites the correct node description, it needs to be disabled in order to expect correct behavior by the **openibd** service.
IFS/Install/Uninstall 131780	Parallel studio 2016 is not supported in this release.	Continue using Parallel studio 2015.
IFS/Install/Uninstall 131946	While installing **IntelIBIFS.<osid>.7.4.1.0.*** on RHEL and SLES systems, non-standard OFED RPMs may need to be removed before running the **INSTALL** script.	Some examples are: • qemu-kvm • gnome-boxes • libvirt-daemon-kvm • glusterfs-rdma • boost-devel • libboost_graph_parallell1_54_0 • libboost_mpi1_54_0 • libibnetdisc5
IFS/Install/Uninstall 131977	• OFED-3.18-1 requires kernel-source package to be "installed" for SLES12SP1. • OFED-3.18-1 fails to start drivers if it does not find the kernel-source package during installation time. This is a requirement in order to use OFED-3.18-1 or IFS 7.4.1 on SLES12SP1.	None

IFS/FastFabric 131984	For a SLES12 host, if **openibd** is disabled using **iba_config**, (**autostart** none) option, after the system is rebooted **systemctl** would report **openibd** status as masked:
	CMZ-DEV-114:~ # systemctl status
	openibd
	openibd.service
	    Loaded: masked (/dev/null)
	    Active: inactive (dead)
		To re-enable it and reboot the system, run the **iba_config** option. If **openibd** needs to be started manually, the user needs to execute the **unmask** and **enable** options of **systemctl** to start **openibd**.
		Commands to manually start **openibd** would be
		/usr/bin/systemctl unmask openibd
		/usr/bin/systemctl enable openibd
		/usr/bin/systemctl start openibd

True Scale/ MPI 135513	Mvapich2 compiled with icc or pgi can fail when using **mpirun_rsh**.	Use the SSH-based version – **mpirun**.
Third Party/RHEL/OFED 139608 141418	The OFED IPoIB driver from the 4.8 kernel is not fully-compatible with the bonding drivers distributed in SLES 12.1 and 12.2 (causing kernel-panic), as well as RHEL 7.4 (bonding fails to activate).	None.
True Scale/MPI 142604	Building mvapich2 with PGI fails.	For RHEL6.9 and RHEL6.8, install binutils version 2.22 (or newer) to rebuild PGI MPI.

§ §

Appendix A Performance Gain Conditions Test

The example shows how to determine if the following conditions are true:

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.

```
$ 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.

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

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

4. 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/
```

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

```
[infiniband]$ cat ../../devices/pci0000:00/0000:00:03.0/
0000:06:00.0/numa_node
```
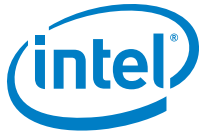
```
0
```

```
[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.

<div align="center">

§ §

</div>