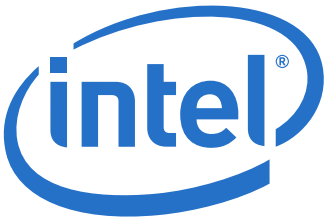


True Scale Fabric Software

Installation Guide

July 2015

Order Number: G91921005US

Contents

Figures

Tables

Revision History

Date	Revision	Description
May 2013	001US	Initial release
October 2013	002US	Updated Platform HPC kit version information Update mpirun information in "LSF Script Administration" on page 140
January 2014	003US	Updated the supported Linux Operating Systems for Fabric Viewer in the section "System Requirements for a Linux* Environment" on page 101.
August 2014	004US	Updated the Support link in Section 1.5.1, "Technical Support" on page 15.
July 2015	005US	Added Appendix F, "Installing IEEL on top of IFS"

§ §

1.0 Introduction

The installation of the Intel® True Scale Fabric Suite (IFS) Software is accomplished using a Text User Interface (TUI) to guide the user through the installation progress. Users also have the ability to use a CLI command to accomplish the installation. Refer to "Overview" for an overview of the software installation packages, and "Installation Recommendations" on page 14 for the installation recommendations.

1.1 Target Audience

This guide is intended to provide network administrators and other qualified personnel a reference for installation and configuration of the Intel® True Scale Fabric OFED+ and True Scale Fabric Suite Software.

1.2 Overview

The following software installation packages are available for an Intel fabric:

- Intel® OFED+ Host Software – This is the basic installation package that installs the OFED+ Host Software components that are needed to set up and control a fabric. The OFED+ Host Software installation is provided with the Intel True Scale Fabric products.
- Intel® True Scale Fabric Suite – The IFS installation package is a set of software products that provide value added features that include the OFED+ Host Software installation package, along with the True Scale Fabric Suite FastFabric Toolset (FF) and the True Scale Fabric Suite Fabric Manager (FM). The IFS is a licensed software package.
- Intel® SHMEM – The SHMEM is part of the OFED+ Host Software and True Scale Fabric Suite Software installation packages. The SHMEM is installed using one of these installation packages and selecting the SHMEM to be installed with that package.
- Rocks Roll for Intel® OFED+ – The Rocks Rol for OFED+ software installation package is provided with the True Scale Fabric products and contains the software components as the OFED+ Host Software installation package contains with the Rocks Roll installation mechanism.
- PCM Kit for Intel® OFED+ – The PCM Kit for OFED+ Software is provided with the Intel® True Scale Fabric Software products and contains the same software components as the OFED+ Host Software with the Platform HPC installation mechanism.
- Intel® True Scale Fabric Suite Fabric Viewer – The True Scale Fabric Suite Fabric Viewer is a licensed software product that provides a number of value added features for viewing the fabric or multiple fabrics.

The following sections discuss the installation packages.

1.2.1 Intel® OFED+ Host Software

The OFED+ Host Software (`IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz`) installation package installs the following components:

- OFED IB Stack
- True Scale HCA Libs
- OFED mlx4 Driver
- IB Tools

- OFED IB Development
- OFED IP over IB
- OFED IB Bonding

Note: OFED IB Bonding will show as not available in the installation menu, when installing the software on OSs that have bonding modules in the OS installed software.

- OFED SDP
- MVAPICH for gcc
- MVAPICH2 for gcc
- OpenMPI for gcc
- MVAPICH/PSM for gcc
- MVAPICH/PSM for PGI
- MVAPICH/PSM for Intel
- MVAPICH2/PSM for gcc
- MVAPICH2/PSM for PGI
- MVAPICH2/PSM for Intel
- OpenMPI/PSM for gcc
- OpenMPI/PSM for PGI
- OpenMPI/PSM for Intel
- SHMEM
- MPI Source
- OFED uDAPL
- OFED RDS
- OFED SRP
- OFED SRP Target
- OFED iSER
- OFED iWARP
- OFED Open SM
- OFED NFS over RDMA
  (Available as a technology preview)
- OFED Debug Info

Note: There is a separate OFED+ Host Software installation package for each of the supported Linux* distributions. Refer to the release notes of the package version being installed for a list of supported Linux* distributions.

1.2.1.1 Installation Prerequisites

In addition to the normal operating system (OS) installation options, the following OS Red Hat* Package Manager (RPMs) must be installed before installing the Intel® OFED+ Host Software.

- pmtools (on SLES)
- dmidecode (on RHEL)
- tcl
- tcl-devel

- pciutils-devel
- binutils-devel
- tk
- libstdc++
- libgfortran
- sysfsutils
- zlib-devel

Depending on which packages you choose, there may be additional prerequisites. Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for additional information.

1.2.2 Intel True Scale Fabric Suite Software

The True Scale Fabric Suite Software (`IntelIB-IFS.`*`DISTRO`*`.`*`VERSION`*`.tgz`) installation package installs the OFED+ Host Software installation package listed in Section 1.2.1 plus:

- FastFabric
- IFS FM

For details on using the True Scale Fabric Suite FastFabric Toolset, refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*. For details on using the True Scale Fabric Suite Fabric Manager (IFS FM), refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide*.

Note: There is a separate IFS installation package for each of the supported Linux* distributions. Refer to the release notes for the version being installed for a list of supported Linux* distributions.

1.2.2.1 Installation Prerequisites

In addition to normal OS installation options, the following OS RPMs must be installed before installing the True Scale Fabric Suite Software installation.

- pmtools (on SLES)
- dmidecode (on RHEL)
- tcl
- tcl-devel
- pciutils-devel
- binutils-devel
- tk
- libstdc++
- libgfortran
- expect
- sysfsutils
- zlib-devel

Depending on which packages you choose, there may be additional prerequisites. Refer to the *Intel® True Scale Fabric Suite Software Release Notes* for additional information.

1.2.3 Intel SHMEM

The OFED+ Host Software (`IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz`) installation package or the IFS software (`IntelIB-IFS.`*`DISTRO`*`.`*`VERSION`*`.tgz`) installation package installs SHMEM when SHMEM is selected to be installed during the installation.

1.2.4 Rocks Roll for Intel OFED+

The Rocks Rol for OFED+ (`intel_ofed-`*`VERSION`*`.x86_64.disk1.iso`) installation package installs the same components as the OFED+ Host Software. Refer to "Intel® OFED+ Host Software" on page 11.

1.2.5 PCM Kit for Intel® OFED+

The Platform HPC Kit for OFED+ Software (`kit-intel_ofed-DISTRO-VERSION.x86_64.iso` for Platform HPC 3.x and `kit-intel_ofed-VERSION-DISTRO-x86_64.tar.bz2` for Platform HPC 4.1.1.1) installation package installs the same components as the OFED+ Host Software. Refer to "Intel® OFED+ Host Software" on page 11.

1.2.6 True Scale Fabric Suite Fabric Viewer

The True Scale Fabric Suite Fabric Viewer (FV) installation package installs the FV application for monitoring the fabric.

1.3 Installation Recommendations

The installation procedures takes you through a step-by-step process to install/upgrade, configure, and verify the Intel® software for a cluster, fabric, multi-fabric, or single node.

Table 1 lists the installation scenarios and the installation procedures that are recommended for each.

Table 1. Installation Recommendations By Type of installation

Type of Installation	Recommended Procedures
Install True Scale Fabric Software on a fabric (Fabric management nodes and Fabric compute nodes).	1. Fabric Management Node Installation 2. Configure Intel Chassis 3. Install and Configure the Fabric Manager 4. Configure Firmware on the Externally Managed Intel Switches 5. Install OFED+ Host Software on the Remaining Servers 6. Verify OFED+ Host Software on the Remaining Servers 7. Installation of additional Fabric Management Nodes 8. Configure and Initialize Health Check Tools 9. Running High Performance Linpack
Install True Scale Fabric Suite Software on a Fabric management node (and configure the switches and chassis).	1. Fabric Management Node Installation 2. Configure Intel Chassis 3. Install and Configure the Fabric Manager 4. Configure Firmware on the Externally Managed Intel Switches 5. Installation of additional Fabric Management Nodes 6. Configure and Initialize Health Check Tools 7. Running High Performance Linpack
Install OFED+ Host Software on a single True Scale node.	Install OFED+ Host Software
Install SHMEM	Install OFED+ Host Software or Install the True Scale Fabric Suite Software

Table 1. Installation Recommendations By Type of installation (Continued)

Type of Installation	Recommended Procedures
Install OFED+ Host Software using Rocks Roll.	Install Front-end and Compute Nodes or Rocks Installation on an Existing Frontend Node
Install OFED+ Host Software using Platform HPC Kit.	Install Intel Software Using the Platform Cluster Manager Kit
Install True Scale Fabric Software on multi-subnet fabrics.	Multi-Subnet Fabrics
Install True Scale Fabric Suite Fabric Viewer	Install True Scale Fabric Suite Fabric Viewer
Upgrade Intel® True Scale Fabric Software software on Fabric management and compute nodes.	1. Upgrade the Management Node 2. Upgrade the Fabric
Upgrade True Scale Fabric Suite Software on a Fabric management node.	Upgrade the Management Node
Upgrade OFED+ Host Software on a single True Scale Fabric node.	Upgrade the Fabric
Install a previous version of software.	Install a Previous Version of Software
Install Lustre Software.	Install Lustre Software

1.4 Supported Languages

English only.

1.5 Additional Information

This section lists additional IFS and OFED+ Host Software related information that will help you use it appropriately.

1.5.1 Technical Support

Intel True Scale Technical Support for products under warranty is available during local standard working hours excluding Intel Observed Holidays. For customers with extended service, consult your plan for available hours. For Support information, see the Support link at www.intel.com/truescale.

1.5.2 Related Materials

The following is a list of the related documentation

- *Intel® True Scale Fabric OFED+ Host Software User Guide*
- *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 Fabric Viewer Online Help*
- *Intel® True Scale Fabric OFED+ Host Software Release Notes*
- *Intel® True Scale Fabric Suite Software Release Notes*
- *Intel® True Scale Fabric Suite Fabric Viewer Release Notes*
- *Installing Platform HPC* guide (for Platform HPC installation only)

- *Installing Fully Automated Platform HPC* guide (for Platform HPC Dell Edition installation only)

1.5.3 Documentation Conventions

This guide uses the following documentation conventions:

- *Note:* provides additional information.
- ***Caution:*** indicates the presence of a hazard that has the potential of causing damage to data or equipment.
- ***Warning:*** indicates the presence of a hazard that has the potential of causing personal injury.
- Text in blue font indicates a hyperlink (jump) to a figure, table, section in this guide, or links to Web sites. For example:
  - Table 9 lists problems related to the user interface and remote agent.
  - See "Installation Checklist" on page 3-6.
  - For more information, visit www.Intel.com.
- Text in **bold** font indicates user interface elements such as a menu items, buttons, check boxes, or column headings. For example:
  - Click the **Start** button, select **Programs**, select **Accessories**, and click **Command Prompt**.
  - Under **Notification Options**, select the **Warning Alarms** check box.
- Text in `Courier` font indicates a file name, directory path, or command line text. For example:
  - To return to the root directory from anywhere in the file structure: Type `cd /root` and press ENTER.
  - Enter the following command: `sh ./install.bin`
- Key names and key strokes are indicated with UPPERCASE:
  - Press CTRL+P.
  - Press the UP ARROW key.
- Text in *italics* indicates terms, emphasis, variables, or document titles. For example:
  - For a complete listing of license agreements, refer to the *Intel® Software End User License Agreement*.
  - What are *shortcut keys*?
  - To enter the date type *mm/dd/yyyy* (where *mm* is the month, *dd* is the day, and *yyyy* is the year).
- Topic titles between quotation marks identify related topics either within this manual or in the online help, which is also referred to as *the help system* throughout this document.

1.5.4 License Agreements

Refer to the *Intel® Software End User License Agreement* for a complete listing of all license agreements affecting this product.

§ §

2.0 Fabric Software Pre-Installation

This section provides the information and procedures needed prior to installing, configuring, and verifying the fabric software. The site implementation engineer must perform the tasks described in this section to ensure that the fabric is ready for the software installation. To aid in keeping track of steps performed for the installation, a checklist is provided in Appendix A that can be copied to use online and/or printed.

The procedures will be marked with one of the following qualifications when required:

- **(Linux) -** Tasks are only applicable when Linux* is being used.
- **(Host) -** Tasks are only applicable when OFED+, the Intel packaging of OFED, or IFS is being used on the hosts.
- **(Switch) -** Tasks which are applicable only when Intel Switches or Intel Chassis, are being used.
- **(All) -** Tasks are generally applicable to all environments.

Note: Some of the Linux* steps may be applicable to other Unix-like operating systems if it is required to enable use of non-True Scale specific FastFabric tools (such as `cmdall`) against the given hosts.

2.1 Installation Prerequisites

2.1.1 Design of the Fabric

Prior to the installation and setup of the fabric, it is important that the design and installation of the hardware be planned carefully. The design plan must include the following information:

- Identification of servers that will function as the administration or Fabric management nodes where the IFS will be installed.
- Server memory requirements based on the software being used:
  - IFS including the True Scale Fabric Suite Fabric Manager, is required to have at least 500 MB of physical memory for each True Scale Fabric Suite Fabric Manager instance. When managing a cluster of 500 nodes or more, 1GB of memory per True Scale Fabric Suite Fabric Manager instance is required.
  - When running multiple True Scale Fabric Suite Fabric Manager instances on a single management node, the physical memory requirements should be multiplied by the number of True Scale Fabric Suite Fabric Manager instances.
- Swap disk space should follow recommendations for the given version of Linux. Swap space should be twice the size of the physical memory on the server running the True Scale Fabric Suite Fabric Manager.
- Ensure at least one central processing unit (CPU) core is available per True Scale Fabric Suite Fabric Manager instance. For example, four True Scale Fabric Suite Fabric Manager instances on a single management node, would require four CPU cores.

2.1.2 Set Up the Fabric

The following steps provide the information to set up the fabric. For information about the configuration files used by FastFabric, refer to Appendix B.

1. **(All)** Ensure all hardware is installed:
   - Servers
   - Core and edge True Scale Fabric switches.

Note: When installing externally managed switches such as the Intel 12200 switch without a management module, the Node GUID is required. The Node GUID can be found on a label on the case of the switch and will be needed to configure and manage the switches with the IFS.

2. **(All)** Ensure an HCA is installed in each server. Refer to the *Intel® True Scale Fabric Adapter Hardware Installation Guide* for instructions.
3. **(All)** The hardware configuration should be reviewed to ensure everything has been installed according to plan. After the software installation, True Scale Fabric Suite FastFabric tools may be used to help verify the installation.
4. **(Linux)** Ensure the required Operating System (OS) version (with the same kernel version) is installed on all hosts. The Fabric Management node(s) (the hosts that will run FastFabric) should have a full install and must include the Tcl and expect OS RPMs.

   For MPI clusters, install the C and Fortran compilers along with their associated tools on the Fabric Management nodes.

Note: Refer To the *Intel® True Scale Fabric Suite Software Release Notes* for a list of supported OS versions.

5. **(Linux)** Enable remote login as root on each host. In order for True Scale Fabric Suite FastFabric to manage the hosts, the Fabric Management Node must be able to securely log in as root to each host. This can be accomplished using ssh.

Note: To simplify the use of FastFabric to setup root access on the nodes in the fabric using ssh, The same root password must be set on all nodes in the fabric. After root access through ssh has been set up using FastFabric, the administrator can change the root passwords.

6. **(All)** Resolve the TCP/IP Host Names.

   FastFabric and TCP/IP must resolve host names to the Management Network and/or IPoIB IP addresses. If the management network is not IPoIB, each host must have both a management network name and an IPoIB network name. To do this, use the actual host name as the management network name and *HOSTNAME*`-ib` as the IPoIB network name, where *HOSTNAME* is the management network name of the given host.

   Name resolution is accomplished by configuring a DNS server on the management network with both management network and IPoIB addresses for each host and each Intel internally-managed chassis. Alternatively, an `/etc/hosts` file can be created on the Fabric Management node; FastFabric can then propagate this `/etc/hosts` file to all the other hosts.

   If using the `/etc/hosts` file approach:

   On the master node, add all the Ethernet and IPoIB addresses into the `/etc/hosts` file. For the IPoIB convention, use *HOSTNAME*`-ib`. The localhost line should not be edited.

The `/etc/hosts` file should not have any node-specific data. Copy the file to every node as described in "Copy /etc/hosts to all hosts" on page 54.

If using DNS:

Refer To the documentation for the domain name system (DNS) server being used. Make sure to edit the `/etc/resolv.conf` configuration file on the Fabric management node to use the proper DNS server. Refer To the Linux* OS documentation for more information about configuring `/etc/resolv.conf` file. This file is typically configured during OS installation.

If `/etc/resolv.conf` must be manually configured for each host, FastFabric can aid in copying the file to all the hosts. The `/etc/resolv.conf` file created on the Fabric management node must not have any node-specific data and must be appropriate for use on all hosts. Copying `/etc/resolv.conf` file to all the nodes is accomplished during the OS installation. If `/etc/resolv.conf` file was not setup on all the hosts during the OS installation, the **FastFabric Copy a file to all hosts** operation can be used during the "Install Intel® OFED+ Host Software on the Remaining Servers" procedures to copy `/etc/resolv.conf` file from the Fabric Management Node to all the other nodes.

7. **(All)** Set up a Network Time Protocol (NTP) server.

   Configure an NTP server for the cluster, and set all Linux* hosts and internally managed chassis to sync to the NTP server. The setup of the internally managed chassis is described in "Configure Intel Chassis" on page 31.

8. **(All)** Install the software.

   When installing a cluster, the next process is to install the True Scale Fabric Suite Software on the Fabric Management node, or when installing single nodes, the next process is to install the OFED+ Host Software. Refer to Table 2 for your installation requirements.

Table 2. Installations

Installation	Section
Install the True Scale Fabric Suite Software	Section 3.0
Install OFED+ Host Software	Section 4.0
Install Intel OFED+ Host Software Using Rocks	Section 6.0
Install Intel Software Using the Platform Cluster Manager Kit	Section 7.0

§ §

3.0 Install the True Scale Fabric Suite Software

This section provides information and procedures to install, configure, and verify the Intel® True Scale Fabric Suite (IFS) Software. The site implementation engineer must perform the tasks described in this section to correctly install and configure the fabric. Appendix A provides a checklist for tracking the installation process. FastFabric (FF) configuration files must be edited or created before you install the IFS software and are described in Appendix B.

The following procedures provide step-by-step installation, configuration, and verification instructions for a typical, single subnet fabric. For information about the installation and verification of multiple subnet fabrics, see Appendix C.

The following labels will be used at the beginning of the steps to specify environments and components:

- **(Linux) -** Tasks are only applicable when Linux* is being used.
- **(Host) -** Tasks are only applicable when OFED+, or IFS is being used on the hosts.
- **(Switch) -** Tasks are only applicable when True Scale Fabric Switches and Chassis are being used.
- **(All) -** Tasks that are generally applicable to all environments.

Note: Some Linux* steps are applicable to other Unix-like operating systems to enable use of non-True Scale specific FastFabric tools (such as `cmdall`) against the given hosts.

3.1 Fabric Management Node Installation

On hosts where the full IFS package has been purchased use the package file, `IntelIB-IFS.`*DISTRO*`.`*VERSION*`.tgz`.

3.1.1 Before You Install

Refer to the release notes for a list of compatible Operating Systems (OS)s.

The IFS software includes a compatible version of OFED+. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information on OFED+.

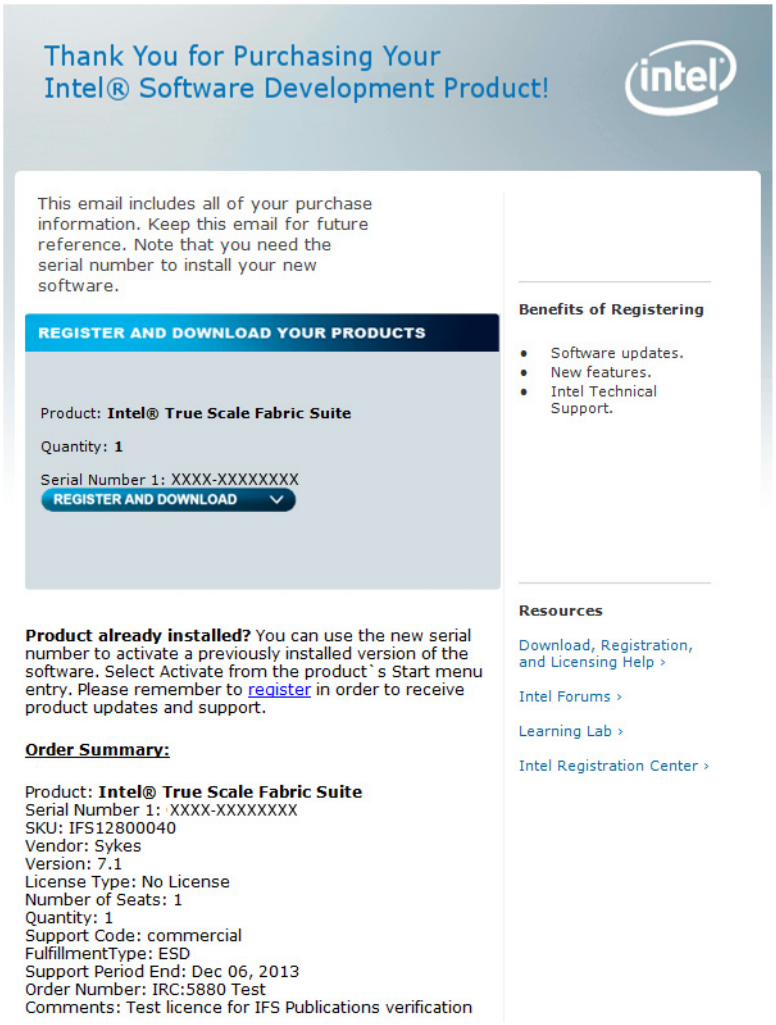
If the managed cluster has IPoIB settings on the compute nodes that are incompatible with the Fabric Management Node, do not run IPoIB on the Fabric Management Nodes. For example, when compute nodes use 4K Maximum Transmission Unit (MTU) and the Fabric Management Nodes use a 2K MTU, you would not run IPoIB settings on the Fabric Management Nodes. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*, Section 3 "True Scale Cluster Setup and Administration", Changing MTU Size subsection for detailed information.

3.1.2 Register and Download the True Scale Fabric Suite Software

Use the following procedure to register and download the True Scale Fabric Suite Software. When you purchased the True Scale Fabric Suite Software an e-mail was sent to the e-mail address provided during the purchase. Refer to that e-mail in the following procedure.

1. Select the **REGISTER AND DOWNLOAD** button in the e-mail received when the True Scale Fabric Suite Software was purchased. Figure 1 shows an example of the e-mail body.

Figure 1. Intel® Registration and Download E-Mail (Example)



The **True Scale Fabric Suite Software, Product Registration** web page will open.

Follow the instructions on the web pages to register and download the product.

3.1.3 Unpack the Tar File

Use the following procedure to unpack the `IntelIB-IFS.`*DISTRO.VERSION*`.tgz` tar file.

1. Copy the tar file to the /root directory.
2. Change directory to /root.

```
cd /root
```

3. Unpack the `IntelIB-IFS.`*DISTRO*`.`*VERSION* tar file to the `IntelIB-IFS.`*DISTRO*`.`*VERSION* directory using the following command:

```
tar xvfz IntelIB-IFS.DISTRO.VERSION.tgz
```

3.1.4 Install Intel IFS

To install the IFS, perform the following procedure:

1. Change directory to `IntelIB-IFS.`*`DISTRO`*`.`*`VERSION`* directory

```
cd IntelIB-IFS.DISTRO.VERSION
```

2. Start the `Install` TUI:

```
./INSTALL
```

Note: If you need 32-bit support on 64-bit OSs, enter the following command:
`./INSTALL --32bit`

The **Intel IB** *VERSION* **Software** main menu appears (Figure 2).

Figure 2. Intel IB Software Main Menu (Example)

```
Intel IB VERSION Software


   1) Install/Uninstall Software

   2) Reconfigure OFED IP over IB

   3) Reconfigure Driver Autostart

   4) Update HCA Firmware

   5) Generate Supporting Information for Problem Report

   6) FastFabric (Host/Chassis/Switch Setup/Admin)


   X) Exit
```

3. Press **1** to select `Install/Uninstall Software`.

Screen 1 of 3 of the **Intel IB Install Menu** appears (Figure 3).

Figure 3. Intel IB Install Menu (Screen 1 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 1 of 3):

0) OFED IB Stack       [   Install   ][Available] VERSION

1) True Scale HCA Libs [   Install   ][Available] VERSION

2) OFED mlx4 Driver    [   Install   ][Available] VERSION

3) IB Tools            [   Install   ][Available] VERSION

4) OFED IB Development [   Install   ][Available] VERSION

5) FastFabric          [   Install   ][Available] VERSION

6) OFED IP over IB     [   Install   ][Available] VERSION

7) OFED IB Bonding     [   Install   ][Available] VERSION

8) OFED SDP            [   Install   ][Available] VERSION

9) IFS FM              [   Install   ][Available] VERSION

a) MVAPICH (gcc)       [   Install   ][Available] VERSION

b) MVAPICH2 (gcc)      [   Install   ][Available] VERSION

c) OpenMPI (gcc)       [   Install   ][Available] VERSION

d) MVAPICH/PSM (gcc)   [   Install   ][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

Note: **True Scale HCA Libs** contains the enhanced Intel Host Channel Adapters (HCA) driver optimized stack for MPI (PSM) on HCAs and OpenMPI, as well as user tools.

Note: **OFED IB Bonding** will show as `[Not Avail]` when installing the software on OSs that have bonding modules in the OS installed software.

4. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
5. Press **N** to go to the next screen.

   Screen 2 of 3 of the **Intel IB Install Menu** appears (Figure 4).

Figure 4. Intel IB Install Menu (Screen 2 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 2 of 3):

0) MVAPICH/PSM (PGI)    [   Install   ][Available] VERSION

1) MVAPICH/PSM (Intel)  [   Install   ][Available] VERSION

2) MVAPICH2/PSM (gcc)   [   Install   ][Available] VERSION.DISTRO

3) MVAPICH2/PSM (PGI)   [   Install   ][Available] VERSION.DISTRO

4) MVAPICH2/PSM (Intel)[   Install   ][Available] VERSION.DISTRO

5) OpenMPI/PSM (gcc)    [   Install   ][Available] VERSION

6) OpenMPI/PSM (PGI)    [   Install   ][Available] VERSION

7) OpenMPI/PSM (Intel)  [   Install   ][Available] VERSION

8) SHMEM                [   Install   ][Available] VERSION.DISTRO

9) MPI Source           [   Install   ][Available] VERSION

a) OFED uDAPL           [   Install   ][Available] VERSION

b) OFED RDS             [   Install   ][Available] VERSION

c) OFED SRP             [   Install   ][Available] VERSION

d) OFED SRP Target      [Don't Install][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

6. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
7. Press **N** to go to the next screen.

   Screen 3 of 3 of the **Intel IB Install Menu** appears (Figure 5).

Figure 5. Intel IB Install Menu (Screen 3 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 3 of 3):

0) OFED iSER           [Don't Install][Available] VERSION

1) OFED iWARP          [Don't Install][Available] VERSION

2) OFED Open SM        [Don't Install][Available] VERSION

3) OFED NFS RDMA       [Don't Install][Available] VERSION

4) OFED Debug Info     [Don't Install][Not Avail]


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

8. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
9. Press **P** to perform the selected actions from all three screens.

   The system prompts:

```
About to Uninstall previous IB Software Installations...
Hit any key to continue...
```

10. Press any key to proceed with the installation.
11. The following system prompts appear. For each prompt, select the default by pressing **Enter**.

```
Rebuild OFED SRPMs (a=all, p=prompt per SRPM, n=only as needed?) [n]:


Permit non-root users to query the fabric? [y]:
```

Note: If you have had a previous version of the software installed and are installing this version after uninstalling a previous version, you might see the following prompt.

```
You have memory locking limits entries for IB drivers from an earlier install

You have a modified //etc/security/limits.conf configuration file

Do you want to keep //etc/security/limits.conf? [y]:


Enable OFED SMI/GSI renice (RENICE_IB_MAD)? [y]:
```

```
Single Port Mode reallocates all Intel HCA resources to HCA Port 1.
Enable Intel HCA Single Port Mode? [y]:
```

Note: Selecting the default by pressing **Enter** causes the dual-port HCAs to act as single-port cards with only port 1 enabled. Enabling Intel HCA Single Port Mode increases performance for environments where the second port is not connected.

Note: If there was a previous version of the software on this host that was uninstalled, you might see a few prompts asking if you want to keep certain files before the following prompt is shown.

```
Installing OFED IP over IB VERSION release...
Enable IPoIB Connected Mode (SET_IPOIB_CM)? [y]:
```

Press **Enter** to select default (y).

The system searches for ifcfg files, which contain IPv4 port IP and netmask addresses.

12. Perform one of the actions in the following table.

If	Then
The system finds the ifcfg files and you want to keep the files.	Answer [y] yes to the system prompt "`Do you want to keep OFED IP over IB ifcfg files:`" The system prompt following this table is shown. Continue with Step 13.
The system finds the ifcfg files and you do not want to keep the files; you want to input new IPv4 addresses.	Answer [n] no to the system prompt "`Do you want to keep OFED IP over IB ifcfg files.`" Proceed through the system prompts to set up the IP addresses. When you finish setting up the IP addresses, the system prompt following this table is shown. Continue with Step 13.
The system does not find the ifcfg files and you want to input new IPv4 addresses.	Answer [y] yes to the system prompt "`Configure OFED IP over IB IPV4 addresses now? [n]:`" Proceed through the system prompts to set up the IP addresses. When you finish setting up the IP addresses, the system prompt following this table is shown. Continue with Step 13.
The system does not find the ifcfg files and you do not want to input new IPv4 addresses at this time.	Answer [n] no to the system prompt "`Configure OFED IP over IB IPV4 addresses now? [n]:`" The system prompt following this table is shown. Continue with Step 13.
You are using IPv6 addresses	Answer [n] no to the system prompt "`Configure OFED IP over IB IPV4 addresses now? [n]:`" The system prompt following this table is shown. Continue with Step 13.

The system prompts:

```
Enable OFED SRP High Availability deamon (SRPHA_ENABLE)? [n]:
```

13. Press **Enter** to select default (n).

If a `ifcfg-ib*` configuration file exist, the following system prompts will be seen:

```
OFED IP over IB will autostart if ifcfg files exists

To fully disable autostart, it's recommended to also remove related ifcfg files
```

```
You have a modified //etc/sysconfig/network/ifcfg-ib[0-9]* configuration file

Do you want to keep OFED IP over IB ifcfg files
(//etc/sysconfig/network/ifcfg-ib[0-9]*)? [y]:
```

14. Press Enter to select default (y).

    The **Intel IB Autostart Menu** appears (Refer to Figure 6).

Figure 6. Intel IB Autostart Menu

```
Intel IB Autostart (VERSION release) Menu


Please Select Autostart Option:

0) OFED IB Stack (openibd)            [Enable ]

1) OFED mlx4 Driver (openibd)         [Enable ]

2) IB Port Monitor (iba_mon)          [Disable]

3) S20 Port Tuner (s20tune)           [Disable]

4) Distributed SA (dist_sa)           [Disable]

5) OFED IP over IB (openibd)          [Enable ]

6) OFED SDP (openibd)                 [Enable ]

7) IFS FM (ifs_fm)                    [Enable ]

8) OFED RDS (openibd)                 [Enable ]

9) OFED SRP (openibd)                 [Enable ]


P) Perform the autostart changes

S) Autostart All                     R) Autostart None

X) Return to Previous Menu (or ESC)
```

15. Review the items to be autostarted; the default value is in brackets (Enable or Disable). To change a value, type the alphanumeric character associated with the item.

    Intel recommends leaving all of the autostart selections as default, unless one of the following scenarios apply:

    — If FastFabric will not monitor the fabric health, performance, and/or check the fabric for errors, change `IB Port Monitor (iba_mon)` to **Enable**.

    Intel recommends changing `Distributed SA (dist_sa)` to **Enable** when installing software in mesh/torus fabrics, or when using Virtual Fabrics with Intel HCAs. The Distributed SA can be installed and run on any node in the fabric. It is only needed on nodes running SHMEM and MPI applications. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.

16. Press **P** to perform the selected actions from the screen.

    The system prompts:

```
Hit any key to continue...
```

17. Press any key to proceed with the installation.

    The system prompts with one of the following:

    — The following lines appear stating the firmware is not required when using Intel HCAs.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log

Firmware is not required for the Intel HCA(s) in this system.


Press any key to continue.
```

    Skip to 22.

    — The following lines appear showing the number of HCAs found.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log

One HCA was found:
```

    When one or more HCA is found, the system prompts with each HCA name and the firmware version installed, and if there is an update available or not. If a firmware update is available or the firmware is up to date, the system prompts to update, install different firmware, or do nothing. Only Connect-X HCAs will have firmware available. Refer to the following bulleted list for an example of the system prompt for each scenario:

    — An update is available (Example):

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware VERSION: Update to VERSION available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

    — The firmware is up to date (Example):

```
0: MT_0BB0110003 (MHQH29-XTC/X4/A0) Firmware VERSION: Okay.  (ConnectX )

1: MT_0D80120009 (MHQH29B-XTR/A2/B0) Firmware VERSION: Okay. (ConnectX-2)


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

    — No firmware is available. This displays if the HCA is not a supported HCA (Example).

```
0: MT_0390140002 (MHGA28-XTC/A4/A0) Firmware : No firmware available.

        Contact your vendor for firmware updates for this HCA.


No firmware available for HCAs in your system.

        Contact your vendor for firmware updates for this system.

Press any key to continue.
```

18. Perform one of the actions in the following table.

If	Then
No firmware is available	Skip to Step 22.
You need to upgrade the firmware	Proceed with Step 19.
You do not need to upgrade the firmware	Skip to Step 21.

19. Select a number corresponding to the HCA that needs to be upgraded.

    The system prompts (Example):

```
MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0

The following firmware revision(s) are available for this HCA:

0: MT_04A0110002: standard firmware

Select firmware version, or Q to cancel:
```

20. Select the number corresponding to the firmware revision required for the HCA.

    The firmware is installed on the HCA.

    The system prompts:

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0: Update to 2.5.0 available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

If	Then
You need to upgrade the firmware in another HCA	Repeat Step 19 and Step 20
You do not need to upgrade the firmware on any other HCAs	Continue with Step 21

21. Press **Q.**

    The installation completes and returns to the main menu.

    Skip to Step 24.

22. Press any key.

    The system prompts:

```
A System Reboot is recommended to activate the software changes

Hit any key to continue...
```

23. Press any key.

    The installation completes and returns to the main menu:

24. Press **X** to exit.

25. If installing IPoIBV6 proceed to Install IPoIB IPV6 section.
    If not installing IPoIBV6, reboot the server.

3.1.5 Install IPoIB IPV6

To install IPoIB for IPV6 on the management node use the following procedures for the OS on the Fabric management node.

3.1.5.1 On Red Hat*:

1. Edit file `/etc/sysconfig/network` to add the following line:

```
NETWORKING_IPV6=yes
```

2. Edit file `ifcfg-`*if-name* to add the following lines:

```
IPV6INIT=yes

IPV6ADDR="ipv6addr/prefix-length"
```

   Ipv6 address should look like the following:

```
3ffe::6/64
```

3. Restart the network.

3.1.5.2 On SUSE* Enterprise:

1. Edit `ifcfg-`*ifname* to add the following line:

```
IPADDR="ipv6addr/prefix-length"
```

   Ipv6 address should look like the following:

```
3ffe::6/64
```

2. Restart the network.

3.2 Configure Intel Chassis

Use FastFabric to install and configure internally managed switches, such as the Intel 12000 series switches. For information about installing and configuring switches made by other manufacturers, see the switch documentation.

3.2.1 Intel Chassis Configuration Pre-requisites

Ensure the internally managed switches are configured to use the True Scale Fabric Suite FastFabric toolset, by performing the following steps. Refer to the *Intel® True Scale Fabric Switches 12000 Series Hardware Installation Guide* and *Intel® 9000 Hardware Installation Guide* for more details:

1. **(Switch)** Connect each chassis to the management network through its Ethernet management port. For chassis with redundant management, connect both Ethernet management ports.
2. **(Switch)** Set up the netmask and gateway addresses on each Intel chassis following the procedures in the *Intel® True Scale Fabric Switches 12000 Series Users Guide.*
3. **(Switch)** Assign each Intel chassis a unique IP address and appropriately configure the Ethernet management port network settings.
4. **(Switch)** For a chassis with redundant management, assign a unique IP address for each Intel Management Module or Intel Management Spine, and configure their Ethernet management port network settings.

5. **(Switch)** Select a unique name for each Intel chassis, Management Module, and Spine. This name should be configured in DNS or `/etc/hosts` as the TCP/IP name for the Ethernet management port.

Note: The chassis node description will be set later.

6. **(Switch)** Configure the administrator password on each Intel chassis.

Note: All versions of Intel® 12000 chassis firmware permit SSH keys to be configured within the chassis for secure password-less login. To simplify the configuration of SSH security using FastFabric, configure all chassis with the same initial administrator password (or leave the default "adminpass" until FastFabric has set up ssh keys), configure the SSH keys, then change the administrator passwords. After ssh has been set up using FastFabric, it is recommended to change the admin passwords.

7. **(Switch)** Copy the relevant chassis firmware files onto the FastFabric management node. During the following steps, the `*.pkg` files will be used to upgrade the firmware on each chassis.

Note: Place all files at a given firmware level into a single directory whose name indicates the firmware revision number.

3.2.2 Configure Chassis Using True Scale Fabric Suite FastFabric

Configure the Chassis using True Scale Fabric Suite FastFabric. Refer to the *Intel® True Scale Fabric Suite FastFabric User Guide* for more information on how to use the True Scale Fabric Suite FastFabric TUI

1. **(Switch)** Type `fastfabric` and press **Enter**.

   The **Intel FastFabric IB Tools** menu is displayed (Figure 7).

Figure 7. IntelFastFabric IB Tools Menu (Example)

```
Intel FastFabric IB Tools

Version: X.X.X.X.X


   1) Chassis Setup/Admin

   2) Externally Managed Switch Setup/Admin

   3) Host Setup

   4) Host Verification/Admin

   5) Fabric Monitoring


   X) Exit
```

2. **(Switch)** Press **1**.

   The **FastFabric IB Chassis Setup/Admin Menu** is displayed (Figure 8).

Figure 8. FastFabric IB Chassis Setup/Admin Menu

```
FastFabric IB Chassis Setup/Admin Menu

Chassis File: /etc/sysconfig/iba/chassis


Setup:

0) Edit Config and Select/Edit Chassis File  [ Skip  ]

1) Verify Chassis via Ethernet ping          [ Skip  ]

2) Update Chassis Firmware                   [ Skip  ]

3) Setup Chassis Basic Configuration         [ Skip  ]

4) Setup Password-less ssh/scp               [ Skip  ]

5) Reboot Chassis                            [ Skip  ]

6) Configure Chassis Fabric Manager          [ Skip  ]

7) Get Basic Chassis Configuration           [ Skip  ]

Admin:

8) Check IB Fabric status                    [ Skip  ]

9) Control Chassis Fabric Manager            [ Skip  ]

a) Generate all Chassis Problem Report Info  [ Skip  ]

b) Run a command on all chassis              [ Skip  ]

Review:

c) View iba_chassis_admin result files       [ Skip  ]


P) Perform the selected actions              N) Select None

X) Return to Previous Menu (or ESC)
```

3. (**Switch**) Select the items in the Setup section that are required. Type the alphanumeric character associated with the item to toggle the selection from Install to Don't Install.
4. Press **P**

   Perform the items that were selected in the sub-sections as follows.

3.2.2.1 Edit the Configuration and Select/Edit Chassis File

(Switch) The **Edit Config and Select/Edit Chassis File** selection will permit the chassis, ports, and FastFabric configuration files to be edited.

- When placed in the editor for `fastfabric.conf`, review the settings.
  - Especially review the `FF_CHASSIS_LOGIN_METHOD` and `FF_CHASSIS_ADMIN_PASSWORD`. Refer to Appendix B for more information about the `fastfabric.conf` file.

Note: FastFabric will provide the opportunity to enter the chassis password interactively when needed. It is not necessary to place it within `fastfabric.conf`. If the Intel chassis admin password is placed in `fastfabric.conf`, change the `fastfabric.conf` permissions to be 0x600 (e.g., root-only access).

Note: All versions of Intel® 12000 chassis firmware permit ssh keys to be configured within the chassis for secure password-less login, There is no need to configure a `FF_CHASSIS_ADMIN_PASSWORD`, and `FF_CHASSIS_LOGIN_METHOD` can be set to ssh (the default) when using the newer versions of the chassis firmware.

— Select the location for the result files from FastFabric through the FF_RESULT_DIR parameter. The default is the directory from which a given session of fastfabric is invoked. Alternatively it can be set to a directory relative to the users home. For example:

```
export FF_RESULT_DIR=${FF_RESULT_DIR:-$HOME/
fastfabric_results}
```

— Refer to Appendix A of the *Intel® True Scale Fabric Suite FastFabric User Guide* for more information about `fastfabric.conf`

- When placed in the editor for `ports`, review the file. For typical single-subnet clusters, the default of "0:0" may be used. This will use the first active port on the Fabric Management Node to access the fabric. For more information on configuring a cluster with multiple subnets, see Appendix C. For further details about the file format, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.
- When placed in the editor for `chassis`, create the file with a list of the chassis names (the TCP/IP Ethernet management port names assigned) or IP addresses (use of names is recommended). Enter one chassis name or IP address per line. For example:

```
Chassis1

Chassis2
```

Note: `Refer to` Section 3.4.2.2, "Generate or Update Switch File" on page 45 to generate a list of the externally managed switches presently in the fabric.

Note: Do not list externally managed switches, such as the Intel 12200 switches in this file. Those will be covered in the next section.

For further details about the file format refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

3.2.2.2 Verify Chassis via Ethernet ping

(Switch) The **Verify Chassis via Ethernet ping** selection will ping each selected chassis over the management network. If all chassis were found, continue to the next step. If some chassis were not found, abort out of the menu and review the following for those chassis which were not found:

- Is chassis powered on and booted
- Is chassis connected to management network
- Are chassis IP address and network settings consistent with DNS or `/etc/hosts`
- Is Management node connected to the management network
- Are Management node IP address and network settings correct
- Is the management network itself up (switches, routers, etc)

- Is correct set of chassis listed in the chassis file (the previous step may be repeated to review and edit the file as needed)?

3.2.2.3 Update Chassis Firmware

(Switch) The **Update Chassis Firmware** selection will permit the chassis firmware version to be verified and updated as needed.

Note: Refer To the relevant chassis firmware release notes to ensure any prerequisites for the upgrade to the new firmware level have been met prior to performing the upgrade through FastFabric.

1. When this procedure is started the following system prompt will be displayed:

```
Multiple Firmware files and/or Directories may be space separated

Shell wildcards may be used

For Directories all .pkg files in the directory tree will be used

Enter Files/Directories to use (or none):
```

2. Specify the directory where the relevant firmware files have been stored and press **Enter**.

   This can be the mount point of the CD or the directory to which the files were copied in a previous step.

   System prompts:

```
Would you like to run the firmware now? [n]:
```

3. Type y and press **Enter** since the fabric is not yet operational.

   FastFabric will ensure that all chassis are running the firmware level provided and install and/or reboot each chassis as needed.

   If any chassis fails to be updated, use the **View iba_chassis_admin result files** option to review the result files from the update. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

3.2.2.4 Set Up Chassis Basic Configuration

(Switch) The **Setup Chassis Basic Configuration** will permit the typical chassis setup operations to be performed for all chassis.

Perform the following procedure:

1. When this procedure is started the following system prompt will be displayed:

```
Would you like to be prompted for chassis' password? [n]:
```

2. Press **Enter** to select default (n).

   System prompts:

```
Do you wish to configure a syslog server? [y]:
```

3. Press **Enter** to select default (y).

   System prompts:

```
Enter IP address for syslog server:
```

4. Enter the IP address of a syslog server which is to receive log messages from all chassis.

   System prompts:

```
Do you wish to configure the syslog TCP/UDP port number? [n]:
```

5. Press ENTER to select default (n).

   System prompts:

```
Do you wish to configure the syslog facility? [n]:
```

6. Press ENTER to select default (n).

   System prompts:

```
Logging to the syslog can be configured in a Quiet Mode, where only user actionable
events are logged. Otherwise, in Normal Mode, all events will be logged

Do you wish to configure Quiet Mode for syslog? [n]:
```

7. Press ENTER to select default (n).

   System prompts:

```
Do you wish to configure an NTP server? [y]:
```

8. Press **Enter** to select default (y).

   System prompts:

```
Enter IP address for NTP server:
```

9. Enter the IP address of an NTP server which can supply a consistent time base for use by all chassis.

   System prompts:

```
Do you wish to configure timezone and DST information? [y]:
```

10. Press **Enter** to select default (y).

    System prompts:

```
Do you want to use the local timezone information from the local server? [y]:
```

11. Press **Enter** to select default (y).

    This will cause the time zone of the local server (e.g., the Fabric Management Node) to be replicated to all the chassis to specify their time zones.

    System prompts:

```
Do you wish to configure the chassis maximum packet MTU size? [n]:
```

12. Press **Enter** to select default (n).

    This will cause the default MTU of 2048 to be used for all chassis. If chassis have previously been manually configured for a different MTU size, this option will keep the previously configured MTU size. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*, Section 3 "True Scale Cluster Setup and Administration", Changing MTU Size subsection for detailed information.

    System prompts:

```
Do you wish to configure the chassis VL Capability? [n]:
```

13. Press **Enter** to select default (n).

    This will cause the default VL Capability of 1 to be used for all chassis. If chassis have previously been manually configured for a different VL Capability, this option will keep the previously configured VL Capability size. See the *Intel® True Scale Fabric Switches 12000 Series Users Guide* for more information.

    System prompts:

```
Do you wish to configure the VL Credit Distribution? [n]:
```

Note: Always select the default (n) for the question "`Do you wish to configure the VL Credit Distribution?`" unless otherwise instructed by Intel support.

14. Press ENTER to select default (n).

    System prompts:

```
Do you wish to configure the chassis link width? [n]:
```

15. Press **Enter** to select default (n).

    This will cause the default link width supported value of 1x/4x/8x to be used for all ports on all chassis. If chassis have previously been manually configured for a different link width supported, those setting will remain unchanged.

    System prompts:

```
Do you wish to configure IB Node Desc to match ethernet chassis name? [y]:
```

16. Press **Enter** to select default (y).

    This will cause the chassis name entered in the `/etc/sysconfig/iba/chassis` file to be used as the IB Node Description for the chassis, making the management network and IB network names for the chassis consistent.

    If the `/etc/sysconfig/iba/chassis` file has IP addresses instead of names, enter `n` to this question

    System prompts:

```
Do you wish to configure IB Node Desc Format? [y]:
```

17. Press **Enter** to select default (y)

    System prompts:

```
Do you wish to use concise IB Node Desc format? [y]:
```

18. Press **Enter** to select default (y).

    This will cause the chassis IB Node Descriptions to use concise naming for the Leafs and Spines such as L01 or S01A (as opposed to "Leaf 1" or "Spine 1, Chip A").

    System prompts:

```
Do you wish to configure the port counter auto-clear feature? [n]:
```

If	Then
You are not using iba_rfm, or iba_top.	Continue with Step 19
You will be using iba_rfm, or iba_top.	Skip to Step 20

19. Press **Enter** to select default (n).

    Selecting the default will leave the auto clear feature as previously configured.

    Skip to Step 21.
20. Press **Y** to select to disable this feature.

    This will cause the port counter auto-clear feature to be disabled on all chassis.
21. Continue with Setup Password-less ssh/scp.

3.2.2.5 Setup Password-less ssh/scp

(Switch) The **Setup Password-less ssh/scp** selection will set up secure password-less ssh such that the Fabric Management Node can securely log in to all the chassis as admin through the management network without requiring a password.

3.2.2.6 Reboot Chassis

(Switch) The **Reboot Chassis** selection will reboot all the selected chassis and ensure they go down and come back up (as verified through ping over the management network). When the chassis come back up, they will be running with all the new configuration settings.

3.2.2.7 Configure Chassis Fabric Manager

(Switch) The **Configure Chassis Fabric Manager** selection will assist in configuring the Fabric Manager for any Intel 12000 chassis with appropriate license keys.

System prompts:

```
Performing Chassis Admin: Configure Chassis Fabric Manager

Enter FM Config file to use (or none or generate):
```

1. Enter generate.

   This will perform the config_generate operation to guide the user through selecting FM configuration options. See the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for more information about config_generate.
2. After responding to the prompts for config_generate, the following system prompt will display:

```
You have selected to use: ./ifs_fm.xml

Syntax Checking ./ifs_fm.xml...

Executing: /opt/iba/fm_tools/config_check -s -c ./ifs_fm.xml

Valid FM Config file: ./ifs_fm.xml


After push, the FM may be started/restarted
```

```
Would you like to restart the FM? [n]:
```

3. Select "y", this will cause the FM to be started with the new configuration.

   System prompts:

```
Would you like to run the FM on slave MMs? [n]:
```

4. Refer to the following If/Then table:

If	Then
Your fabric has a single chassis running the FM, you can run the FM on the slave management module (MM). This will cause the FM to be started in the applicable chassis.	Type **y**
Your fabric has multiple chassis running the FM, Intel recommends to run FM on the master MM. This will cause the FM to only be started on the master MM in the applicable chassis	Type **N**

Note: Ensure that there is a License Key for each management module that will be running the Fabric Manager

   System prompts:

```
There will be a disruption as FMs are restarted

Doing the operation in parallel (on multiple chassis) will finish the fastest

Doing it serially may reduce disruption

Would you like to do the operation in parallel? [y]:
```

5. Intel recommends to do the operation in parallel. Press **Enter** to select the default y.

   System prompts:

```
You have selected to perform the push, and FM restart in parallel


Would you like to enable FM start at boot? [n]:
```

6. Select "y", this will cause the FM to be started on all applicable chassis each time those chassis boot.

   System prompts:

```
Would you like to enable FM start on slave MMs at boot? [n]:
```

7. Refer to the following If/Then table:

If	Then
Your fabric has a single chassis running the FM, you can run the FM on the slave management module (MM). This will cause the FM to be started in the applicable chassis.	Type **y**
Your fabric has multiple chassis running the FM, Intel recommends to run FM on the master MM. This will cause the FM to only be started on the master MM in the applicable chassis	Type **N**

System prompts:

```
Would you like to be prompted for chassis' password? [n]:
```

8. Press **ENTER** to select the default "n".

System prompts:

```
Are you sure you want to proceed? [n]:
```

9. Select "y", this will update the FM.

System prompts:

```
Hit any key to continue (or ESC to abort)...
```

3.2.2.8 Get Basic Chassis Configuration

(Switch) The **Get Basic Chassis Configuration** selection will retrieve basic information from chassis such as syslog, NTP configuration, timezone info, MTU Capability, VL Capability, VL Credit Distribution, Link Width and node description. The following is an example of the output from this selection:

```
Executing: /sbin/iba_chassis_admin  -F /etc/sysconfig/iba/chassis getconfig

Executing getconfig Test Suite (getconfig) DAY mth dd hh:mm:ss CST yyyy ...

Executing TEST SUITE getconfig CASE (getconfig.00.000.00.00.getconfig) get
00.000.00.00 ...

TEST SUITE getconfig CASE (getconfig.00.000.00.00.getconfig) get 10.228.72.90

00.000.00.00:

    Product Family          : 9000

    Firmware Active         : 4.2.X.X.X

    Firmware Primary        : 4.2.X.X.X

    Syslog Configuration    : Syslog host set to: 0.0.0.0 port 514 facility 22

    NTP                     : Configured to use the local clock

    time zone               : Time zone offset has not been configured

    MTU Capability          : 2048 Bytes

    VL Capability           : 8 VLs

    LinkWidth Support       : 12x mode is DISABLED

    Node Description        : SilverStorm 9080 GUID=0x00066a00da000131

    Auto clear status       : Auto clear is enabled


PASSED

Executing TEST SUITE getconfig CASE (getconfig.00.000.00.00.getconfig) get
00.000.00.00 ...

TEST SUITE getconfig CASE (getconfig.00.000.00.00.getconfig) get 00.000.00.00

00.000.00.00:
```

```
    Product Family          : 12000
    Firmware Active         : 7.2.X.X.X
    Firmware Primary        : 7.2.X.X.X
    Syslog Configuration    : Syslog host set to: 0.0.0.0 port 514 facility 22
    NTP                     : Configured to use NTP server: 00.000.00.01
    time zone               : Current time zone offset is: -6
    MTU Capability          : 2048 Bytes
    VL Capability           : 1 VLs
    VL Credit Distribution : 4
    LinkWidth Support       : 1X/4X/8X
    Node Description        : Intel 12800-360 GUID=0x00066a00e9000111
    Auto clear status       : Auto clear is disabled


PASSED
Summary:
count - configuration
    1 - Auto clear status       : Auto clear is disabled
    1 - Auto clear status       : Auto clear is enabled
    1 - Firmware Active         : 4.2.X.X.X
    1 - Firmware Active         : 7.2.X.X.X
    1 - Firmware Primary        : 4.2.X.X.X
    1 - Firmware Primary        : 7.2.X.X.X
    1 - LinkWidth Support       : 12x mode is DISABLED
    1 - LinkWidth Support       : 1X/4X/8X
    2 - MTU Capability          : 2048 Bytes
    1 - NTP                     : Configured to use NTP server: 00.000.00.01
    1 - NTP                     : Configured to use the local clock
    1 - Product Family          : 12000
    1 - Product Family          : 9000
    2 - Syslog Configuration    : Syslog host set to: 0.0.0.0 port 514 facility 22
    1 - time zone               : Current time zone offset is: -6
    1 - time zone               : Time zone offset has not been configured
    1 - VL Capability           : 1 VLs
    1 - VL Capability           : 8 VLs
```

```
    1 - VL Credit Distribution : 4

TEST SUITE getconfig: 2 Cases; 2 PASSED

TEST SUITE getconfig PASSED

Done getconfig Test Suite DAY mth dd hh:mm:ss CST yyyy
```

3.2.2.9 Check IB Fabric status

(**All**) The **Check IB Fabric status** selection prompts the user to:

- Perform a fabric error analysis,
- Clear the error counters after generating a report
- Perform a fabric link speed error analysis
- Check for links configured to run slower than supported
- Check for links connected with mismatched speed potential
- Enter a filename for the results or save the results to the default location (/root/ffres/linkanalysis.res

For more information refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*.

3.2.2.10 Control Chassis Fabric Manager

(**Switch**) The **Control Chassis Fabric Manager** selection prompts the user to:

- Restart the Fabric Manager
- Run the Fabric Manager on slave MMs
- Restart the Fabric Manager on all MMs
- Perform this operation in parallel
- Change the Fabric Manager boot state to enable the Fabric Manager to start at boot
- Enable Fabric Manager to start on slave MMs at boot
- Be prompted for chassis' password

For more information refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*.

3.2.2.11 Generate all Chassis Problem Report Information

(**Switch**) The **Generate all Chassis Problem Report Info** selection generates the chassis problem report. For more information refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*.

3.2.2.12 Run a command on all chassis

(Switch) If there are any other operations which need to be performed on all chassis, they may be performed using the **Run a command on all chassis** option. Each time this is executed, a single chassis CLI command may be specified to be executed against all selected chassis. When using these commands, additional setup or verification of the chassis may be performed.

3.2.2.13 View iba_chassis_admin result files

(**Switch**) The **View iba_chassis_admin result files** selection opens the `/root/ffres/punchlist.csv`, `/root/ffres/test.res`, and `/root/ffres/test.log` files to be viewed. For more information refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*.

3.3 Install and Configure the Fabric Manager

(All) At this point the True Scale Fabric Suite Fabric Manager for the fabric must be enabled. If using a Host FM, section "Install Intel IFS" on page 23 will have installed and enabled the True Scale Fabric Suite Fabric Manager using the default configuration file. If using an embedded True Scale Fabric Suite Fabric Manager the, "Configure Chassis Fabric Manager" on page 38 will have configured the True Scale Fabric Suite Fabric Manager. Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for information on how to configure the True Scale Fabric Suite Fabric Manager.

When using the True Scale Fabric Suite Fabric Manager, a typical True Scale Fabric Suite Software installation will place the True Scale Fabric Suite FastFabric and the True Scale Fabric Suite Fabric Manager on the same Fabric Management Node. If required, it is also valid to place True Scale Fabric Suite FastFabric on its own independent management node, perhaps along with other 3rd party management applications (such as MPI job schedulers, etc).

The following procedures require that a subnet manager be operational within the fabric.

3.4 Configure Firmware on the Externally Managed Intel Switches

If the fabric contains Intel 12200 series externally managed switches, True Scale Fabric Suite FastFabric is used to aid in the installation and configuration of the switches. If the fabric contains another vendor's switches please refer to the vendor's documentation to configure the firmware on the externally managed switches.

3.4.1 Switch Configuration Pre-Requisites

Prior to using True Scale Fabric Suite FastFabric, the following minimal steps need to be performed:

1. (**Switch**) Select a unique name to be used for each switch. This name will be configured as the IB Node Description for the switch in the following steps.

Note: Externally managed switches do not have an Ethernet port and therefore will not have a TCP/IP name.

2. (**Switch**) Copy the relevant switch firmware files onto the True Scale Fabric Suite FastFabric management node. For the following steps, the *.emfw files will be used to upgrade the firmware on each switch.

Note: When copying files, it is best to place all files at a given firmware level into a single directory whose name indicates the firmware revision number

3.4.2 Configure Externally Managed Switches

Once the pre-requisites have been completed, configure the switches using True Scale Fabric Suite FastFabric in the following procedure.

1. **(Switch)** If the **Intel FastFabric IB Tools** menu is not displayed type `fastfabric` and press **Enter**.
2. (**Switch**) Press **2**.

   Displays the **FastFabric IB Switch Setup/Admin Menu** (Figure 9).

Figure 9. FastFabric IB Switch Setup/Admin Menu

```
FastFabric IB Switch Setup/Admin Menu

Externally Managed Switch File: /etc/sysconfig/iba/ibnodes


Setup:

0) Edit Config and Select/Edit Switch File    [ Skip  ]

1) Generate or Update Switch File             [ Skip  ]

2) Test for Switch Presence                   [ Skip  ]

3) Verify Switch Firmware                     [ Skip  ]

4) Update Switch Firmware                     [ Skip  ]

5) Setup Switch Basic Configuration           [ Skip  ]

6) Reboot Switch                              [ Skip  ]

7) Report Switch Firmware & Hardware Info     [ Skip  ]

8) Get Basic Switch configuration             [ Skip  ]

Admin:

9) Report Switch VPD Information              [ Skip  ]

a) Generate all Switch Problem Report Info    [ Skip  ]

Review:

b) View iba_switch_admin result files         [ Skip  ]


P) Perform the selected actions              N) Select None

X) Return to Previous Menu (or ESC)
```

3. (**Switch**) Select the items **0** through **4** in the Setup section of the menu.
4. Press **P**.
5. Perform the items selected using the following sections.

3.4.2.1 Edit Config and Select/Edit Switch File

(**Switch**) The **Edit Config and Select/Edit Switch File** selection will permit the ibnodes, ports, and FastFabric configuration files to be edited. When placed in the editor for `fastfabric.conf`, review all the settings. Refer to Appendix B for more information about fastfabric.conf.

When placed in the editor for ports, review the file. For typical single-subnet clusters, the default of "0:0" may be used. This will use the first active port on the Fabric Management Node to access all externally managed switches. For more information on configuring a cluster with multiple subnets, see Appendix C. For further details about the file format, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

When placed in the editor for ibnodes, create the file with a list of the switch node GUID and required switch names, Enter one switch node GUID and required switch name per line. For example:

```
0x00066a00d9000138,edge1

0x00066a00d9000139,edge2
```

Note: Per the previous example, when typing a new name, do not use any spaces before or after the comma.

Note: The Generate or Update Switch File menu item or `iba_gen_ibnodes` may be used to generate a list of the externally managed switches presently in the fabric. For example when using the vi editor, the command `:r ! iba_gen_ibnodes` may be used to add the output from this command to the file.

Note: Do not list internally managed chassis, such as the Intel 12000 chassis in this file. Those were covered in a previous section.

Note: FastFabric is topology aware when updating externally managed switch firmware or resetting the switches. The update or restart will start at the switches furthest from the FastFabric node and then work toward the FastFabric node. This way switches which are rebooted will not be in the path between the FastFabric node and others which are being updated or reset.

For further details about the file format, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

If needed, an SA query such as the following can be used to get a list of all switches. This includes both internally and externally managed switches, and consequently the output must be edited to leave only the Intel externally managed switches:

```
iba_saquery -t sw -o nodeguid
```

3.4.2.2 Generate or Update Switch File

(**Switch**) The **Generate or Update Switch File** selection generates or updates the ibnodes file. It can also update switch names in the ibnodes file by comparing the actual fabric to a topology xml data.

3.4.2.3 Test for Switch Presence

(**Switch**) The **Test for Switch Presence** selection will verify that each Externally Managed Switch specified in the ibnodes file can be accessed by the Fabric Management Node through the Fabric Network. If all switches were found, continue to the next step. If some switches were not found, abort out of the menu and review the following for those switches which were not found:

- Is switch powered on and booted
- Is switch connected to True Scale Fabric
- Is Subnet Manager running
- Is Fabric Management node's Port active

- Is Fabric Management node connected to the correct True Scale Fabric
- Is correct set of switches listed in the ibnodes file (the previous step may be repeated to review and edit the file as needed)?

3.4.2.4 Verify Switch Firmware

(**Switch**) The **Verify Switch Firmware** selection will verify each externally managed switch is operational and its firmware is valid and accessible.

3.4.2.5 Update Switch Firmware

(**Switch**) The **Update Switch Firmware** selection will permit the switch firmware version to be updated and the switch node name set.

Note: Refer To the relevant switch firmware release notes to ensure any prerequisites for the upgrade to the new firmware level have been met prior to performing the upgrade through FastFabric.

Perform the following procedure:

1. When this procedure is started the following message will be displayed:

```
Multiple Firmware files and/or Directories may be space separated

Shell wildcards may be used

For Directories all .emfw files in the directory tree will be used

Enter Files/Directories to use (or none):
```

2. Specify the directory where the relevant firmware files have been stored. This can be the mount point of the CD or the directory to which the files were copied in a previous step.

   The following message will display:

```
After upgrade, the switch may be optionally rebooted

Would you like to reboot the switch after the update? [n]:
```

3. Enter `y`

   The following message will display:

```
The firmware on the switch will be checked, and if the running version is the same
as the version being used for the update, the update operation will be skipped

Would you like to override this check, and force the update to occur? [n]:
```

4. Press **Enter** to select default (n).

   The fabric is not yet operational

   The following message will display:

```
You have selected to update the switch firmware and reboot.

There will be a disruption as switch or switches are rebooted

Doing the operation in parallel (on multiple switches) will finish the fastest

Doing it serially may reduce disruption

Would you like to do the operation in parallel? [y]:
```

5. Since the True Scale Fabric itself is used to update externally managed switches, updating multiple switches with the reboot option may disrupt parallel update operations.If there are not any selected externally managed switches in the path from the Fabric Management node to any other externally managed switch (for example, if the Fabric Management node is connected directly to a core switch and externally managed switches are only at the edges), parallel operations can be established. To control the order of the rebooting of externally managed switches by FastFabric, see the discussion of the `distance` option for `ibnodes` file or command in the *Intel® True Scale Fabric Suite FastFabric User Guide*, Appendix A, subsection "Externally Managed Switch List File."

   Press **Enter**.

   or

   Type `n` and press **Enter** if in doubt.

Note: Be aware that non-parallel operation for a fabric with many externally managed switches could take a significant amount of time.

FastFabric will update the firmware on all switches and set the node names as per the ibnodes file created in a previous step. Each switch will then be rebooted.

If any switch fails to be updated, use the **View iba_switch_admin result files** option to review the result files from the update. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

3.4.2.6 Set Up Switch Basic Configuration

(Switch) The **Setup Switch Basic Configuration** will permit the typical switch setup operations to be performed for all switch.

Perform the following procedure:

1. When this procedure is started the following message will be displayed:

```
Do you wish to configure the switch maximum MTU size? [n]:
```

2. Press **ENTER** to select default (n).

   This will cause the default MTU of 2048 to be used for all switches. If the switches have previously been manually configured for a different MTU size, this option will keep the previously configured MTU size. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*, Section 3 "True Scale Cluster Setup and Administration", Changing MTU Size subsection for detailed information.

   The following message will display:

```
Do you wish to configure the switch VL Capability? [n]:
```

3. Press **ENTER** to select default (n).

   This will cause the default VL Capability of 1 data VL to be used for all switches. If switches have previously been manually configured for a different VL Capability, this option will keep the previously configured VL Capability. See the *Intel® True Scale Fabric Switches 12000 Series Users Guide* for more information.

Note: This operation is only applicable to Intel 12000 switches.

   The following message will display:

```
Do you wish to configure the switch VL Credit Distribution Meathod? [n]:
```

4. Press **ENTER** to select default (n).

   This will cause the default VL Credit Distribution Method setting of 4 to be used for all switches. If switches have previously been manually configured for a different VL Credit Distribution Method setting, those settings will remain unchanged. See the *Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide* for more information.

Note: This operation is only applicable to Intel 12000 switches.

The following message will display:

```
Do you wish to configure the switch Link Width Options? [n]:
```

5. Press **ENTER** to select default (n).

   This will cause the default switch Link Width Options to be used for all switches. If switches have previously been manually configured for different switch Link Width Options, this option will keep the previously configured switch Link Width Options. See the *Intel® True Scale Fabric Switches 12000 Series Users Guide* for more information.

   Selecting (y) will prompt for setting the switch link width supported setting for all ports on all switches.

Note: This operation is only applicable to Intel 12000 switches.

The following message will display:

```
Do you wish to configure the switch Link Speed Options? [n]:
```

6. Press **ENTER** to select default (n).

   This will cause the default switch Link Speed Options to be used for all switches. If switches have previously been manually configured for different switch Link Speed Options, this option will keep the previously configured switch Link Speed Options. See the *Intel® True Scale Fabric Switches 12000 Series Users Guide* for more information.

   Selecting (y) will prompt for setting the switch link speed supported setting for all ports on all switches.

Note: This operation is only applicable to Intel 12000 switches.

The following message will display:

```
Do you wish to configure the switch Node Description as it is set in the ibnodes
file? [n]:
```

7. Press **ENTER** to select default (n).

   This will cause the default switch Node Description on each switch to be used. If the switches have previously been manually configured for a customized switch Node Description, this option will keep the previously configured switch Node Descriptions. See the *Intel® True Scale Fabric Switches 12000 Series Users Guide* for more information.

   Selecting (y) will cause the Node Description on each switch to be updated as specified by the ibnodes file referenced in Appendix B.

Note: Only node descriptions on Intel 12000 switches can be changed in this step.

3.4.2.7 Reboot Switch

(Switch) The **Reboot Switch** will reboot all switches, this will ensure that all the configuration changes become effective and are discovered by the Fabric Manager.

3.4.2.8 Report Switch Firmware and Hardware Info

(Switch) The **Report Switch Firmware and Hardware Info** selection will report the firmware and hardware versions for each switch, along with the Capability (QDR, DDR, or SDR), Fan Status, and Power Supply Status. Review the results against the expected models and firmware versions.

(Switch) If any 12200 were purposely skipped, these sections should be repeated for those switches. In this case it is recommended to create a separate file with a name other than ibnodes. An alternate name may be specified at the prompt:

```
Select Switch File to Use/Edit

[/etc/sysconfig/iba/ibnodes]:
```

3.4.2.9 Get Basic Switch configuration

(Switch) The **Get Basic Switch configuration** selection executes the report switch get config Test Suite (`switchgetportconfig`) for all of the ports. The results show how many cases, how many of the cases passed, and how many of the cases failed, it also gives an overall summary of configuration and passed or failed as shown in the following example:

```
MTU                                 :  4096

        VL Capability                       :  1+1

        VL Credit Distribution Method       :  0

        Link Width                          :  1-4x

        Link Speed                          :  2.5-10Gb

        Node Description                    :  12200-18


PASSED

Summary:

count - configuration

    1 - Link Speed                          :  2.5-10Gb

    1 - Link Width                          :  1-4x

    1 - MTU                                 :  4096

    1 - VL Capability                       :  1+1

    1 - VL Credit Distribution Method       :  0

TEST SUITE report switch getconfig: 1 Cases; 1 PASSED

TEST SUITE report switch getconfig PASSED

Done report switch getconfig Test Suite Mon Mar 19 10:16:24 EDT 2012
```

3.4.2.10 Report Switch VPD Information

(Switch) The **Report Switch VPD** (vital product data) **Information** selection executes the report switch hwvpd Test Suite command (`iba_switch_admin`) for all of the nodes listed in `/etc/sysconfig/iba/ibnodes`. The results show the VPD hardware information as shown in the following example:

```
    0x00066a00e3000100: H/W VPD serial number: USF1011100108

    0x00066a00e3000100: H/W VPD part number  : 220058-604-B

    0x00066a00e3000100: H/W VPD model        : 12200-18

    0x00066a00e3000100: H/W VPD h/w version  : 604-B

    0x00066a00e3000100: H/W VPD manufacturer : Intel

    0x00066a00e3000100: H/W VPD prod desc    : 12200 Fixed EDGE - Push Air

    0x00066a00e3000100: H/W VPD mfg id       : 00066a

    0x00066a00e3000100: H/W VPD mfg date     : 1-01-2011

    0x00066a00e3000100: H/W VPD mfg time     : 10:00

PASSED

TEST SUITE report switch hwvpd: 1 Cases; 1 PASSED

TEST SUITE report switch hwvpd PASSED

Done report switch hwvpd Test Suite Mon Mar 19 10:17:20 EDT 2012
```

3.4.2.11 Generate all Switch Problem Report Info

(Switch) The **Generate all Switch Problem Report Info** selection executes the capture all command (`captureall`) to capture supporting information for a problem report for all of the Fabric nodes listed in `/etc/sysconfig/iba/ibnodes` in parallel on all hosts. The supporting information is captured and a `switchcapture.all.tgz` file is created in the `/uploads` directory as shown in the following test example:

```
TEST SUITE capture switch state CASE
(switchcapture.0x00066a00e3000100.i2c.extmgd.switchcapture) capture switch
0x00066a00e3000100 PASSED

TEST SUITE capture switch state: 1 Cases; 1 PASSED

TEST SUITE capture switch state PASSED

Done capture switch state Test Suite Mon Mar 19 10:29:31 EDT 2012



Combining captured files into ./uploads/switchcapture.all.tgz ...

Done.
```

3.4.2.12 View iba_switch_admin result files

(Switch) The **View iba_switch_admin result files** selection starts the editor to view iba_switch_admin result files. The two files that are opened in the editor are `/root/test.res` and `/root/test.log`. The following is an example of part of the first screen using vi editor, showing the `/root/test.res`:

```
2 files to edit


Executing report switch getconfig Test Suite (switchgetportconfig) Mon Mar 19
10:16:19 EDT 2012 ...

Executing TEST SUITE report switch getconfig CASE
(switchgetportconfig.0x00066a00e3000100.i2c.extmgd.switchgetportconfig) retrieve
switch 0x00066a00e3000100 ...

TEST SUITE report switch getconfig CASE
(switchgetportconfig.0x00066a00e3000100.i2c.extmgd.switchgetportconfig) retrieve
switch 0x00066a00e3000100

        MTU                                 :  4096

        VL Capability                       :  1+1

        VL Credit Distribution Method       :  0

        Link Width                          :  1-4x

        Link Speed                          :  2.5-10Gb

        Node Description                    :  12200-18


PASSED

Summary:

count - configuration

    1 - Link Speed                          :  2.5-10Gb

    1 - Link Width                          :  1-4x

    1 - MTU                                 :  4096

    1 - VL Capability                       :  1+1

    1 - VL Credit Distribution Method       :  0

TEST SUITE report switch getconfig: 1 Cases; 1 PASSED

TEST SUITE report switch getconfig PASSED

Done report switch getconfig Test Suite Mon Mar 19 10:16:24 EDT 2012


Executing report switch hwvpd Test Suite (switchhwvpd) Mon Mar 19 10:17:12 EDT 2012
...

Executing TEST SUITE report switch hwvpd CASE
(switchhwvpd.0x00066a00e3000100.i2c.extmgd.switchhwvpd) retrieve switch
0x00066a00e3000100 ...

.

.

.
```

When you exit the vi editor, the following question is shown:

```
Would you like to remove test.res test.log test_tmp* and save_tmp

in /root ? [n]:
```

3.5 Install OFED+ Host Software on the Remaining Servers

FastFabric may now be used to install and configure the remaining hosts and verify overall operation of the fabric.

Note: The following procedure is for the OFED+ Host Software packaging of OFED+ or the True Scale Fabric Stack. FastFabric may also be used to install the True Scale Fabric Stack Tools on the remaining hosts when using other variations of OFED. In this case, OFED must be installed on each host manually.

1. **(All)** If the **Intel FastFabric Tools** menu is not displayed, type `fastfabric` and press **Enter**.
2. **(Linux)** Press **3**.

   Displays the **FastFabric IB Host Setup Menu** (Figure 10).

Figure 10. FastFabric IB Host Setup Menu

```
FastFabric IB Host Setup Menu

Host File: /etc/sysconfig/iba/hosts


Setup:

0) Edit Config and Select/Edit Host File     [ Skip  ]

1) Verify hosts pingable                     [ Skip  ]

2) Setup Password-less ssh/scp               [ Skip  ]

3) Copy /etc/hosts to all hosts              [ Skip  ]

4) Show uname -a for all hosts               [ Skip  ]

5) Install/Upgrade IB Software               [ Skip  ]

6) Configure IPoIB IP Address                [ Skip  ]

7) Build Test Apps and Copy to Hosts         [ Skip  ]

8) Reboot Hosts                              [ Skip  ]

Admin:

9) Refresh ssh Known Hosts                   [ Skip  ]

a) Rebuild MPI Library and Tools             [ Skip  ]

b) Run a command on all hosts                [ Skip  ]

c) Copy a file to all hosts                  [ Skip  ]

Review:

d) View iba_host_admin result files          [ Skip  ]


P) Perform the selected actions              N) Select None

X) Return to Previous Menu (or ESC)
```

1. Select items **0** through **2** and **4** through **8**.
2. Press **P**.

Note: If `/etc/hosts` will be used for name resolution (as opposed to using DNS), also select `Copy /etc/hosts to all hosts`

3. Perform the items selected using the following sections.

3.5.1 Edit Config and Select/Edit Host File

(All) The **Edit Config and Select/Edit Host File** selection will permit the hosts and FastFabric configuration files to be edited. When placed in the editor for `fastfabric.conf`, review all the settings. Especially review the `FF_IPOIB_SUFFIX`, `ff_host_basename_to_ipoib`, `ff_host_basename`,

FF_IPOIB_NETMASK, FF_PRODUCT, FF_PACKAGES, FF_INSTALL_OPTIONS, FF_UPGRADE_OPTIONS, and FF_ALL_ANALYSIS files. Refer to Appendix B for more information about fastfabric.conf.

Note: During setup of password-less ssh, FastFabric will provide the opportunity to enter the host root password interactively when needed. Therefore, it is recommended not to place it within fastfabric.conf file. If it is required to keep the root password for the hosts in the fastfabric.conf file, its recommended to change the fastfabric.conf permissions to be 0x600 (e.g. root-only access).

When placed in the editor for hosts, create the file with a list of the hosts names (the TCP/IP management network names) except the Fabric management node from which FastFabric is presently being run, Enter one host's name per line. For example:

```
host1

host2
```

Note: Do not list the Fabric management node itself (the node where FastFabric is currently running).

If additional Fabric Management Nodes are to be used, they may be listed at this time and FastFabric can aid in their initial installation and verification.

For further details about the file format, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

3.5.2 Verify hosts pingable

(All) The **Verify hosts pingable** selection will ping each selected host over the management network. If all hosts were found, continue to the next step. If some hosts were not found, abort out of the menu and review the following for those hosts which were not found:

- Host powered on and booted
- Host connected to management network
- Host management network IP address and network settings consistent with DNS or /etc/hosts
- Management node connected to the management network
- Management node IP address and network settings correct
- Management network itself up (switches, routers, etc)
- Correct set of hosts listed in the hosts file (the previous step may be repeated to review and edit the file as needed)?

3.5.3 Setup Password-less ssh/scp

(Linux) The **Setup Password-less ssh/scp** section will set up secure password-less ssh such that the Fabric Management Node can securely log in to all the other hosts as root through the management network without requiring a password.

Password-less ssh is required by FastFabric, MPI test applications and most versions of MPI (including OpenMPI, MVAPICH, and MVAPICH2).

3.5.4 Copy /etc/hosts to all hosts

(Linux) The **Copy /etc/hosts to all hosts** section will copy the /etc/hosts file on this host to all the other selected hosts.

Note: If DNS is being used, this step should be skipped.

Note: Typically, `/etc/resolv.conf` is set up as part of OS installation for each host. However, if `/etc/resolv.conf` was not setup on all the hosts during OS installation, the **FastFabric Copy a file to all hosts** operation could be used at this time to copy `/etc/resolv.conf` from the Fabric Management Node to all the other nodes.

3.5.5 Show uname -a for all hosts

(Linux) The **Show uname -a for all hosts** selection will show the OS version on all the hosts. Review the results carefully to verify all the hosts have the expected OS version. In typical clusters, all hosts will be running the same OS and kernel version.

If any hosts are identified with an incorrect OS version, the OS on those hosts should be corrected at this time and operation of this sequence should be aborted when prompted. As necessary, all the preceding setup steps should then be repeated for those hosts (there is no harm in repeating them for all the hosts).

3.5.6 Install/Upgrade Intel IB Software

(Host) The **Install/Upgrade Intel IB Software** selection will install the OFED+ Host Software on all the hosts. By default it will look in the current directory for the `IntelIB-Basic.DISTRO.VERSION.tgz` file. If the tarball is not found in the current directory, the installer application will prompt for input of a directory name where this file can be found.

Note: An initial installation will uninstall any existing OFED+ or stock OFED software. Initial installs must be performed when installing on a clean system or on a system which has stock OFED installed. For upgrading the fabric refer to Section 10.0, "Upgrade the Fabric" on page 117.

Perform the following steps to install the selected hosts:

1. The **"**Install/Upgrade Intel IB Software**"** will start with the following:
   System prompts:

```
Performing Host Setup: Install/Upgrade Intel IB Software

Do you want to use ./IntelIB-Basic.DISTRO.VERSION.tgz? [y]:
```

2. Press **ENTER** to accept the default (y).
   System prompts:

```
Would you like to do an upgrade/reinstall? [y]:
```

3. Type n and press **ENTER**.
   System prompts:

```
Would you like to do an initial installation? [n]:
```

4. Type **y** and press **ENTER** to proceed.
   System prompts:

```
You have selected to perform an initial installation

This will uninstall any existing Intel IB software on the selected nodes

Are you sure you want to proceed? [n]:
```

5. Type **y** and press **ENTER** to proceed.
   System prompts:

```
Executing: /sbin/iba_host_admin -f /etc/sysconfig/iba/hosts -d .. load

.

.

.

Hit any key to continue (or ESC to abort)...
```

6. Press any key to proceed.

   The selected hosts have completed rebooting the **FastFabric IB Host Setup Menu** appears. The installation is complete.

If any hosts fail to be installed, use the "View iba_host_admin result files" option to review the result files from the update. For more details, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

3.5.7 Configure IPoIB IP Address

(Host) The **Configure IPoIB IP Address** selection will create the `ifcfg-ib0` files on each host (previous non-OFED releases created the `ifcfg-ib1` file). The file will be created with a statically assigned IPv4 address. The IPoIB IP address for each host will be determined by the resolver (Linux* host command). If not found through the resolver, `/etc/hosts` on the given host will be checked.

3.5.8 Build Test Apps and Copy to Hosts

(Host) The **Build Test Apps and Copy to Hosts** selection will build the MPI and/or SHMEM sample applications on the IB Management Node and copy the resulting object files to all the hosts. This is in preparation for execution of MPI and/or SHMEM performance tests and benchmarks in a later step.

Note: This option is only available when using with the True Scale Fabric OFED+ packaging of OFED.

3.5.9 Reboot Hosts

(Linux) The **Reboot Hosts** selection will reboot all the selected hosts and ensure they go down and come back up (as verified through ping over the management network). When the hosts come back up, they will be running the newly installed Fabric software.

3.5.10 Refresh ssh Known Hosts

(Linux) The **Refresh ssh Known Hosts** will run the `setup_ssh -U "”` command to refresh the ssh known hosts list on this server for the Management Network. This may be used to update security for this host if hosts are replaced, reinstalled, renamed, or repaired.

3.5.11 Rebuild MPI Library and Tools

(Host) The **Rebuild MPI Library and Tools** will rebuild the MPI Library itself and related tools (such as `mpirun`) and install the resulting rpms on all the hosts. This will be performed using the `do_build` tool supplied with the MPI Source. When rebuilding MPI, `do_build` will prompt the user for selection of which MPI (openmpi, mvapich or mvapich2) to rebuild and provide choices as to which available compiler to use. Refer to the *Intel® True Scale Adapter Hardware User Guide* and *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.

Note: This option is only available when using with the True Scale Fabric OFED+ packaging of OFED.

3.5.12 Run a command on all hosts

(**Linux**) If there are any other setup operations which need to be performed on all hosts, they may be performed using the **Run a command on all hosts** option. Each time this is executed a Linux* shell command (or sequence of commands separated by semicolons) may be specified to be executed against all selected hosts.

Note: It is recommended at this time to run the "date" command to verify the date and time is consistent on all hosts. If needed **Copy a file to all hosts** option may be used to copy the appropriate files to all hosts to enable and configure NTP.

3.5.13 Copy a file to all hosts

(Linux) The **Copy a file to all hosts** will run the `scpall` command. A file on the local host may be specified to be copied to all selected hosts.

3.5.14 View iba_host_admin result files

(All) The **View iba_host_admin result file** permits viewing of the `test.log` and `test.res` files that reflect the results from `iba_host_admin` runs (such as for installing Fabric software or rebooting all hosts per menu items above). The user is also given the option to remove these files after viewing them.

If not removed, subsequent runs of `iba_chassis_admin`, `iba_host_admin` or `iba_switch_admin` from within the current directory will continue to append to these files.

3.6 Verify OFED+ Host Software on the Remaining Servers

Upon completion of the preceding sections, the hosts are all booted, installed and operational. The subsequent steps will verify the operation of the hosts and fabric.

1. **(All)** If the **FastFabric IB Host Verification/Admin Menu** is not displayed type `fastfabric` and press **Enter**.
2. **(All)** Press **4**.

   Displays the **FastFabric IB Host Verification/Admin Menu** (Figure 11).

Figure 11. FastFabric IB Host Verification/Admin Menu

```
FastFabric IB Host Verification/Admin Menu

Host List: /etc/sysconfig/iba/allhosts


Validation:

0) Edit Config and Select/Edit Host File     [ Skip  ]

1) Summary of Fabric Components              [ Skip  ]

2) Verify hosts pingable, sshable and active [ Skip  ]

3) Perform Single Host verification          [ Skip  ]

4) Verify IB Fabric status and topology      [ Skip  ]

5) Verify Hosts see each other               [ Skip  ]

6) Verify Hosts ping via IPoIB               [ Skip  ]

7) Refresh ssh Known Hosts                   [ Skip  ]

8) Check MPI Performance                     [ Skip  ]

9) Check Overall Fabric Health               [ Skip  ]

a) Start or Stop Bit Error Rate Cable Test   [ Skip  ]

Admin:

b) Generate all Hosts Problem Report Info    [ Skip  ]

c) Run a command on all hosts                [ Skip  ]

Review:

d) View iba_host_admin result files          [ Skip  ]


P) Perform the selected actions              N) Select None

X) Return to Previous Menu (or ESC)
```

3. Select the items **0** through **8** in the **Validation** section of the menu
4. Press **P**.

3.6.1 Edit Config and Select/Edit Host File

(**All**) The **Edit Config and Select/Edit Host File** section will permit the hosts, ports, and FastFabric configuration files to be edited. When placed in the editor for `fastfabric.conf`, review all the settings. Especially review the `FF_TOPOLOGY_FILE`, `FF_IPOIB_SUFFIX`, `ff_host_basename_to_ipoib`, and `ff_host_basename`. Refer to Appendix B for more information about `fastfabric.conf`. If required, a FastFabric topology file may be created as `/etc/sysconfig/iba/topology.0:0.xml` to describe the intended topology of the fabric and augment assorted fabric reports with customer-specific information such as

cable labels and additional details about nodes, SMs, links, ports and cables. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information about topology verification files.

Review the following parameters which will be used for overall fabric health checks: `FF_ANALYSIS_DIR`, `FF_ALL_ANALYSIS`, `FF_FABRIC_HEALTH`, `FF_CHASSIS_CMDS`, `FF_CHASSIS_HEALTH`, and `FF_ESM_CMDS`. `FF_ALL_ANALYSIS` should be updated to reflect the type of SM (esm or hostsm).

When placed in the editor for ports, review the file. For typical single-subnet clusters, the default of 0:0 may be used. This will use the first active port on the Fabric Management node to access the fabric. For more information on configuring a cluster with multiple subnets, see Appendix C. For further details about the file format, refer to the Selection of Ports section in the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

When placed in the editor for allhosts, create the file with the Fabric Management node's hosts names (the TCP/IP management network names) (shown as `mgmthost`, for example) and include the hosts file previously created, enter one per line. For example:

```
mgmthost

include /etc/sysconfig/iba/hosts
```

For further details about the file format refer to the Selection of Hosts section in the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

3.6.2 Summary of Fabric Components

(**All**) The **Summary of Fabric Components** selection will provide a brief summary of the counts of components in the fabric, including how many switch chips, hosts, and links are in the fabric. It will also indicate if any 1x links were found (which could indicate a poorly seated or bad cable). Review the results against the expected configuration of the cluster.

If components are missing or 1x links are found, they should be corrected. Subsequent steps will aid in locating any 1x links.

3.6.3 Verify hosts pingable, sshable and active

(**All**) The **Verify hosts pingable, sshable and active** selection will verify each host and provide a concise summary of the bad hosts found.Interactive prompts allow the user to select ping, ssh and port active verification. After completion of this test the user will have the option of using the resulting good hosts file for the remainder of their operations within this TUI session.

3.6.4 Perform Single Host verification

(**All**) The **Perform Single Host verification** uses the `iba_verifyhosts` to perform a single host test on all hosts. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for information on `iba_verifyhosts`.

3.6.5 Verify IB Fabric status and topology

(**All**) The **Verify IB Fabric status and topology** selection can run the following checks:

- Perform a fabric error analysis

- Clear error counters after generating the report
- Perform a fabric link speed error analysis
- Check for links that are configured to run slower than supported
- Check links that are connected with mismatched speed potential
- Verify the fabric topology
- Verify all aspects of the topology including links, nodes, and sms
- Include unexpected devices in the punch-list

3.6.6 Verify Hosts see each other

(**Host**) The **Verify Hosts see each other** selection will verify that each host can see all the others through queries to the Subnet Administrator.

3.6.7 Verify Hosts ping via IPoIB

(**Host**) The **Verify Hosts ping via IPoIB** selection will verify that IPoIB is properly configured and running on all the hosts. This is accomplished through the Fabric management node pinging each host through IPoIB.

Note: Use of this operation requires that IPoIB be enabled on the Fabric Management Node as well as each host selected for verification.

1. The management host needs to have ipoib configured
2. Depending on the MTU of the fabric, this may not be successful.

3.6.8 Refresh ssh Known Hosts

(**Linux**) The **Refresh ssh Known Hosts** selection will refresh the ssh `known_hosts` file on the Fabric management node to include the IPoIB hostnames of all the hosts.

Note: Use of this operation requires that IPoIB be enabled on the Fabric Management Node as well as each host selected for verification.

3.6.9 Check MPI Performance

(**Host**) The **MPI Performance** selection will do a quick check of PCIe and MPI performance through end-to-end latency and bandwidth tests.

Note: This option is available for the OFED+, but is not presently available for other packagings of OFED.

When MPI Performance is selected it displays a prompt as follows:

```
Test Latency and Bandwidth deviation between all hosts? [y]:
```

At the prompt press Enter to select default (y)

This displays the results of pairwise analysis of latency and bandwidth for the selected hosts and reports pairs outside an acceptable tolerance range. By default performance is compared relative to other hosts in the fabric (with the assumption that all hosts selected for a given run should have comparable fabric performance). Failing hosts will be clearly indicated.

The results are also written to the `test.res` file which may be viewed through the **View iba_host_admin result files** option. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

If any hosts fail, carefully examine the failing hosts to verify the HCA models, PCIe slot used, BIOS settings and any motherboard jumpers related to devices on PCIe buses or slot speeds. Also verify the HCA and riser cards are properly seated.

The bandwidth that is reported should also be checked against the practical PCIe speeds in the Performance Impact table (Table 3). If all pairs are not in the expected performance range, carefully examine all hosts to verify the HCA models, PCIe slot used, BIOS settings and any motherboard jumpers related to devices on PCIe buses or slot speeds.

Table 3. Performance Impact

PCIe Speed	Fabric Speed	Theoretical Max	Practical Bandwidth
PCIe Gen 2 x16	QDR	8000MB/sec	5200-6000 MB/sec
PCIe Gen 1 x16	QDR	4000MB/sec	2600-3000 MB/sec
PCIe Gen 2 x8	QDR	4000MB/sec	2600-3000 MB/sec
PCIe x8	DDR	2000MB/sec	1300-1500 MB/sec
PCIe x4	DDR	1000MB/sec	800-900 MB/sec
PCIe x8	SDR	1000MB/sec	900-1000 MB/sec
PCIe x4	SDR	1000MB/sec	800-900 MB/sec
133	SDR	1064MB/sec	800-900 MB/sec
100	SDR	8000MB/sec	600-680 MB/sec
66	SDR	532MB/sec	400-450 MB/sec

3.6.10 Check Overall Fabric Health

(**ALL**) The **Check Overall Fabric Health** selection will permit the present fabric configuration to be baselined for use in future fabric health checks. This should be performed after configuring any additional Fabric management Nodes. Refer to "Configure and Initialize Health Check Tools" on page 63 for more information.

3.6.11 Start or Stop Bit Error Rate Cable Test

(**ALL**) The **Start or Stop Bit Error Rate Cable Test** selection performs host and/or ISL cable testing. The test allows for starting and stopping an extended Bit Error Rate test. Intel recommends that this test be run for seven hours for a thorough test.

3.6.12 Generate all Hosts Problem Report Info

(Host) The **Generate all Hosts Problem Report Info** will run the captureall command to collect configuration and status information from all hosts and generate a single *.tgz file which can be sent to the Support Representative.

Based on the answer to the prompt shown below, various levels of detail about the fabric can be included in the capture.

```
Capture detail level (1-Normal, 2-Fabric, 3-Fabric+FDB, 4-Analysis):
```

The Details levels are:

`1-Normal` — obtains local information from each host

`2-Fabric` — in addition to "Normal", also obtains basic fabric information by queries to the SM and fabric error analysis using `iba_report`.

3-Fabric+FDB — in addition to "Fabric", also obtains all the switch forwarding tables and IB multicast membership lists from the SM.

4-Analysis — in addition to "Fabric+FDB", also obtains all_analysis results. If all_analysis has not yet been run, it is run as part of the capture.

Note: Detail levels 2-4 can be used when fabric operational problems occur. If the problem is most likely node specific, detail level 1 should be sufficient. Detail levels 2-4 require an operational Fabric Manager. Typically your support representative will request a given detail level. If a given detail level takes excessively long or fails to be gathered, try a lower detail level.

For detail levels 2-4, the additional information is only gathered on the node running the captureall command. The information is gathered for every fabric specified in the /etc/sysconfig/iba/ports file.

3.6.13 Run a command on all hosts

(Linux) The **Run a command on all hosts** will run the cmdall command. A Linux* shell command (or sequence of commands separated by semicolons) may be specified to be executed against all selected hosts.

3.6.14 View iba_host_admin result files

(All) The **View iba_host_admin result files** permits viewing of the test.log and test.res files which reflect the results from iba_host_admin runs (such as those for installing Fabric software or rebooting all hosts per menu items above). The user is also given the option to remove these files after viewing them.

If not removed, subsequent runs of iba_chassis_admin, iba_host_admin or iba_switch_admin from within the current directory will continue to append to these files.

3.7 Installation of additional Fabric Management Nodes

If the fabric is to have more than one Fabric Management Node, the setup of the additional management nodes may be completed using the **Installation of additional Fabric Management Nodes**. The previous steps will have performed basic software installation, setup and verification on those nodes. Now the management software itself must be installed and configured.

Note: The following steps assume a symmetrical configuration where all Fabric management nodes have the same connectivity and capabilities. In asymmetrical configurations where the Fabric management nodes are not all connected to the same set of management networks and subnets, the files copied to each management node may need to be slightly different. For example, configuration files for fabric_analysis may indicate different port numbers, or host files used for FastFabric and MPI may need to list different hosts. For multiple-subnet configurations, refer to "Multi-Subnet Fabrics" on page 151.

Repeat the following steps on each additional Fabric Management Node:

1. (All) Upgrade the IntelIB- Basic to IntelIB-IFS software to add additional components using the procedure documented in Section 11.0, "Upgrade from OFED+ Host Software to Intel IFS" on page 121. The Fabric Management node must have at least FastFabric, the True Scale Fabric Stack and should have IPoIB installed and configured. For MPI clusters the Fabric Management node should also include at least OFED openmpi, OFED mvapich, or OFED mvapich2, and if the user

desires to rebuild MPI itself, the OFED True Scale Fabric Development package and MPI Source packages will also be required.

Note: Do not uninstall or replace existing configuration files which were previous created, especially IPoIB-related configuration files.

2. (**All**) Copy the FastFabric configuration files from the initial Fabric Management Node. At least the following files should be copied:

```
/etc/sysconfig/fastfabric.conf

/etc/sysconfig/iba/ports

/etc/sysconfig/iba/topology*.xml

/etc/sysconfig/iba/hosts

/etc/sysconfig/iba/allhosts

/etc/sysconfig/iba/ibnodes

/etc/sysconfig/iba/chassis
```

   After copying the files, edit the hosts and allhosts files such that the file on each Fabric Management node omits itself from the hosts files (but lists all other Fabric Management nodes) and specifies itself in the allhosts file.

   See Appendix B for a complete list of FastFabric configuration files.

3. (All) If the Fabric Manager is also going to be run, copy the FM configuration file (`/etc/sysconfig/ifs_fm.xml`) from the initial Fabric Management Node. After copying the file, edit the file on each Fabric Management node as needed.

   Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for more information on how to configure the FM.

4. (**Linux**) Perform **Setup Password-less ssh/scp** option in the **Host Setup via FastFabric** menu and **Refresh ssh Known Hosts** option in the **Host Admin via FastFabric** menu.

3.8 Configure and Initialize Health Check Tools

For more information on the health check tools, see the detailed discussion in *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*. The Health check tools may be run on one or more Fabric management nodes within the cluster. This procedure should be followed on each Fabric management node from which the health check tools will be used.

1. (**All**) Edit `fastfabric.conf` and review the following parameters: `FF_ANALYSIS_DIR`, `FF_ALL_ANALYSIS`, `FF_FABRIC_HEALTH`, `FF_CHASSIS_CMDS`, `FF_CHASSIS_HEALTH`, and `FF_ESM_CMDS`. `FF_ALL_ANALYSIS` should be updated to reflect the type of SM (esm or hsm).
2. (**All**) If using Embedded SM(s) in True Scale Fabric Chassis, create `/etc/sysconfig/iba/esm_chassis` listing the chassis which are running SMs.

   Create the file with a list of the chassis names (the TCP/IP Ethernet management port names assigned) or IP addresses (use of names is recommended). Enter one name or IP address per line. For example:

```
Chassis1

Chassis2
```

For further details about the file format, refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

3. (**All**) Perform a health check using: `all_analysis -e`. If any errors are encountered, resolve the errors and rerun `all_analysis -e` until a clean run occurs.
4. (**All**) Create a cluster configuration baseline using: `all_analysis -b`.

Note: This may also be done using the FastFabric menu system by selecting **Check Overall Fabric Health** and enter `y` to `Baseline present configuration? [n]`**:**

5. (**All**) If required, schedule regular runs of `all_analysis` through cron or other mechanisms. Refer To the Linux* OS documentation for more information on cron. Also refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information about `all_analysis` and its automated use.

3.9 Running High Performance Linpack

As part of the installation process, a set of common MPI benchmarks have been installed. One of the more popular measures of overall performance is High Performance Linpack (HPL). This is the application used to rate systems on the Top 500 list. The steps allow some initial runs of HPL to be made and provide some initial baseline numbers. The defaults provided should perform within 10 — 20% of optimal HPL results for the cluster. Tuning for that additional 10 — 20% is beyond the scope of this document.

6. (**Host**) To run HPL, first select a configuration file appropriate to your cluster. It is best to start with a small configuration to verify HPL has been properly compiled:

```
cd /opt/iba/src/mpi_apps

./config_hpl 2t
```

This command will configure a two process test run of HPL.

7. (**Host**) Create the file `/opt/iba/src/mpi_apps/mpi_hosts` listing the host names of all the hosts.
8. (**Host**) Run HPL:

```
./run_hpl 2
```

Since this is a very small problem size, the performance of the run will be much lower than the potential of the machine. So do not worry about performance, just whether or not the run was successful.

At this point the user is ready to move onto full scale HPL runs. Assorted sample `HPL.dat` files are provided in `/opt/iba/src/mpi_apps/hpl-config`. These files are a good starting point for most clusters and should get within 10 — 20% of the optimal performance for the cluster. The problem sizes used assume a cluster with 1GB of physical memory per processor (e.g., for a 2 processor node, 2GB of node memory is assumed). For each cluster size, 4 files are provided:

t - a very small test run (5000 problem size)

s - a small problem size on the low end of optimal problem sizes

m - a medium problem size

l - a large problem size

These can be selected using `config_hpl`. The following command displays the pre-configured problem sizes available:

```
./config_hpl
```

For example, to do a small run for a 256 processor cluster (i.e., 128 nodes of dual CPU systems):

1. Type `./config_hpl 256s` and press **Enter**.
2. Type `./run_hpl 256` and press **Enter**.

   During these runs the user should use top on a node to monitor memory and CPU usage. The xhpl should use 98 — 99% of the CPU. If any other processes are taking more than 1 — 2%, review the host configuration and stop these extra processes if possible. HPL is very sensitive to swapping. If a lot of swapping is seen, and xhpl is dropping below 97% for long durations, this may indicate a problem size that is too large for the memory and OS configuration.

   At this point the user can continue to tune HPL to refine performance. Parameters in `HPL.dat` can all affect HPL performance. In addition, the selection of compiler and BLAS Math library may also significantly affect performance. The new `HPL.dat` files may be placed in `/opt/iba/src/mpi_apps/hpl-config`. Use `config_hpl` to select them and copy them to all nodes in the run. Alternately, `scpall` may be used to copy the file to all nodes. Refer to *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information on `scpall`.

§ §

4.0 Install OFED+ Host Software

The Intel® OFED+ Host Software package has a Text User Interface (TUI) for easy installation of the software. Use this method of software installation if you downloaded the OFED+ Host Software package for installation on a host. Appendix A provides a checklist for tracking the installation process.

Note: This method is suitable for use on small clusters. For large clusters it is recommended to purchase the IFS and install the IFS on the Fabric Management Node. Once the IFS has been installed on the Fabric Management Node, FastFabric may be used to install multiple hosts with the OFED+ Host Software simultaneously.

- Use the package file, `IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz`.
- Using the menus, install the required components (at least **OFED IB Stack**, **True Scale HCA Libs**, **IB Tools**, and **OFED IB Fabric Development**) as described in the following procedures.
- FastFabric and IFS FM are displayed as not available

4.0.1 Download the Intel® OFED+ Host Software

Use the following procedure to download the Intel® OFED+ Host Software.

1. Open downloadcenter.intel.com/.
2. Type Intel True Scale in the ***Search downloads*** field.
3. Click the **Search** button.
4. Select the **Intel® True Scale Fabric Host Channel Adapter Host Drivers & Software vX.X.X.X.X** with the correct version number, in the ***XX* result(s) matching: *"Intel True Scale"* sorted by relevance** section of the web page.
5. Click **Download** beside the IntelIB-Basic tar file with the correct distribution for the installation.
6. Read and agree to the *Intel Software License Agreement*.
7. Save the tar file to the local drive.

4.0.2 Unpack the Tar File

Use the following procedure to unpack the `IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz` tar file.

1. Copy the tar file to the `/root` directory.
2. Change directory to `/root`.

```
cd /root
```

3. Unpack the `IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz` tar file to the `IntelIB-Basic.`*`DISTRO`*`.`*`VERSION` directory.

```
tar xvfz IntelIB-Basic.DISTRO.VERSION.tgz
```

4.1 Install OFED+ Host Software

To install the OFED+ Host Software, perform the following procedure:

1. Change directory to `IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*.

```
cd IntelIB-Basic.DISTRO.VERSION
```

2. Start the `INSTALL` TUI.

```
./INSTALL
```

Note: If you need 32-bit support on 64-bit OSs, enter the following command:

```
./INSTALL --32bit
```

The **Intel IB *VERSION* Software** main menu appears (Figure 12).

Figure 12. Intel IB Main Menu (Example)

```
Intel IB VERSION Software


   1) Install/Uninstall Software
   2) Reconfigure OFED IP over IB
   3) Reconfigure Driver Autostart
   4) Update HCA Firmware
   5) Generate Supporting Information for Problem Report
   6) FastFabric (Host/Chassis/Switch Setup/Admin)


   X) Exit
```

3. Press **1** to select `Install/Uninstall Software`.

Screen 1 of 3 of the **Intel IB Install** menu appears (Figure 13).

Figure 13. Intel IB Install Menu (Screen 1 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 1 of 3):

0) OFED IB Stack       [   Install   ][Available] VERSION

1) True Scale HCA Libs [   Install   ][Available] VERSION

2) OFED mlx4 Driver    [   Install   ][Available] VERSION

3) IB Tools            [   Install   ][Available] VERSION

4) OFED IB Development [   Install   ][Available] VERSION

5) FastFabric          [Don't Install][Not Avail]

6) OFED IP over IB     [   Install   ][Available] VERSION

7) OFED IB Bonding     [   Install   ][Available] VERSION

8) OFED SDP            [   Install   ][Available] VERSION

9) IFS FM              [Don't Install][Not Avail]

a) MVAPICH (gcc)       [   Install   ][Available] VERSION

b) MVAPICH2 (gcc)      [   Install   ][Available] VERSION

c) OpenMPI (gcc)       [   Install   ][Available] VERSION

d) MVAPICH/PSM (gcc)   [   Install   ][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

Note: **True Scale HCA Libs** contains the enhanced HCA driver optimized stack for MPI (PSM) on HCAs and OpenMPI, as well as user tools.

Note: **OFED IB Bonding** will show as `[Not Avail]` when installing the software on OSs that have bonding modules in the OS installed software.

4. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
5. Press **N** to go to the next screen.

   Screen 2 of 3 of the **Intel IB Install** menu appears (Figure 14).

Figure 14. Intel IB Install Menu (Screen 2 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 2 of 3):

0) MVAPICH/PSM (PGI)    [   Install   ][Available] VERSION

2) MVAPICH/PSM (Intel) [   Install   ][Available] VERSION

3) MVAPICH2/PSM (gcc)  [   Install   ][Available] VERSION.DISTRO

4) MVAPICH2/PSM (PGI)  [   Install   ][Available] VERSION.DISTRO

5) MVAPICH2/PSM (Intel)[   Install   ][Available] VERSION.DISTRO

6) OpenMPI/PSM (gcc)    [   Install   ][Available] VERSION

7) OpenMPI/PSM (PGI)    [   Install   ][Available] VERSION

8) OpenMPI/PSM (Intel) [   Install   ][Available] VERSION

9) Intel SHMEM        [   Install   ][Available] VERSION.DISTRO

a) MPI Source          [   Install   ][Available] VERSION

b) OFED uDAPL          [   Install   ][Available] VERSION

c) OFED RDS            [   Install   ][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

6. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
7. Press **N** to go to the next screen.

   Screen 3 of 3 of the **Intel IB Install** menu appears (Figure 15).

Figure 15. Intel IB Install Menu (Screen 3 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 3 of 3):

0) OFED SRP            [   Install   ][Available] VERSION

1) OFED SRP Target     [Don't Install][Available] VERSION

2) OFED iSER           [Don't Install][Not Avail]

3) OFED iWARP          [Don't Install][Available] VERSION

4) OFED Open SM        [Don't Install][Available] VERSION

5) OFED NFS RDMA       [Don't Install][Not Avail]

6) OFED Debug Info     [Don't Install][Not Avail]

N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

8. Review the items to be installed; the default value is in brackets (Install or Don't Install). To change a value, type the alphanumeric character associated with the item.
9. Press **P** to perform the selected actions from all three screens.

   The system prompts:

```
About to Uninstall previous InfiniBand Software Installations...

Hit any key to continue...
```

10. Press any key to proceed with the installation.
11. The following system prompts are displayed. For each prompt, select the default by pressing ENTER.

```
Preparing OFED VERSION release for Install...

Rebuild OFED SRPMs (a=all, p=prompt per SRPM, n=only as needed?) [n]:


Installing OFED IB Stack VERSION release...

Permit non-root users to query the fabric? [y]:
```

Note: If you have had a previous version of the software installed and are installing this version after uninstalling a previous version, you might see the following prompt.

```
You have memory locking limits entries for IB drivers from an earlier install

Do you want to keep //etc/security/limits.conf? [y]:
```

```
Enable OFED SMI/GSI renice (RENICE_IB_MAD))? [y]:


Single Port Mode reallocates all Intel HCA resources to HCA Port 1.

Enable Intel HCA Single Port Mode? [y]:
```

Note: Selecting the default by pressing **enter** causes the dual-port HCAs to act as single-port cards with only port 1 enabled. Enabling `Intel HCA Single Port Mode` increases performance for environments where the second port is not connected.

Note: If there was a previous version of the software on this host that was uninstalled, you might see the following prompt.

```
You have a Distributed SA configuration file from an earlier install

Do you want to keep //etc/sysconfig/iba/dist_sa.conf? [y]:


Installing OFED IP over IB VERSION release...

Enable IPoIB Connected Mode (SET_IPOIB_CM)? [y]:
```

The system searches for ifcfg files, which contain IPv4 port IP and netmask addresses.

12. Perform one of the actions in the following table.

If	Then
The system finds the ifcfg files and you want to keep the files.	Answer yes to the system prompt question "`Do you want to keep OFED IP over IB ifcfg files.`"
The system finds the ifcfg files and you do not want to keep the files, you want to input new IPv4 addresses.	Answer no to the system prompt question "`Do you want to keep OFED IP over IB ifcfg files`" and proceed through the system prompts to set up the IP addresses.
The system does not find the ifcfg files and you want to input new IPv4 addresses.	Proceed through the system prompts to set up the IP addresses.
The system does not find the ifcfg files and you do not want to input new IPv4 addresses at this time.	The system prompt following this table is shown.
You are using IPv6 addresses	The system prompt following this table is shown.

The system prompts:

```
Enable OFED SRP High Availability deamon (SRPHA_ENABLE)? [n]:
```

13. Press **Enter** to select default (n).

The **IB Autostart Menu** appears (Refer to Figure 16).

Figure 16. Intel IB Autostart Menu

```
Intel IB Autostart (VERSION release) Menu


Please Select Autostart Option:

0) OFED IB Stack (openibd)            [Enable ]

1) OFED mlx4 Driver (openibd)         [Enable ]

2) IB Port Monitor (iba_mon)          [Disable]

3) S20 Port Tuner (s20tune)           [Disable]

4) Distributed SA (dist_sa)           [Disable]

5) OFED IP over IB (openibd)          [Enable ]

6) OFED SDP (openibd)                 [Enable ]

7) IFS FM (ifs_fm)                    [Enable ]

8) OFED RDS (openibd)                 [Enable ]

9) OFED SRP (openibd)                 [Enable ]


P) Perform the autostart changes

S) Autostart All                      R) Autostart None

X) Return to Previous Menu (or ESC)
```

14. Review the items to be autostarted; the default value is in brackets (`Enable` or `Disable`). To change a value, type the alphanumeric character associated with the item.

    Intel recommends leaving all of the autostart selections as default, unless one of the following scenarios apply:

- If FastFabric will not monitor the fabric health, performance, and/or check the fabric for errors, change `IB Port Monitor (iba_mon)` to `Enable`.
- Intel recommends changing `Distributed SA (dist_sa)` to `Enable` when installing software in mesh/torus fabrics, or when using Virtual Fabrics with Intel HCAs, `dist_sa` must be enabled on management nodes only. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.

15. Press **P** to perform the selected actions from all three screens.

    The system prompts:

```
Hit any key to continue...
```

16. Press any key.

    The system prompts with one of the following:

    — The following lines appear stating the firmware is not required when using HCAs.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log
```

```
Firmware is not required for the Intel HCA(s) in this system.


Press any key to continue.
```

Skip to Step 21.

— The following lines appear showing the number of HCAs found.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log

One HCA was found:
```

When one or more HCA is found, the system prompts with each HCA name and the firmware version installed, and if there is an update available or not. If a firmware update is available or the firmware is up to date, the system prompts to update, install different firmware, or do nothing. Only Connect-X HCAs will have firmware available. Refer to the following bulleted list for an example of the system prompt for each scenario:

— An update is available (Example):

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0: Update to 2.5.0 available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

— The firmware is up to date (Example):

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.5.0: Okay.




To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

— No firmware is available. This displays if the HCA is not a Connect-X HCA (Example).

```
0: MT_0390140002 (MHGA28-XTC/A4/A0) Firmware : No firmware available.

        Contact your vendor for firmware updates for this HCA.


No firmware available for HCAs in your system.

        Contact your vendor for firmware updates for this system.

Press any key to continue.
```

17. Perform one of the actions in the following table.

If	Then
No firmware is available	Skip to Step 21.
You need to upgrade the firmware	Proceed with Step 18.
You do not need to upgrade the firmware	Skip to Step 20.

18. Select a number corresponding to the HCA that needs to be upgraded.

The system prompts (Example):

```
MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0

The following firmware revision(s) are available for this HCA:

0: MT_04A0110002: standard firmware

Select firmware version, or Q to cancel:
```

19. Select the number corresponding to the firmware revision required for the HCA.

    The firmware is installed on the HCA.

    The system prompts:

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0: Update to 2.5.0 available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

If	Then
You need to upgrade the firmware in another HCA	Repeat Step 18 and Step 19.
You do not need to upgrade the firmware on any other HCAs	Continue with Step 20.

20. Press **Q**

    The installation completes and returns to the main menu

    Skip to Step 23

21. Press any key.

    The system prompts:

```
A System Reboot is recommended to activate the software changes

Hit any key to continue...
```

22. Press any key.

    The installation completes and returns to the main menu:

23. Press **X** to exit.
24. If installing IPoIBV6 proceed to Section 4.1.1, "Install IPoIB IPV6" on page 75, then return to this procedure Step 25. If not installing IPoIBV6, reboot the server.
25. Repeat this procedure for each host.

4.1.1 Install IPoIB IPV6

To install IPoIBV6 on the management node use the following procedures for the OS on the Fabric management node.

4.1.1.1 On Red Hat*:

1. Edit file `/etc/sysconfig/network` to add the following line:

```
NETWORKING_IPV6=yes
```

2. Edit file `ifcfg-`*if-name* to add the following lines:

```
IPV6INIT=yes

IPV6ADDR="ipv6addr/prefix-length"
```

Ipv6 address should look like the following

```
3ffe::6/64
```

3. Reboot the server

4.1.1.2 On SUSE* Enterprise:

1. Edit `ifcfg-ifname` to add the following line:

```
IPADDR="ipv6addr/prefix-length"
```

Ipv6 address should look like the following:

```
3ffe::6/64
```

2. Reboot the server.

§ §

5.0 Install Intel® SHMEM

This section provides procedures for installing SHMEM on the cluster.

5.0.1 Requirements

SHMEM requires the following:

- Red Hat* Enterprise Linux* (RHEL). Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for the latest supported OS releases.
- Every node must have a Intel HCA.
- Intel and AMD x86 processors running in 64-bit mode are supported
- Host systems are expected to have reasonable amounts of RAM, typically 1 GB or more per processor core
- One (or more) Message Passing Interface (MPI) implementations are required and Performance Scaled Messaging (PSM) support must be enabled within the MPI. The following MPI implementations are supported:
  - OpenMPI
  - MVAPICH
  - MVAPICH2

5.1 Install SHMEM

To install the SHMEM, perform the following procedure:

1. Perform normal installation using one of the following sections:.
   - For IFS Installation use Section 3.0
   - For OFED+ Host Software installation use Section 4.0
2. In step 6 of both procedures, when reviewing the items to be installed, ensure that the value for line `8) SHMEM` is `Install` as shown in Figure 17

Figure 17. Intel IB Install Menu (Screen 2 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 2 of 3):

0) MVAPICH/PSM (PGI)    [   Install   ][Available] VERSION.DISTRO

1) MVAPICH/PSM (Intel)  [   Install   ][Available] VERSION.DISTRO

2) MVAPICH2/PSM (gcc)   [   Install   ][Available] VERSION

3) MVAPICH2/PSM (PGI)   [   Install   ][Available] VERSION

4) MVAPICH2/PSM (Intel)[   Install   ][Available] VERSION

5) OpenMPI/PSM (gcc)    [   Install   ][Available] VERSION.DISTRO

6) OpenMPI/PSM (PGI)    [   Install   ][Available] VERSION.DISTRO

7) OpenMPI/PSM (Intel)  [   Install   ][Available] VERSION.DISTRO

8) Intel SHMEM        [   Install   ][Available] VERSION.DISTRO

9) MPI Source           [   Install   ][Available] VERSION.DISTRO

a) OFED uDAPL           [   Install   ][Available] VERSION.DISTRO

b) OFED RDS             [   Install   ][Available] VERSION.DISTRO


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

SHMEM is installed with the rest of the installation.

For information on SHMEM configuration and use, refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

§ §

6.0 Install Intel OFED+ Host Software Using Rocks

Rocks is a distribution designed for managing clusters from the San Diego Supercomputer Center (SDSC).

Rocks+ High Performance Computing (HPC), available from StackIQ, is the commercial edition of Rocks that provides an end-to-end cluster operating environment to manage the Linux* operating system, cluster management middleware, libraries, compilers, and monitoring tools.

Rocks is a way to manage the "kickstart automated installation" method created by Red Hat*. By using the Rocks conventions, the installation process can be automated for clusters of any size. A Roll is an extension to the Rocks base distribution that supports different cluster types or provides additional capabilities.

Intel extends the normal Rocks compute node appliance `.xml` file by adding two functions; one function installs the OFED+ Host Software, and the other function loads the drivers after the machine is rebooted.

Note: A kickstart is no longer used to reboot the machine.

Note: Only one OFED+ Roll can be enabled at a time due to compatibility issues

6.1 Install Front-end and Compute Nodes

Rocks is based on a set of kickstart graphs (`.xml` files) that tell the frontend node which pieces need to be installed on which type of compute nodes. The frontend node installs from local RPMs, then the compute nodes collect the RPMs from the frontend. Install a Rocks frontend node first, if you do not already have one.

To install a Rocks frontend node:

1. Open downloadcenter.intel.com/.
2. Type Intel True Scale in the ***Search downloads*** field.
3. Click the **Search** button.
4. Select the **Intel® True Scale Fabric Host Channel Adapter Host Drivers & Software vX.X.X.X.X** with the correct version number, in the ***XX* result(s) matching: *"Intel True Scale"* sorted by relevance** section of the web page.
5. Click **Download** by the OFED+ Roll image file (`intel_ofed-`*`VERSION`*`.x86_64.disk1.iso`) for the Rocks version required.
6. Read and agree to the *Intel Software License Agreement*.
7. Save the `.iso` image to the local drive.
8. Burn the `.iso` image to a CD.
9. Download the required rolls from the Rocks web site: http://www.stackiq.com
10. Follow the links to get the following `.iso` images that can be burned to a CD or DVD:

- Rocks image (boot up to begin installation)
- OS Roll (available from RedHat, equivalent to the RHEL6 installation DVD)
- Other Rolls required by your system

Note that you may also need updates; look for the latest files with the service-pack prefix. Make sure you downloaded the .iso images correctly; verify by checking the md5 checksum from the web site.

11. Burn the .iso image(s) to separate CDs or DVDs.
12. Build the frontend node with the CDs and/or DVDs (Steps 8. and 9):

    Insert the Rocks CD into your frontend node. After the frontend boots from the CD, follow the instructions on the screen. Insert the OS Roll, the Intel® OFED+ Roll, and any other of the Rocks Rolls you need when prompted.

    When the build is complete the frontend node will reboot automatically. The OFED+ software installation completes during this reboot.

Note: Deselect OFED and HPC when Installing Rocks.

Note: If using your own RHEL6 roll, deselect the OS from the Rocks disk.

13. Reboot the frontend node at the login screen to allow OFED+ to start.
14. Install the compute nodes. Login to the frontend node as a root user, and run the command:

```
# insert-ethers
```

    This command launches a program that captures compute node DHCP requests and puts the information into the Rocks MySQL database. Follow the instructions on the Rocks web site: http://www.stackiq.com

15. Reboot the compute nodes to allow OFED+ to start.

```
# rocks run host reboot
```

16. Once Rocks is up and running, test the rocks cluster according to your own testing procedures.

6.2 Rocks Installation on an Existing Frontend Node

If the frontend node has already been installed, you can add the OFED+ Roll for Rocks to the repository on the head node, update the master graph .xml file, and re-install all of the compute nodes as described in the following paragraphs. You must be logged in as a root user to perform these tasks.

If you need to burn a CD version of OFED+ Roll for Rocks from the .iso image, use the following script:

```
# su - root

# mount /dev/cdrom /mnt/cdrom

# rocks add roll

# umount /mnt/cdrom

# rocks enable roll intel_ofed

# rocks create distro

# rocks run roll intel_ofed | bash

# init 6
```

If you download the .iso image without burning a CD, use the following script:

```
# su - root
```

```
# rocks add roll package.iso

# rocks enable roll intel_ofed

# rocks create distro

# rocks run roll intel_ofed | bash

# init 6
```

Use the following command for each node:

```
# shoot-node compute_node_name
```

or use the following command to rebuild the entire cluster:

```
# rocks run host compute '/boot/kickstart/cluster_kickstart'
```

6.3 Add IPoIB Interfaces

Use the Rocks commands to add the IPoIB network to the network for the ib0 interface. For example:

```
rocks add network ib subnet=10.240.20.0 netmask=255.255.255.0
```

The default network MTU for IPoIB that Rocks creates is 1500.

Set the network as follows:

- IPoIB Connected Mode (CM), use MTU size 65520
- IPoIB Unreliable Datagram (UD), use MTU is 2044

```
#rocks set network mtu ipoib value (value = 2044 or 65520)
```

The firewall settings need to be edited to allow the frontend node to login to the compute nodes using the ipoib interface

```
# rocks add host firewall compute action=accept service=ssh chain=input
network=ipoib protocol=all flags="-m state --state NEW"
```

The firewall must be synced

```
# rocks sync host firewall compute
```

Tell Rocks to serve DNS for the ipoib network, and then push out the new configuration to DNS.

```
# rocks set network servedns ipoib true

# rocks sync config

Use the Rocks commands to add the host interfaces to the network for the ib0
interface. For example:

rocks add host interface head node name iface=ib0 subnet=ipoib ip=10.240.20.101
```

Use more `add host interface` commands to set the IP address for each compute node (this may be scripted if desired). For example:

```
rocks add host interface compute-0-0 iface=ib0 subnet=ib ip=10.240.20.102
```

If the error `interface "ib0" exists` is encountered:

1. Remove the ib0 interface on all the hosts.

```
#rocks remove host interface frontend_node ib0

#rocks remove host interface compute_node ib0
```

2. Sync the network.

```
#rocks sync host network
```

3. Add the host interfaces.

```
rocks add host interface head node name iface=ib0 subnet=ipoib ip=10.240.20.101
name=head node name-ib

rocks add host interface compute-0-0 iface=ib0 subnet=ib ip=10.240.20.102
name=compute-0-0-ib
```

Once the IP addresses have been set, use the sync commands to sync the fabric:

```
rocks sync host network localhost

rocks sync host network compute
```

6.4 Upgrade Instructions

Currently the Intel roll cannot be upgraded in one operation. The existing roll needs to be removed manually and then the new roll installed. Use the following "Roll Removal Instructions" and then use the "Rocks Installation on an Existing Frontend Node" instructions above to add the new roll.

6.4.1 Roll Removal Instructions

Uninstallation of Intel IB Software components must be performed using the following steps:

1. `# iba_config -u`
2. `# rpm -e intel_ofed`
3. `# rpm -e roll-intel_ofed-usersguide`
4. `# rocks remove roll intel_ofed`
5. `# rocks create distro`

6.4.2 Rocks Installation Instructions

Refer to Section 6.2, "Rocks Installation on an Existing Frontend Node" on page 80 for installation procedures.

§ §

7.0 Install Intel Software Using the Platform Cluster Manager Kit

7.1 Platform HPC, Kits, and Nodegroups

The Platform HPC Kit automates the provisioning of a cluster. The kits are a mechanism for packaging install scripts and applications for easy installation onto a Platform HPC cluster. Examples of kits include application software and OS software.

The Intel kits are:

- Intel® OFED+ kit – kit version of OFED+ software for installation in a Platform HPC cluster
- Intel® Fabric kit – kit version of IFS software for installation in a Platform HPC cluster

Platform HPC builds a cluster using the concept of nodegroups, which are groupings of nodes in the cluster. Two important nodegroups are relevant to the Intel software: installer and compute. The installer node group normally comprises the management node(s). The host nodegroup comprises nodes in the cluster that are to perform computing applications.

The Intel kits are targeted for the following default nodegroups in Platform HPC:

- Intel OFED+ kit – compute nodegroup
- Intel Fabric kit – installer nodegroup

7.2 New Installation for Platform HPC 3.X

The following procedures are for installing the Platform HPC 3.x kit on a new nodegroup or a nodegroup that does not have a previous version of Platform HPC installed. The installation procedures are provided in the Installing IBM Platform HPC guide. The following procedure provides the steps to integrate the Intel OFED+ kit or Intel Fabric kit into the Platform HPC 3.x:

Warning: During initial installation of master and host nodes, to avoid conflicts, **DO NOT** include os-ofed kits.

1. Obtain the *Installing IBM Platform HPC* guide from IBM.
2. Choose network addressing schemes for the following (Refer to the *Installing IBM Platform HPC* guide for a complete list of information required during the installation):

- The **public network interface**. This interface connects the master node to the main public network.
- The **private network interface**. This interface is used inside the Platform HPC cluster between the nodes.

3. Follow the procedures in the *Installing Platform HPC* guide with the following added steps.
4. Before performing the Start IBM Platform HPC installation in the *Installing Platform HPC* guide perform the following procedure to copy the Intel kit files that will be used for installation to the root directory.
   a. Open registrationcenter.intel.com/.
   b. Type Intel True Scale in the ***Search downloads*** field.
   c. Click the **Search** button.

d. Select the **Intel® True Scale Fabric Host Channel Adapter Host Drivers & Software vX.X.X.X.X** with the correct version number, in the ***XX* result(s) matching: *"Intel True Scale"* sorted by relevance** section of the web page.

e. For Platform HPC3.x kit, Click **Download** by the Intel® OFED+ Kit iso file (kit-intel_ofed-VERSION.x86_64.DISTRO.iso) and/or the Intel IFS iso file (kit-intel_ifs-VERSION.x86_64.DISTRO.iso) for the Platform HPC version required.
For Platform HPC4.1.1 kit, Click **Download** by Intel® OFED+ Kit .tar.bz2 file (kit-intel_ofed-VERSION-DISTRO-x86_64.tar.bz2) and/or the Intel IFS .tar.bz2 file (kit-intel_ifs-VERSION-DISTRO-x86_64.tar.bz2) for the new Platform HPC version required.

f. Read and agree to the *Intel Software License Agreement*.

g. For Platform HPC 3.x kit, Save the `.iso` image(s) to the local drive. For Platform HPC 4.1.1 kit, save the `.tar.bz2` image(s) to the local drive.

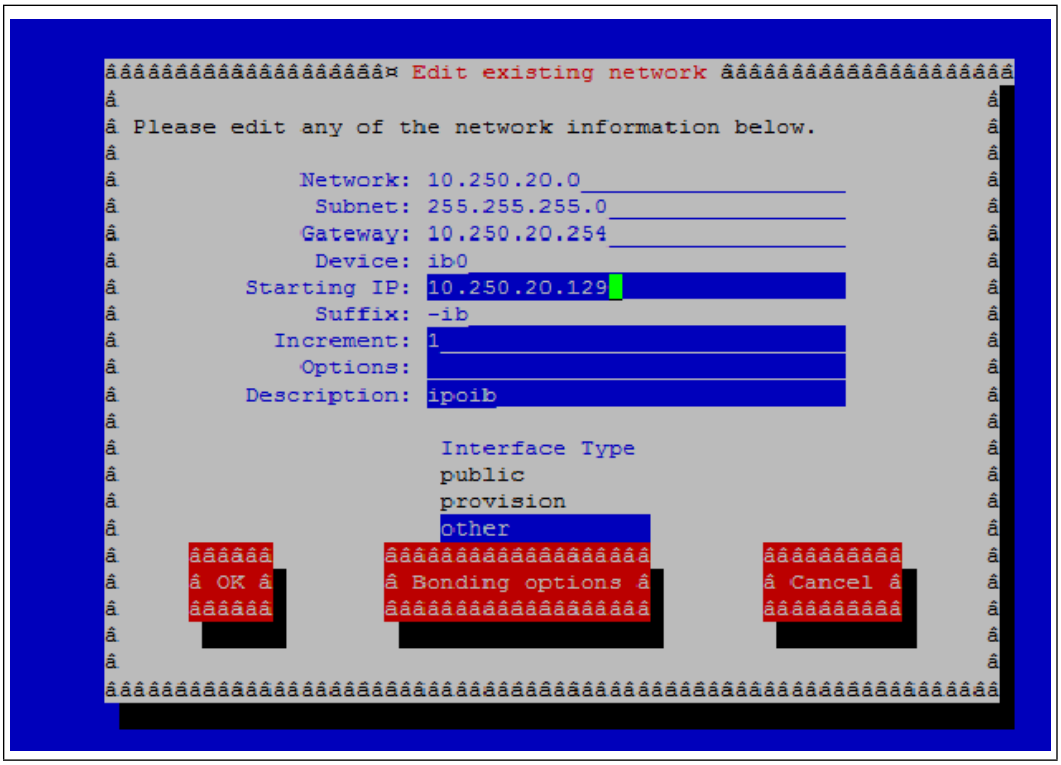
5. Continue with the procedures in the *Installing Platform HPC* guide, when prompted to add or delete kits, delete the following kits:
   — `os-ofed-*`
6. Add the Intel kits downloaded in Step 4.
7. Continue the procedures in the *Installing Platform HPC* guide.
8. Add the IPoIB interface following the steps in "Set up the IPoIB Interface for Platform HPC 3.x" on page 84.
9. Reboot the installer node

   Rebooting the installer node will start the Intel OFED+ and IFS software.

7.2.1 Set up the IPoIB Interface for Platform HPC 3.x

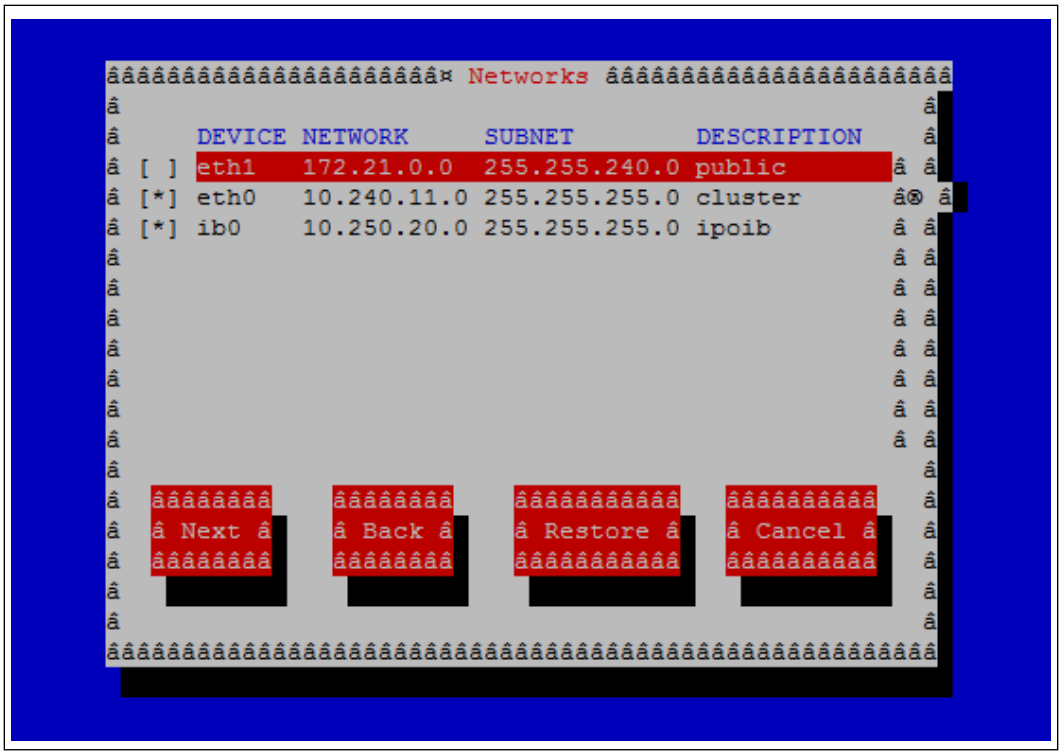
1. Add the IPoIB interface.
   a. On the installer node, use the kusu-netedit TUI (Figure 18) to add the interface.
   b. Designate the device as `ib0`.
   c. Use the other interfaces as examples for the other fields.
   d. Indicate the suffix to be `-ib`.
   e. Use `ipoib` as the description.

Figure 18. `kusu-netedit` **TUI**



2. Add the new network interface to the installer and compute nodegroups.
   a. Run the kusu-ngedit TUI (Figure 19).
   b. Select the nodegroup and then page ahead to the **Networks** screen.
   c. Add the network.

Figure 19. `kusu-ngedit` Tool



3. Run `kusu-net-tool updinstnic ib0` on installer node.

7.3 Existing Platform HPC Installation for Platform HPC 3.x Kits

If the Intel kits are not present in the cluster, use the following procedures to install the Platform HPC 3.x kits into an existing Platform HPC cluster:

1. Copy the Intel kit files that will be used for installation to the root directory.
   a. Open downloadcenter.intel.com/.
   b. Type Intel True Scale in the ***Search downloads*** field.
   c. Click the **Search** button.
   d. Select the **Intel® True Scale Fabric Host Channel Adapter Host Drivers & Software vX.X.X.X.X** with the correct version number, in the ***XX* result(s) matching: "Intel *True Scale*" sorted by relevance** section of the web page.
   e. Click **Download** by the Intel OFED+ Kit iso file (`kit-intel_ofed-`*`VERSION`*`.x86_64.`*`DISTRO`*`.iso`) and/or the Intel IFS iso file (`kit-intel_ifs-`*`VERSION`*`.x86_64.`*`DISTRO`*`.iso`) for the Platform HPC version required.
   f. Read and agree to the *Intel Software License Agreement*.
   g. Save the `.iso` image(s) to the local drive.

Note: If the installer node has Platform OFED and HPC kits installed you will need to remove these kits before you proceed.

2. Add the Intel kit(s) using the `kusu-kitops` command for each file.

```
kusu-kitops –a –m file.iso
```

3. Add the Intel kit(s) to the repository using the `kusu-repoman` tool:

```
kusu-repoman –r reponame –a –k <kit-name>
```

Where `kit-name = intel_ofed` or `intel_ifs`.

4. Sync the nodegroups using the `kusu-cfmsync` command.

```
kusu-cfmsync –n nodegroup –p
```

5. Verify, using the kusu-ngedit TUI, that the `intel_ofed` component is enabled for the host nodes and that the `intel_ifs` kit is enabled for the installer node if applicable.
6. Run `kusu-boothost` command to trigger the host nodes to PXE boot/reinstall.

```
kusu-boothost –n nodegroup –r
```

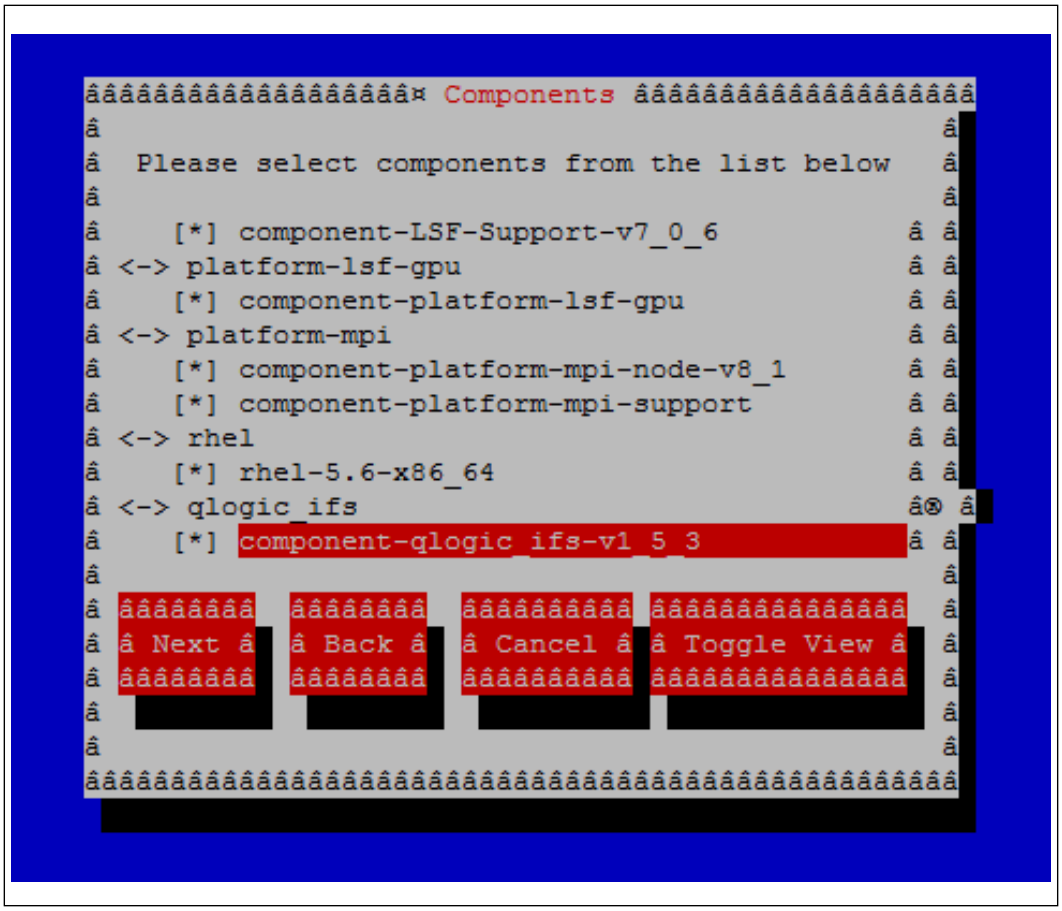
7. Wait for the host nodes to be accessible again.
8. Reboot the installer node with the `reboot` command and wait for it to come up.

7.4 Removing Kits From an Existing Platform HPC

To remove a kit from the cluster, use the following steps. For the Intel Fabric kit, use the installer nodegroup. For the Intel OFED+ kit, use the compute nodegroup.

1. Run the kusu-ngedit TUI to remove the kit component from the nodegroup.
2. Select the appropriate nodegroup on the main screen
3. Navigate to the **Components** screen (Figure 20),
4. Navigate down to the Intel kit, and uncheck the select box by pressing the space bar:

Figure 20. `kusu-ngedit` TUI Components Screen



5. Navigate through to the end with the **Next** keys and allow kusu-ngedit TUI to update the nodegroup.
6. Exit the kusu-ngedit utility.

   For the installer nodegroup, kusu-ngedit will pause to wait for the complete kit uninstall process to complete.

   For the compute nodegroup, kusu-ngedit will issue a message to the compute nodes and then return. When uninstalling the kits from the compute nodegroup, give the process 10-15 minutes to complete, then check a few random nodes to ensure that the Intel files have been deleted from the directories before proceeding (check for the existence of `iba_*` files in `/sbin`).
7. Remove the Intel kit(s) from the repository using the `kusu-repoman` tool:

```
kusu-repoman -r reponame -e -k <kit-name, e.g. intel_ofed or intel_ifs>
```

8. Remove the Intel kit(s) using the `kusu-kitops` command for each kit:

```
kusu-kitops -e -k <kit-name, e.g. intel_ofed or intel_ifs>
```

9. Check that the kits have been removed by issuing `kusu-kitops -l` command to list all the kits.

7.5 New Installation for Platform HPC 4.1.1

The following procedure is for installing the Platform HPC 4.1.1 kit on a new odegroup or a nodegroup that does not have a previous version of HPC installed. The installation procedure is provided in the Installing *IBM Platform HPC* guide. The following procedure provides the steps to integrate the Intel® OFED+ kit or Intel® Fabric kit into the Platform HPC 4.1.1 Kit installation.

7.6 Existing Platform HPC 4.1.1 Kits

If the Intel kits are not present in the cluster, use the following procedures to install the Platform HPC 4.1.1 kits into an existing Platform HPC cluster:

1. Copy the Intel kit files that will be used for installation to the root directory.
   a. Open downloadcenter.intel.com/.
   b. Type Intel True Scale in the ***Search downloads*** field.
   c. Click the **Search** button.
   d. Select the **Intel® True Scale Fabric Host Channel Adapter Host Drivers & Software vX.X.X.X.X** with the correct version number, in the ***XX* result(s) matching: "Intel *True Scale"* sorted by relevance** section of the web page.
   e. Click **Download** by the Intel® OFED+ Kit `.tar.bz2` file (`kit-intel_ofed-VERSION-DISTRO-x86_64.tar.bz2`) and/or the Intel IFS `.tar.bz2` file (`kit-intel_ifs-VERSION-DISTRO-x86_64.tar.bz2`) for the new Platform HPC version required.
   f. Read and agree to the *Intel Software License Agreement*.
   g. Save the `.tar.bz2` image(s) to the local drive.

Note: If the installer node has Platform OFED and HPC kits installed, you will need to remove these kits before you proceed.

2. Add the Intel® IFS kit(s) using the `addkit` command to XCAT cluster:

```
addkit kit-intel_ifs-1.5.4.1-7.2.1.1.20-rhels-6-x86_64.tar.bz2
```

or using Platform Management Console (PMC).

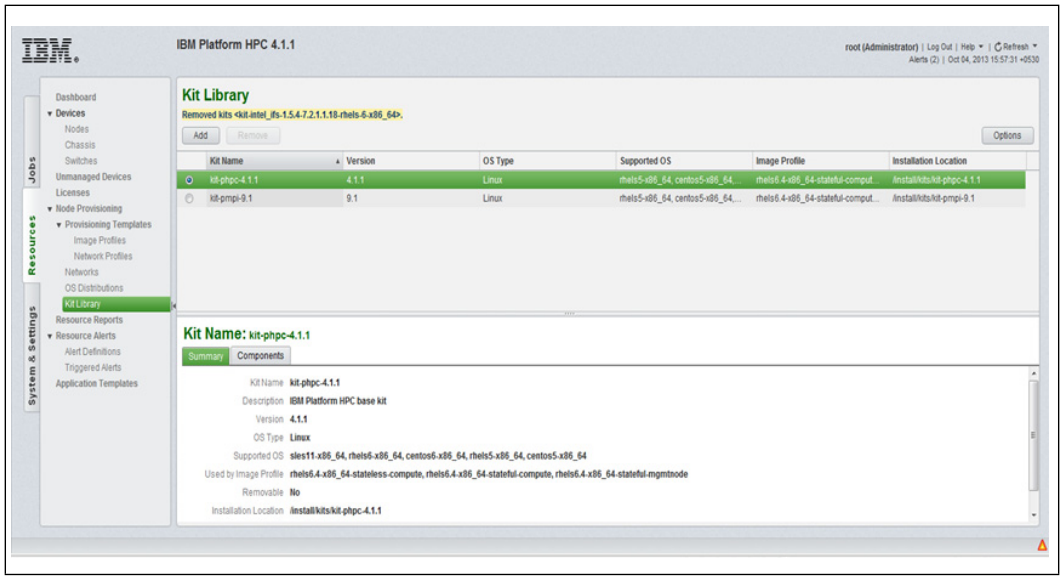
```
Navigate Node Provisioning>Kit Library -> Click Add.

Select location where kit *.tar.bz2 is stored.

Select next

Select add
```

Figure 21. Platform HPC 4.1.1 Install



3. Associate IFS kit component to osimage on management node using CLI:

```
addkitcomp -i rhels6.4-x86_64-stateful-mgmtnode compo-intel_ifs

updatenode < management node name >
```

The IFS kit is installed and ready for use.

4. Add the Intel OFED kit(s) using the `addkit` command to the XCAT cluster on the Installer Node using the CLI command

```
addkit kit-intel_ofed-1.5.4.1-7.2.1.1.20-rhels-6-x86_64.tar.bz2.
```

It can also be added using PMC as follows:

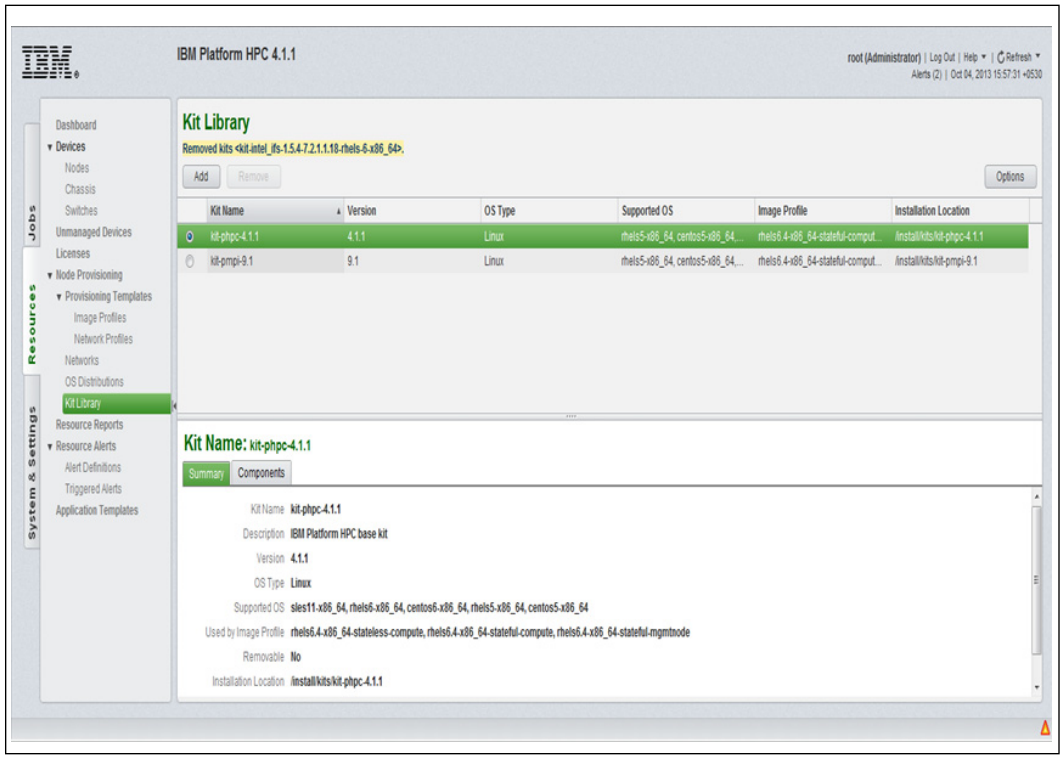
```
Navigate Node Provisioning>Kit Library -> Click Add.

Select location where kit *.tar.bz2 is stored.

Select next

Select add
```

Figure 22. Platform HPC 4.1.1 Install (cont.)



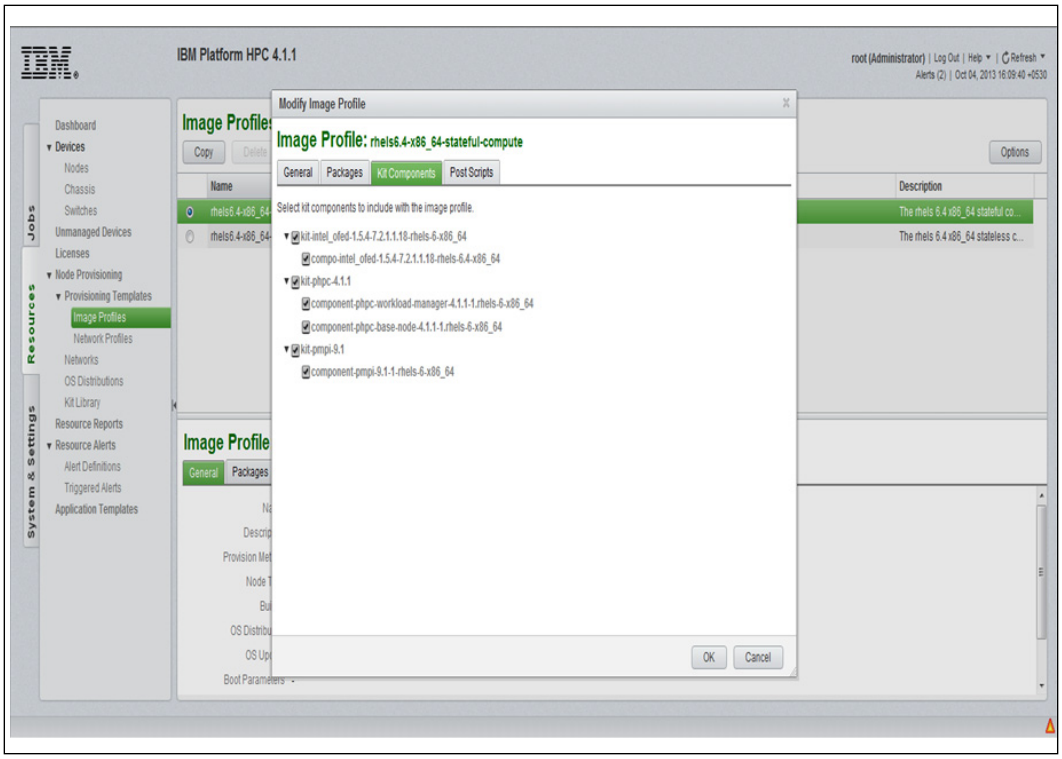
5. Associate the OFED kit component to `osimage` on the compute node using the CLI on the Installer Node.

```
addkitcomp -i rhels6.4-x86_64-stateful-compute compo-intel_ofed

updatenode <compute node range>
```

The kit component can also be associated using the GUI by selecting `OFED component` using a check mark.

Figure 23. Platform HPC 4.1.1 Install (cont.)



6. Associate the OFED help kit component to osimage on the management node using the CLI:

```
addkitcomp -i rhels6.4-x86_64-stateful-mgmtnode compo-kit-intel_ofed

updatenode < management node name >
```

7. Reboot the installer node, which will start the Intel® OFED+ and IFS software.

The OFED kit is ready to use.

7.6.1 Set up the IPoIB Interface for Platform HPC 4.1.1

After the True Scale Fabric IFS kit is installed on a Management node, FastFabric commands can be used to set up the IPoIB interface.

1. Add the IPoIB interface on the Installer node using the iba_config TUI to configure the IPoIB interface. Choose option 2, Reconfigure OFED IP over IB.
2. On the compute nodes, use the FastFabric TUI to configure IPoIB on all compute nodes. Choose option 6, Configure IPoIB IP Address to setup IPoIB.

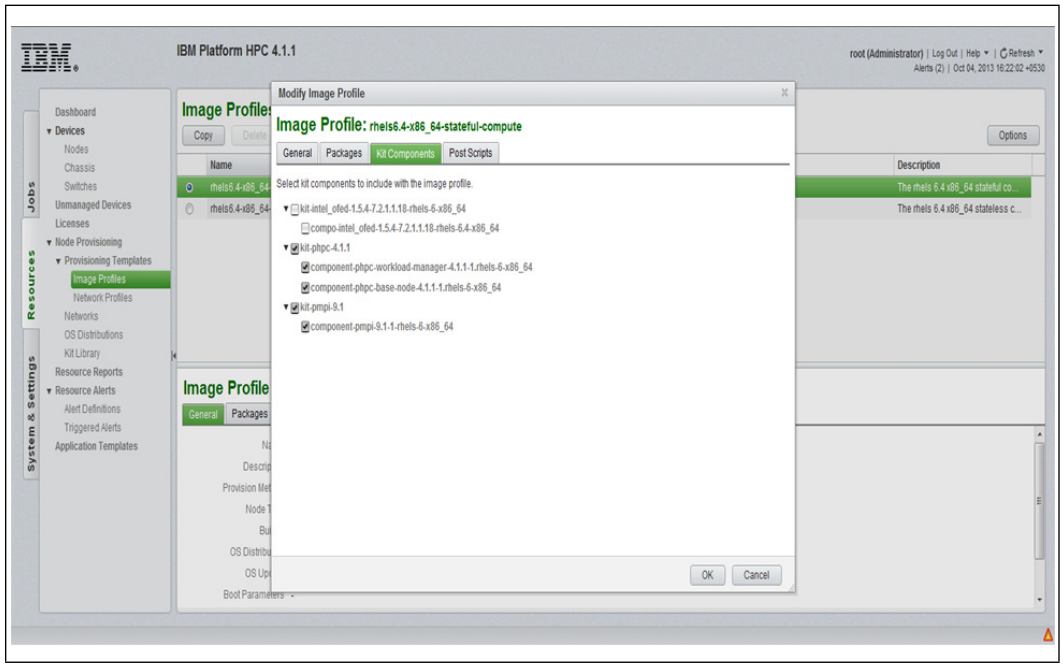
For more details please refer to the **configipoib** section of the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide*.

7.7 Removing Kits From an Existing Platform HPC 4.1.1

To remove a kit from the cluster, use the following steps. For the Intel® OFED+ kit, use the GUI. Note that the `rmkitcomp` and `updatenode` commands still need to run on Installer node to remove the OFED help component.

1. Log into the PMC.
2. Navigate to Node Provisioning->Provisioning Templates->Image Profiles.
3. Click Modify-> Kit Components tab ->Deselect `kit-intel_ofed-1.5.4.1-7.2.1.1.20-rhels-6-x86_64` using a check mark

Figure 24. Removing Kits from Platform HPC 4.1.1



Once the kit is disassociated from osimage, the following message is displayed; **Image profile update completed**.

4. Remove the OFED kit help component from the management node using the CLI commands:

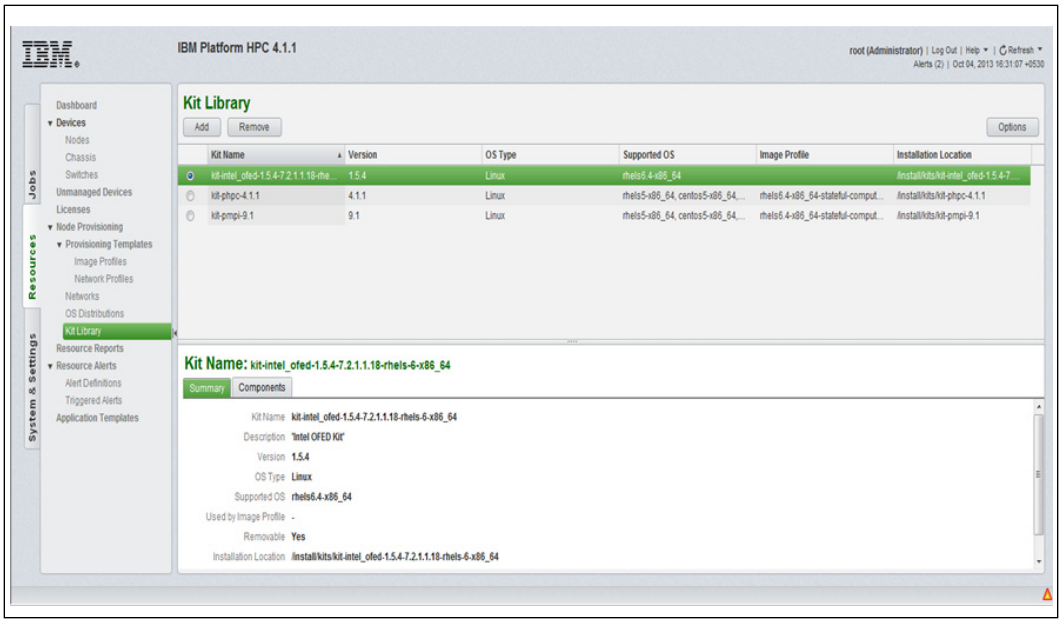
```
rmkitcomp -u -i rhels6.4-x86_64-stateful-mgmtnode compo-kit-intel_ofed.
```

5. Run `updatenode < management node >`
6. Remove the OFED kit using the CLI command:

```
rmkit kit-intel_ofed
```

or by using the **Remove** option from **Kit Library** using PMC.

Figure 25. Removing Kits from Platform HPC 4.1.1 (cont.)



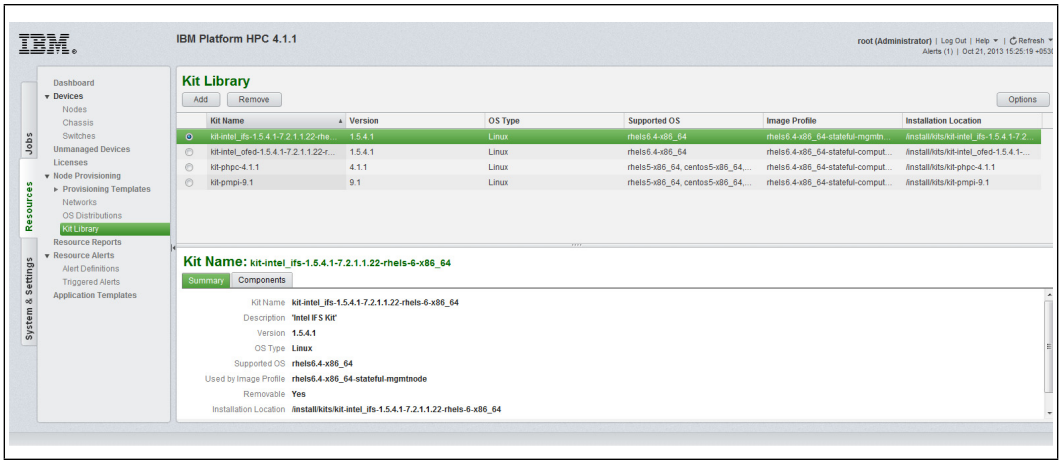
A message is displayed that kit is successfully removed.

7. Confirm the deletion of the selected kit(s).

To remove an Intel IFS kit from the cluster, use the `rmkitcomp` and `updatenode` CLI commands to disassociate the Kit from the OS image profile. Then use PMC to remove the kit.

1. Disassociate the IFS kit component from the management node using the CLI commands `rmkitcomp -u -i rhels6.4-x86_64-stateful-mgmtnode compo-intel_ifs`.
2. Run `updatenode < management node >`.
3. Remove the IFS kit using the CLI command `rmkit kit-intel_ifs`, or log into PMC and remove the kit using the **Remove** option from the **Kit Library**.

Figure 26. Removing Kits from Platform HPC 4.1.1 (cont.)



7.8 Updating Kits With an Existing Platform HPC Installation

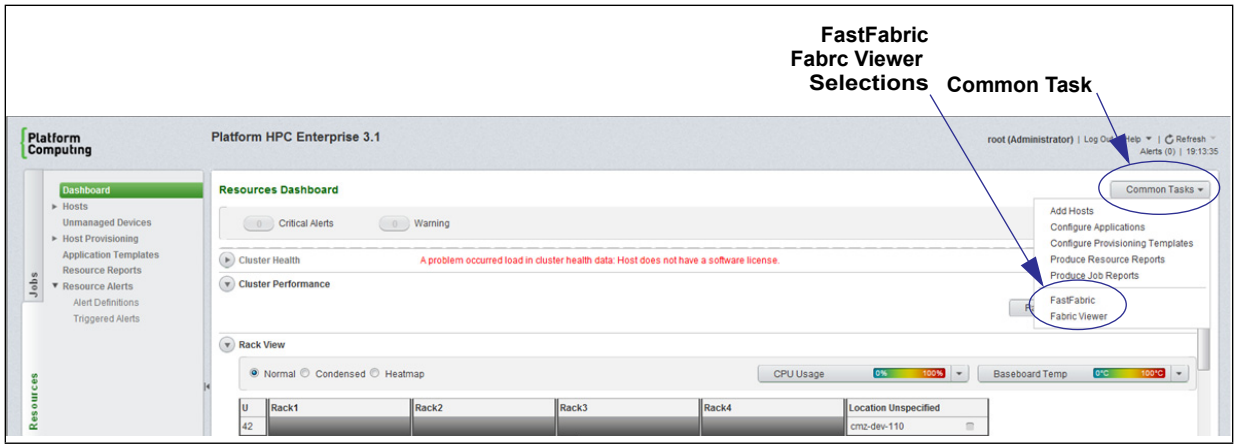
To update the Intel Platform HPC 3.2 kits that have been installed into an existing Platform HPC cluster, the kits must first be removed using the procedures in "Removing Kits From an Existing Platform HPC" on page 87. Then the new kits can be added using the procedures in "Set up the IPoIB Interface for Platform HPC 3.x" on page 84.

To update the Intel Platform HPC 4.1 kits that have been installed into an existing Platform HPC cluster, the kits must first be removed using the procedures in "Set up the IPoIB Interface for Platform HPC 4.1.1" on page 92. The new kits can then be added using the procedures in "New Installation for Platform HPC 4.1.1" on page 89.

7.9 Platform HPC GUI Integration—Intel Drop Down Menu Items

When the Intel Fabric kit is installed on the installer node, access to the Intel management tools is provided through the Platform HPC GUI. The Intel management tools can be selected from the Common Tasks drop down menu when the Common Tasks button at the upper right side of the GUI window is selected (Figure 27).

Figure 27. Platform HPC GUI Window

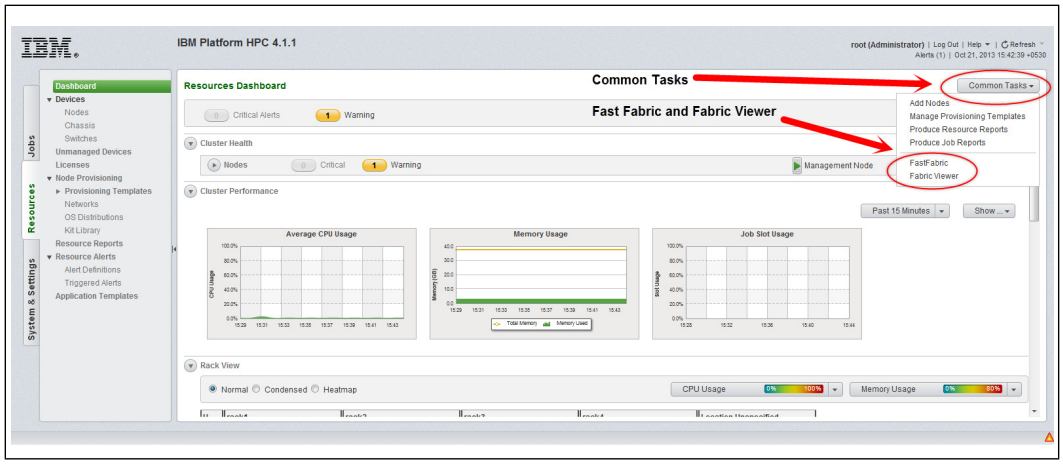


The following lists the two Intel management tools and their description:

- FastFabric – Selecting **FastFabric** opens an SSH applet to the installer node. After entering the root user name and password, the FastFabric TUI is available. To use the FastFabric TUI refer to the *Intel® True Scale Fabric Suite FastFabric User Guide*. After FastFabric is exited, close the browser window or tab.
- Fabric Viewer – Selecting the **Fabric Viewer** opens the Fabric Viewer applet in a browser window. The applet is equivalent in functionality to the Fabric Viewer application. To use the Fabric Viewer applet refer to the *Intel® True Scale Fabric Suite Fabric Viewer Online Help*. After the Fabric Viewer is exited, close the browser window or tab manually.

Note: Disable the fire wall on the installer node before opening the Fabric Viewer Applet so that you will be able to open port 3245 for FV-to-FM connectivity. Refer to the OS document for information on disabling the fire wall on the installer node.

Figure 28. Platform HPC 4.1.1 GUI Window



§ §

8.0 Install True Scale Fabric Suite Fabric Viewer

This section provides the procedures to install the True Scale Fabric Suite Fabric Viewer (FV) on both Windows* and Linux* platform. Refer to Section 8.1 for the Windows* Installation and Section 8.2 for the Linux* Installation.

8.1 Windows* Installation

8.1.1 System Requirements for a Windows* Environment

The following are minimum system requirements:

- Windows* operating system
- Internet Explorer* approved version for Windows*
- x86 processor architecture
- Ethernet card/local network access
- 2GB or greater of RAM
- 800x600 resolution (65K color depth)

One of the following browsers is also required when using the FV applet with Platform HPC:

- Internet Explorer* approved version for Windows*
- FireFox (latest version supported by the OS)

Note: Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for the latest supported OS and browser releases.

8.1.2 Install the True Scale Fabric Suite Fabric Viewer on a Windows* OS

Perform the following procedures to install the FV in a Windows* environment.

Caution: When re-installing the FV, the following files will be overwritten:

- **rules** – Contains event handler configurations.
- **preferences** – Contains start up user preferences.

To prevent these files from being overwritten, move them out of the `Program Files\Intel\Fabric_Viewer` folder.

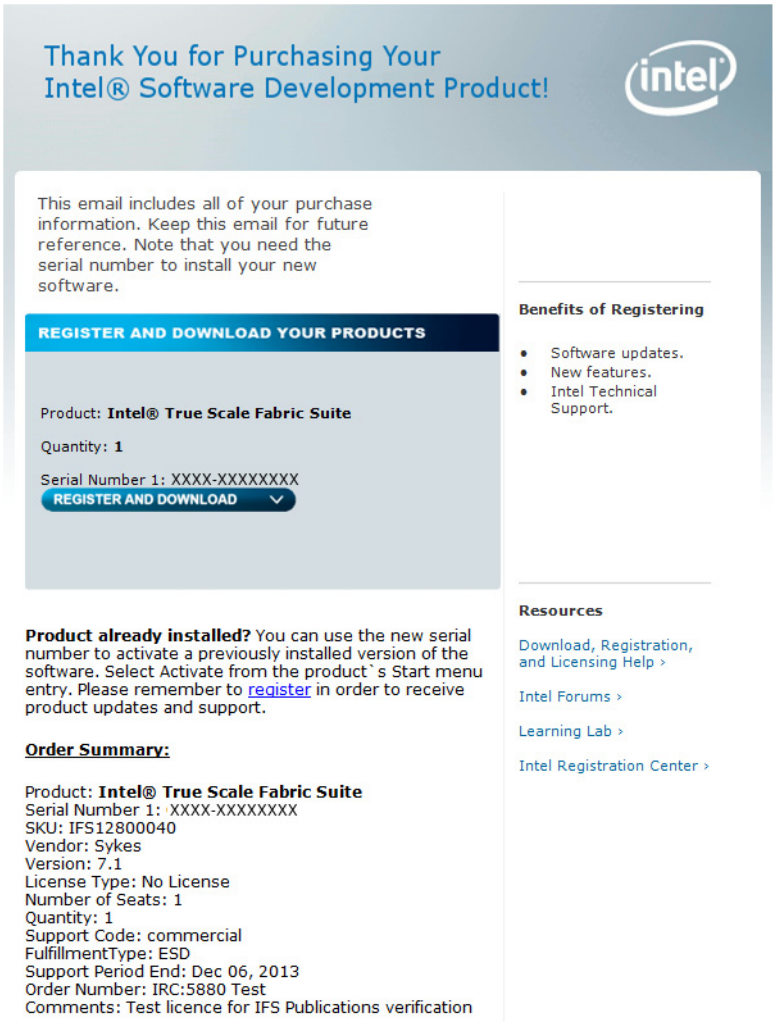
Once the installation is complete, move the files back into the `Program Files\Intel\Fabric_Viewer` folder.

8.1.3 Register and Download the True Scale Fabric Suite Software

Use the following procedure to register and download the True Scale Fabric Suite Software. When you purchased the True Scale Fabric Suite Software an e-mail was sent to the e-mail address provided during the purchase. Refer to that e-mail in the following procedure.

1. Select the **REGISTER AND DOWNLOAD** button in the e-mail received when the True Scale Fabric Suite Software was purchased. Figure 29 shows an example of the e-mail body.

Figure 29. Intel® Registration and Download E-Mail (Example)



The **True Scale Fabric Suite Software, Product Registration** web page will open.

2. Follow the instructions on the web pages to register and download the product.

8.1.4 Extract the .exe File

Use the following procedure to extract the `FabricViewer_x_x_x_x_x.exe` file.

1. Log in to the server where the FV will be installed.
2. Open the `FabricViewer_x_x_x_x_x.zip` file.
3. Extract the `FabricViewer_x_x_x_x_x.exe` file to the desktop.

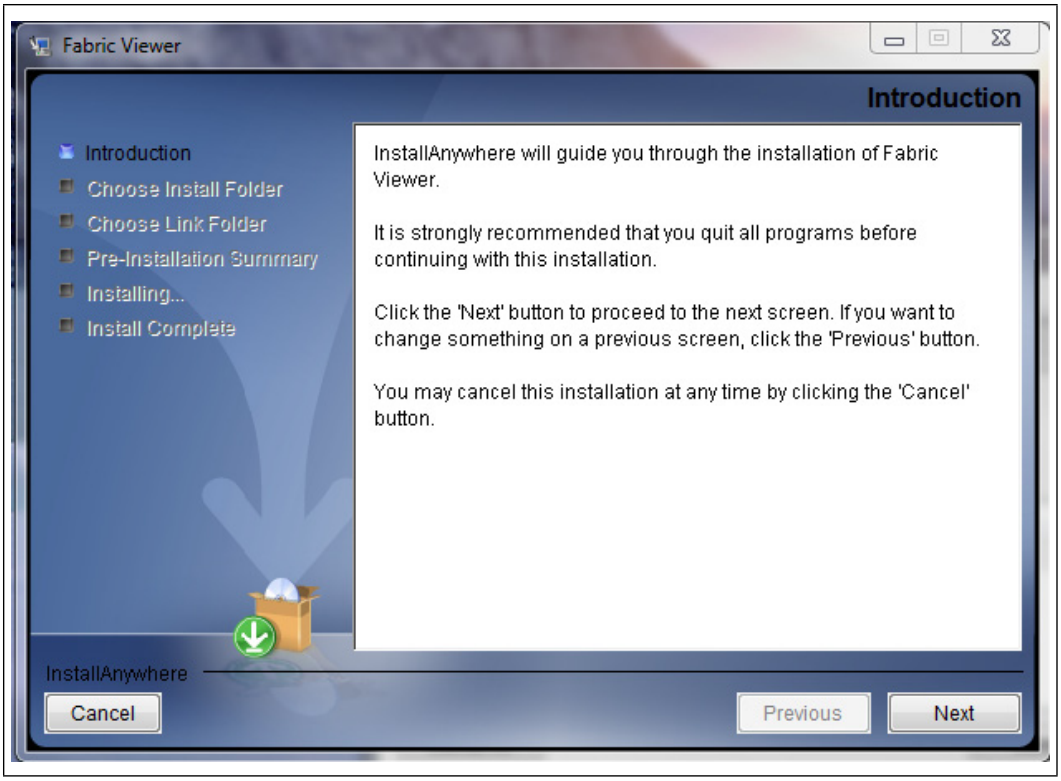
8.1.4.1 Using the Installation Wizard to Install the Fabric Viewer

1. Double-click the `FabricViewer_x_x_x_x_x.exe` file on the desktop.

Where *x_x_x_x_x* is the version number of the Fabric Viewer application being installed.

The InstallAnywhere installation program is installed and the **True Scale Fabric Suite Fabric Viewer Introduction** window opens (Figure 30).

Figure 30. True Scale Fabric Suite Fabric Viewer Introduction Window



2. Click **Next**.

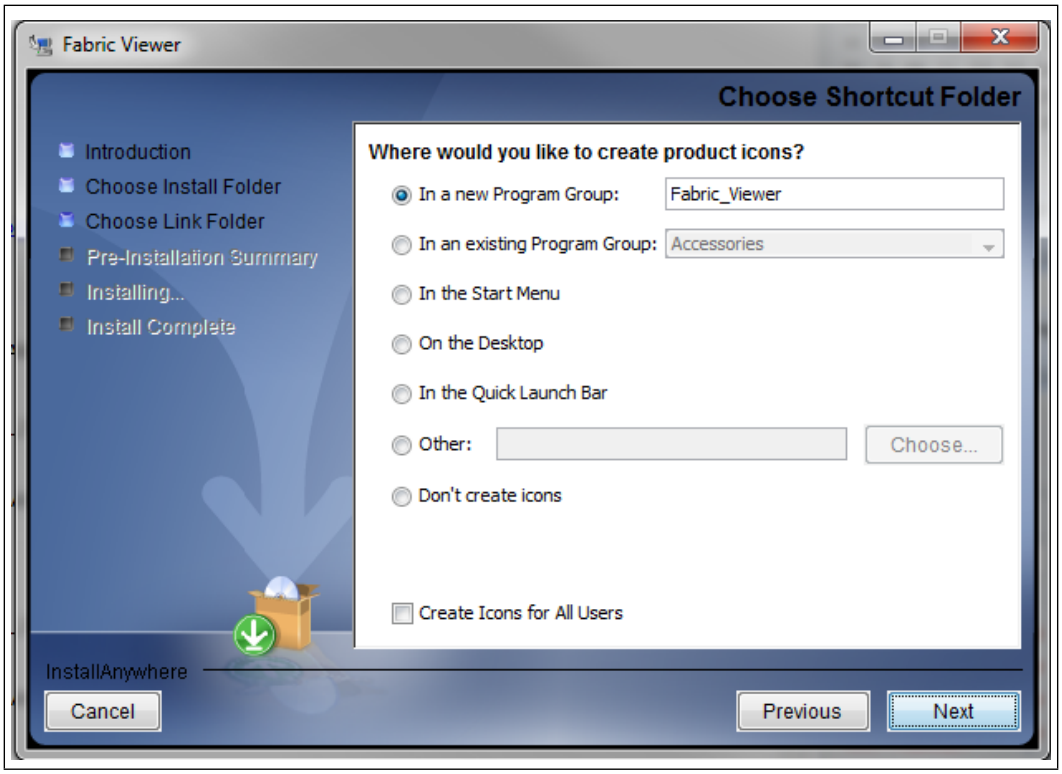
   The **Choose Install Folder** window appears.

Note: Intel recommends to use the default file location.

3. Click **Next**.

   The **Choose Shortcut Folder** window appears (Figure 31).

Figure 31. Choose Shortcut Folder Window



4. Select the shortcut folder where the program icons should be located.

Note: Intel recommends to use the default shortcut folder.

5. Click the **Create icons for All Users** checkbox, if all users should have access to this application.
6. Click **Next**.

   The **Pre-Installation Summary** window appears.
7. Click **Install**.

   The **Installing True Scale Fabric Suite Fabric Viewer** window appears.

   When the installation is complete the **Install Complete** window appears.
8. Click **Done**.

   The FV is installed.

8.2 Linux* Installation

8.2.1 System Requirements for a Linux* Environment

The following are minimum system requirements for the Linux* Installation:

- Linux* OS. Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for the latest supported OS releases.
- FireFox*, approved version for OS
- x86_64 processor architecture
- Ethernet card/local network access
- 2GB or greater of RAM
- X-Windows System

8.2.2 Install the True Scale Fabric Suite Fabric Viewer on a Linux* OS

Perform the following steps to install the FV in a Linux* environment.

Caution: When reinstalling the FV, the following files will be overwritten:

- **rules**: contains event handler configurations.
- **preferences**: contains start up user preferences.

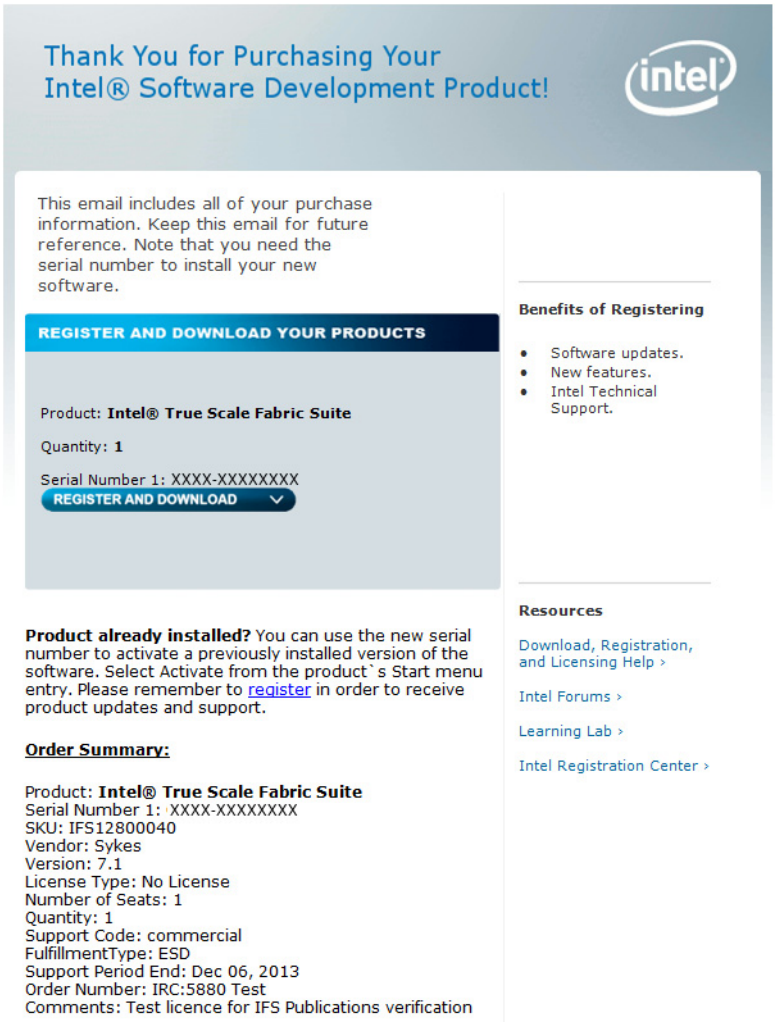
To prevent these files from being overwritten, move them out of the True Scale Fabric Suite Fabric Viewer directory. Once the installation is complete, move them back into the directory.

8.2.3 Register and Download the True Scale Fabric Suite Software

Use the following procedure to register and download the True Scale Fabric Suite Software. When you purchased the True Scale Fabric Suite Software an e-mail was sent to the e-mail address provided during the purchase. Refer to that e-mail in the following procedure.

1. Select the **REGISTER AND DOWNLOAD** button in the e-mail received when the True Scale Fabric Suite Software was purchased. Figure 32 shows an example of the e-mail body.

Figure 32. Intel® Registration and Download E-Mail (Example)



The **True Scale Fabric Suite Software, Product Registration** web page will open.

2. Follow the instructions on the web pages to register and download the product.

8.2.4 Extract the .bin File

Use the following procedure to copy the `FabricViewer_x_x_x_x_x.bin` file.

1. Log in to the server where FV will be installed.
2. Open an X Windows session.
3. Open a Terminal window in X Windows.
4. Make a temporary directory for the `Fabric Viewer` file.

```
mkdir Temporary_Directory
```

5. Open the `FabricViewer_x_x_x_x_x.zip` file.

6. Extract the `FabricViewer_x_x_x_x_x.bin` file to the temporary directory made in Step 4.
7. Change directories to the temporary directory made in Step 4.

```
cd Temporary_Directory
```

8. Change the modifiers for the bin file.

```
chmod 755 FabricViewer_x_x_x_x_x.bin
```

8.2.4.1 Using the Installation Wizard to Install the Fabric Viewer

1. Install the FV application.

```
./FabricViewer_x_x_x_x_x.bin
```

Where: *x_x_x_x_x* = Version number being installed.

The InstallAnywhere installation program is installed and the **True Scale Fabric Suite Fabric Viewer Introduction** window appears (Figure 33).

Figure 33. True Scale Fabric Suite Fabric Viewer Introduction Window

2. Click **Next**.

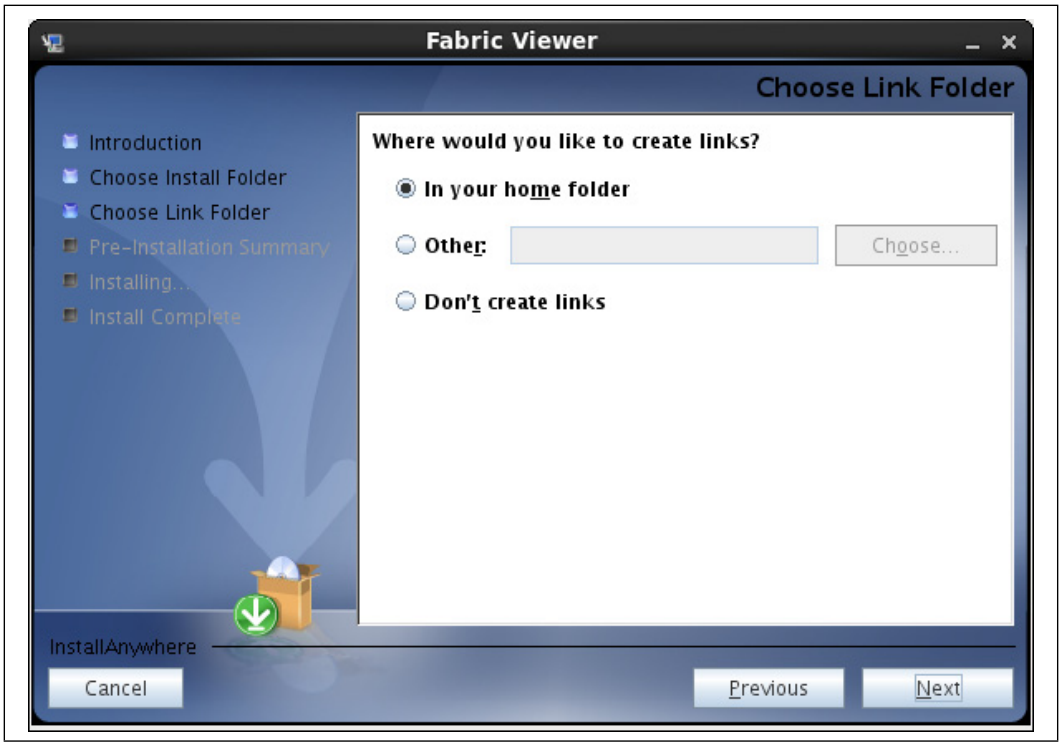
The **Choose Install Folder** window appears.

Note: Intel recommends to use the default file location.

3. Click **Next**.

The **Choose Link Folder** window appears (Figure 34).

Figure 34. Choose Link Folder Window



4. Select the link folder where the links should be located.

Note: Intel recommends to use the default selection.

5. Click **Next**.

   The **Pre-Installation Summary** window appears.

6. Click **Install**.

   The **Installing True Scale Fabric Suite Fabric Viewer** window appears.

   When the installation is complete the **Install Complete** window appears.

7. Click **Done**.

   The FV is installed.

8.3 Start the True Scale Fabric Suite Fabric Viewer Application

8.3.1 Windows* Procedure

The following procedure provides the steps to start the True Scale Fabric Suite Fabric Viewer application from the **Start** menu.

1. From the Windows* **Start** menu, select **All Programs**.
2. Select **Fabric_Viewer** folder.
3. Select **Fabric_Viewer.**

Note: By default, Windows sets its display properties to **Show icons using all possible colors**. If this property is not checked, the FV desktop and Start menu icons may appear distorted.

8.3.2 Linux* Procedure

The following steps are performed in a terminal window on X Windows and assume that the default installation for the FV was used.

1. Change to the main directory for the FV:

```
# cd /opt/Intel/Fabric_Viewer
```

2. Start the FV.

```
./Fabric_Viewer
```

Note: If an error is received, verify that the correct command `./Fabric_Viewer` was executed from the correct directory `/opt/Intel/Fabric_Viewer`.

8.4 Configure Startup Options

Note: Intel recommends that the user accepts the default settings.

Refer to the *Intel® True Scale Fabric Suite Fabric Viewer Online Help* for procedures to set the user preferences.

8.5 Uninstall the True Scale Fabric Suite Fabric Viewer

8.5.1 Windows* Procedure

The following instructions assume that the default installation for the FV was used.

Note: The FV must be closed for the uninstall to be successful. The uninstall program will not warn the user if the application is open. Warnings may be received at the end of the uninstall process stating that certain files have not been removed.

1. From the Windows* **Start** menu, select **All Programs**.
2. Select **Fabric_Viewer** folder.
3. Select **Uninstall_Fabric_Viewer**.
4. Follow the instructions on the uninstall windows.

8.5.2 Linux* Procedure

The following steps assume that the default installation for the FV was used. The following steps are performed in a terminal window on X Windows.

Note: The FV must be closed for the uninstall to be successful. The uninstall program will not warn the user if the application is open. Warnings may be received at the end of the uninstall process stating that certain files have not been removed.

1. Quit the FV.
2. Change to the home directory:

```
# cd
```

3. Execute the following command:

```
# ./Uninstall_Fabric_Viewer
```

The FV is uninstalled.

§ §

9.0 Upgrade the Management Node

This procedure provides discussion and step-by-step directions to upgrade an Fabric Management Node from a previous Intel® True Scale Fabric Suite (IFS) software version to the latest IFS software version.

9.1 Preinstallation

Prior to upgrading to IFS software, ensure the following have been performed:

- Review the Release Notes for a list of compatible software.
- Uninstall all versions of 3rd party IB stacks.
- Back up the following configuration files in case the upgrade fails:
  - `/etc/sysconfig/ifs_fm.xml`
  - `/etc/sysconfig/fastfabric.conf`
  - `/etc/sysconfig/iba/*`
  - `/var/opt/iba/analysis/baseline/*`
  - Refer To the OS documentation for a list of any other OS specific files that should be included in any backups.

Note: When managing a cluster where the IPoIB settings on the compute nodes are incompatible with the Fabric Management node (for example when a 4K MTU is used on the compute nodes and a 2K MTU is used on the management nodes), it is recommended not to run IPoIB on the Fabric management nodes.

9.2 Intel True Scale Fabric Suite Upgrade

To install the IFS software in a node with existing IFS software perform the following steps. Use the package file, `IntelIB-IFS.`*`DISTRO`*`.`*`VERSION`*`.tgz` on host where the full IFS package has been purchased.

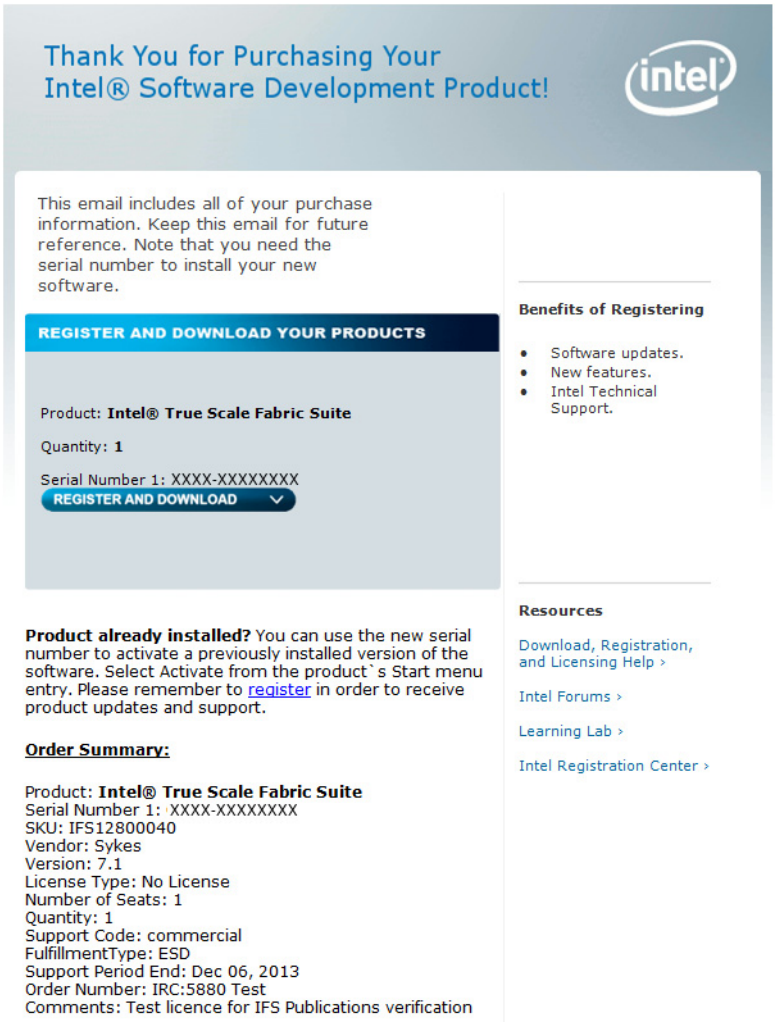
Using the menus, select to install the required components (at least OFED IB Stack, IB Tools and FastFabric) as described in the following procedures.

9.2.1 Register and Download the Intel® True Scale Fabric Suite

Use the following procedure to register and download the IFS. When you purchased the IFS software an e-mail was sent to the e-mail address provided during the purchase. Refer to that e-mail in the following procedure.

1. Select the **REGISTER AND DOWNLOAD** button in the e-mail received when the IFS software was purchased. Figure 35 shows an example of the e-mail body.

Figure 35. Intel® Registration and Download E-Mail (Example)



The **True Scale Fabric Suite Software, Product Registration** web page will open.

2. Follow the instructions on the web pages to register and download the product.

9.2.2 Unpack the Tar File

Use the following procedure to unpack the `IntelIBIFS.`*DISTRO.VERSION*`.tgz` tar file.

1. Copy the tar file to the /root directory.
2. Change directory to /root.

```
cd /root
```

3. Unpack the `IntelIB-IFS.`*DISTRO*`.`*VERSION* tar file to the `IntelIB-IFS.`*DISTRO*`.`*VERSION* directory using the following command:

```
tar xvfz IntelIB-IFS.DISTRO.VERSION.tgz
```

9.2.3 Upgrade IntelIB-IFS

To upgrade the IFS, perform the following procedure:

1. Change directory to `IntelIB-IFS.`*`DISTRO.VERSION`* directory

```
cd IntelIB-IFS.DISTRO.VERSION
```

2. Start the `Install` TUI:

```
./INSTALL
```

Note: If you need 32-bit support on 64-bit OSs, enter the following command:

```
./INSTALL --32bit
```

The **Intel IB** *VERSION* **Software** main menu appears (Figure 36).

Figure 36. Intel IB Software Main Menu (Example)

```
Intel IB VERSION Software


   1) Install/Uninstall Software
   2) Reconfigure OFED IP over IB
   3) Reconfigure Driver Autostart
   4) Update HCA Firmware
   5) Generate Supporting Information for Problem Report
   6) FastFabric (Host/Chassis/Switch Setup/Admin)


   X) Exit
```

3. Press **1** to select `Install/Uninstall Software`.

Screen 1 of 3 of the **Intel IB Install Menu** appears (Figure 37). Items that were installed in a previous installation are denoted by `[ Upgrade ]`.

Figure 37. Intel IB Install Menu (Screen 1 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 1 of 3):

0) OFED IB Stack       [   Upgrade   ][Available] VERSION

1) True Scale HCA Libs [   Upgrade   ][Available] VERSION

2) OFED mlx4 Driver    [   Upgrade   ][Available] VERSION

3) IB Tools            [   Upgrade   ][Available] VERSION

4) OFED IB Development [   Upgrade   ][Available] VERSION

5) FastFabric          [   Upgrade   ][Available] VERSION

6) OFED IP over IB     [   Upgrade   ][Available] VERSION

7) OFED IB Bonding     [   Upgrade   ][Available] VERSION

8) OFED SDP            [   Upgrade   ][Available] VERSION

9) Intel FM            [   Upgrade   ][Available] VERSION

a) MVAPICH (gcc)       [   Upgrade   ][Available] VERSION

b) MVAPICH2 (gcc)      [   Upgrade   ][Available] VERSION

c) OpenMPI (gcc)       [   Upgrade   ][Available] VERSION

d) MVAPICH/PSM (gcc)   [   Upgrade   ][Available] VERSION-DISTRO


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

Note: `OFED IB Bonding` will show as `[Not Avail]` when installing the software on OSs that have bonding modules in the OS installed software.

4. Review the items to be upgraded; the default value is in brackets (Upgrade, Uninstall, or Don't Install). To change a value, type the alphanumeric character associated with the item.
5. Press **N** to go to the next screen.

   Screen 2 of 3 of the **Intel IB Install Menu** appears (Figure 38)

Figure 38. Intel IB Install Menu (Screen 2 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 2 of 3):

0) MVAPICH/PSM (PGI)    [   Upgrade   ][Available] VERSION-DISTRO

1) MVAPICH/PSM (Intel) [   Upgrade   ][Available] VERSION-DISTRO

2) MVAPICH2/PSM (gcc)   [   Upgrade   ][Available] VERSION

3) MVAPICH2/PSM (PGI)   [   Upgrade   ][Available] VERSION

4) MVAPICH2/PSM (Intel)[   Upgrade   ][Available] VERSION

5) OpenMPI/PSM (gcc)    [   Upgrade   ][Available] VERSION-DISTRO

6) OpenMPI/PSM (PGI)    [   Upgrade   ][Available] VERSION-DISTRO

7) OpenMPI/PSM (Intel) [   Upgrade   ][Available] VERSION-DISTRO

8) Intel SHMEM          [   Upgrade   ][Available] VERSION.DISTRO

9) MPI Source           [   Upgrade   ][Available] VERSION

a) OFED uDAPL           [   Upgrade   ][Available] VERSION

b) OFED RDS             [   Upgrade   ][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

6. Review the items to be upgraded; the default value is in brackets (Upgrade or Don't Install). To change a value, type the alphanumeric character associated with the item.
7. Press **N** to go to the next screen.

   Screen 3 of 3 of the **Intel IB Install Menu** appears (Figure 39)

Figure 39. Intel IB Install Menu (Screen 3 of 3) Example

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 3 of 3):

0) OFED SRP            [   Upgrade   ][Available] VERSION

1) OFED SRP Target     [Don't Install][Available] VERSION

2) OFED iSER           [Don't Install][Not Avail]

3) OFED iWARP          [Don't Install][Available] VERSION

4) OFED Open SM        [Don't Install][Available] VERSION

5) OFED NFS RDMA       [Don't Install][Not Avail]

6) OFED Debug Info     [Don't Install][Not Avail]


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

8. Review the items to be upgraded; the default value is in brackets (Upgrade or Don't Install). To change a value, type the alphanumeric character associated with the item.
9. Press **P** to perform the selected actions from all three screens.

   The system prompts:

```
About to Uninstall previous InfiniBand Software Installations...

Hit any key to continue...
```

10. Press any key to proceed with the installation.
11. The following prompts will occur. For each prompt, select the default by pressing **Enter**

```
Preparing OFED VERSION release for Install...

Rebuild OFED SRPMs (a=all, p=prompt per SRPM, n=only as needed?) [n]:


Installing OFED IB Stack VERSION release...

Permit non-root users to query the fabric? [y]:


Enable OFED SMI/GSI renice (RENICE_IB_MAD)? [y]:
```

```
Single Port Mode reallocates all Intel HCA resources to HCA Port 1.

Enable Intel HCA Single Port Mode? [y]:
```

Note: Selecting the default by pressing **enter** causes the dual-port HCAs to act as single-port cards with only port 1 enabled. Enabling Intel HCA Single Port Mode increases performance for environments where the second port is not connected.

```
Do you want to keep //etc/sysconfig/iba/ibanodes? [y]:


Enable IPoIB Connected Mode (SET_IPOIB_CM)? [y]:


Do you want to keep OFED IP over IB ifcfg files
(//etc/sysconfig/network/ifcfg-ib[0-9]*)? [y]:


Enable OFED SRP High Availability deamon (SRPHA_ENABLE)? [n]:
```

12. Press **Enter** to select default (n).

    The **IB Autostart Menu** appears (Refer to Figure 40).

Figure 40. Intel IB Autostart Menu

```
Intel IB Autostart (VERSION release) Menu


Please Select Autostart Option:

0) OFED IB Stack (openibd)            [Enable ]

1) OFED mlx4 Driver (openibd)         [Enable ]

2) IB Port Monitor (iba_mon)          [Disable]

3) S20 Port Tuner (s20tune)           [Disable]

4) Distributed SA (dist_sa)           [Disable]

5) OFED IP over IB (openibd)          [Enable ]

6) OFED SDP (openibd)                 [Enable ]

7) IFS FM (ifs_fm)                    [Enable ]

8) OFED RDS (openibd)                 [Enable ]

a) OFED SRP (openibd)                 [Enable ]


P) Perform the autostart changes

S) Autostart All                     R) Autostart None

X) Return to Previous Menu (or ESC)
```

13. Review the items to be autostarted; the default value is in brackets (Enable or Disable). To change a value, type the alphanumeric character associated with the item.

    Intel recommends leaving all of the autostart selections as default, unless one of the following scenarios apply:

- If FastFabric will not monitor the fabric health, performance, and/or check the fabric for errors, change `IB Port Monitor (iba_mon)` to `Enable`.
- Intel recommends changing `Distributed SA (dist_sa)` to `Enable` when installing software in mesh/torus fabrics, or when using Virtual Fabrics with Intel HCAs, `dist_sa` must be enabled on management nodes only. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.

14. Press **P** to perform the selected actions from the screen.

    The system prompts:

```
Hit any key to continue...
```

15. Press any key.

    The system prompts with one of the following:

- The following lines appear stating the firmware is not required when using Intel HCAs.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log

Firmware is not required for the intel HCA(s) in this system.

Press any key to continue.
```

    Skip to Step 20.

- The following lines appear showing the number of HCAs found.

```
/usr/bin/iba_hca_firmware_tool    -i -l //var/log/iba.log

One HCA was found:
```

    When one or more HCA is found, the system prompts with each HCA name and the firmware version installed, and if there is an update available or not. If a firmware update is available or the firmware is up to date, the system prompts to update, install different firmware, or do nothing. Only Connect-X HCAs will have firmware available. Refer to the following bullet list for an example of the system prompt for each scenario:

- An update is available (Example):

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0: Update to 2.5.0 available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

- The firmware is up to date (Example):

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.5.0: Okay.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

- No firmware is available. This displays if the HCA is not a Connect-X HCA (Example).

```
0: MT_0390140002 (MHGA28-XTC/A4/A0) Firmware : No firmware available.

        Contact your vendor for firmware updates for this HCA.


No firmware available for HCAs in your system.

        Contact your vendor for firmware updates for this system.

Press any key to continue.
```

16. Perform one of the actions in the following table.

If	Then
No firmware is available	Skip to Step 20.
You need to upgrade the firmware	Proceed with Step 17.
You do not need to upgrade the firmware	Skip to Step 19.

17. Select a number corresponding to the HCA that needs upgraded.

    The system prompts (Example):

```
MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0

The following firmware revision(s) are available for this HCA:

0: MT_04A0110002: standard firmware

Select firmware version, or Q to cancel:
```

18. Select the number corresponding to the firmware revision required for the HCA.

    The firmware is installed on the HCA

    The system prompts:

```
0: MT_04A0110002 (MHGH28-XTC/X4/A0) Firmware 2.2.0: Update to 2.5.0 available.


To update an HCA, or to install different firmware on an HCA, type its number. To
quit, enter 'Q':
```

If	Then
You need to upgrade the firmware in another HCA	Repeat Step 17 and Step 18.
You do not need to upgrade the firmware on any other HCAs	Continue with Step 19.

19. Press **Q**

    The installation completes and displays the main menu

    Skip to Step 22.

20. Press any key.

```
Hit any key to continue...
```

21. Press any key.

    The installation completes and returns to the main menu:

22. Press **X** to exit.
23. Reboot the server.
24. **(All)** Compare all configuration files with the `-sample` configuration files to ensure they have the latest information and data.

Note: If FastFabric is being used, after the upgrade review the `FF_PRODUCT` parameter in `/etc/sysconfig/fastfabric.conf`. This parameter must be adjusted to match value shown in `/etc/sysconfig/fastfabric.conf-sample`.

§ §

10.0 Upgrade the Fabric

10.1 Upgrade OFED+ Host Software

If an existing fabric which has been installed and verified, needs to have Intel True Scale Fabric software upgraded, the following steps may be followed.

1. **(All)** On each Fabric Management Node, perform an upgrade installation of the IFS software using the procedure documented in the Section 9.0, "Upgrade the Management Node" on page 107. Each Fabric Management Node must have at least FastFabric, the OFED IB Stack and OFED IP over IB installed and configured.

   For MPI clusters using OFED+ software, the Fabric Management Nodes should also include the MPI Runtime and MPI Development packages. If the user desires to rebuild MPI itself, the True Scale Fabric Development package and MPI Source packages will also be required.

   After completing the install, reboot each of the Fabric Management Nodes to ensure they are running the new True Scale Fabric software.
2. **(All)** Start the FastFabric application from the login directory on the Management Node.
3. **(All)** Select the **Host Setup** option from the **FastFabric IB Host Setup Menu** (Figure 41).

Figure 41. FastFabric IB Host Setup Menu (Example)

```
FastFabric IB Host Setup Menu

Host List: /etc/sysconfig/iba/hosts


Setup:

0) Edit Config and Select/Edit Hosts Files   [Perform]

1) Verify Hosts via Ethernet ping            [ Skip  ]

2) Setup Password-less ssh/scp               [ Skip  ]

3) Copy /etc/hosts to all hosts              [ Skip  ]

4) Show uname -a for all hosts               [ Skip  ]

5) Install/Upgrade Intel IB Software         [Perform]

6) Configure IPoIB IP Address                [ Skip  ]

7) Build MPI Test Apps and Copy to Hosts     [ Skip  ]

8) Reboot Hosts                              [Perform]

Admin:

9) Refresh ssh Known Hosts                   [ Skip  ]

a) Rebuild MPI Library and Tools             [ Skip  ]

b) Run a command on all hosts                [ Skip  ]

c) Copy a file to all hosts                  [ Skip  ]

Review:

d) View iba_host_admin result files          [ Skip  ]


P) Perform the selected actions              N) Select None

X) Return to Previous Menu (or ESC)
```

4. Select the items that need to be performed in the menu and press **P** to perform them:
5. **(All)** `Edit Config and Select/Edit Hosts Files` will permit the hosts and `fastfabric.conf` files to be edited. When placed in the editor for `fastfabric.conf`, review all the settings. Especially review the `FF_PRODUCT`, `FF_PACKAGES`, and `FF_UPGRADE_OPTIONS`. See Appendix B, "Configuration Files" for more information about `fastfabric.conf`.

   Select a hosts list file which lists all the hosts except the Fabric Management nodes. If necessary, create a new file at this time, potentially based on the existing `/etc/sysconfig/iba/hosts file`.

Note: Do not list any of the Fabric Management Nodes in the host file (for example, the nodes which have FastFabric installed).

Note: The file may list the Management Network or IPoIB hostnames for the selected hosts

6. **(Host)** `Install/Upgrade Intel IB Software` will upgrade the True Scale Fabric software on all the selected hosts. By default it will look in the current directory for the `FF_PRODUCT.$FF_PRODUCT_VERSION.tgz` file. If it is not found in the current directory, it will prompt for input of a directory name where this file can be found.

Note: An upgrade installation will update any existing OFED+ software and is only valid for hosts which already have a previous version of OFED+ software installed.

Perform the following steps to upgrade the selected hosts:

The `Install/Upgrade Intel IB Software` will start with the following system prompts:

```
Performing Host Setup: Install/Upgrade Intel IB Software

Do you want to use ../IntelIB-Basic.DISTRO.VERSION.tgz? [y]:
```

a. Press **Enter** to accept the default (y).
System prompts:

```
Would you like to do an upgrade/reinstall? [y]:
```

b. Press **Enter** to accept the dealt (y).
System prompts:

```
You have selected to perform an upgrade installation

Are you sure you want to proceed? [n]:
```

c. Type **y** and press **Enter** to proceed.
System prompts:

```
Executing: /sbin/iba_host_admin -f /etc/sysconfig/iba/hosts -d .. upgrade

.

.

.

Hit any key to continue (or ESC to abort)...
```

d. Press any key to proceed.
System prompts:

```
Performing Host Setup: Reboot Hosts

.

.

.

Hit any key to continue (or ESC to abort)...
```

e. Press any key to proceed.
The selected hosts have completed rebooting the **FastFabric IB Host Setup Menu** appears. The upgrade is complete.

If any hosts fail to be updated, use the `View iba_host_admin result files` option to review the result files from the update. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Note: When using the True Scale Fabric packaging of OFED, FastFabric may be used to upgrade the True Scale Fabric stack on the remaining hosts. When using other packaging of OFED, FastFabric may be used to upgrade the True Scale Fabric Stack Tools (`IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz`) on the remaining hosts.

7. **(Linux)** If any other setup operations need to be performed on all hosts, use the `Run a command on all hosts` menu option. This option executes a the specified Linux* shell command (or sequence of commands separated by semicolons) against all selected hosts.

Note: Check the relevant release notes for the new OFED+ Host Software release being installed for any such additional required steps.

8. **(Linux)** `Reboot Hosts` will reboot all the selected hosts and ensure they go down and come back up (as verified through ping over the management network). When the hosts come back up, they will be running the True Scale Fabric software installed.
9. Repeat the verification steps for the fabric as discussed in "Verify OFED+ Host Software on the Remaining Servers" on page 57.

§ §

11.0 Upgrade from OFED+ Host Software to Intel IFS

This procedure provides discussion and step-by-step directions to upgrade a Fabric Management Node from OFED+ Host Software to the IFS.

To install the IFS in a node with existing OFED+ Host Software perform the following steps.

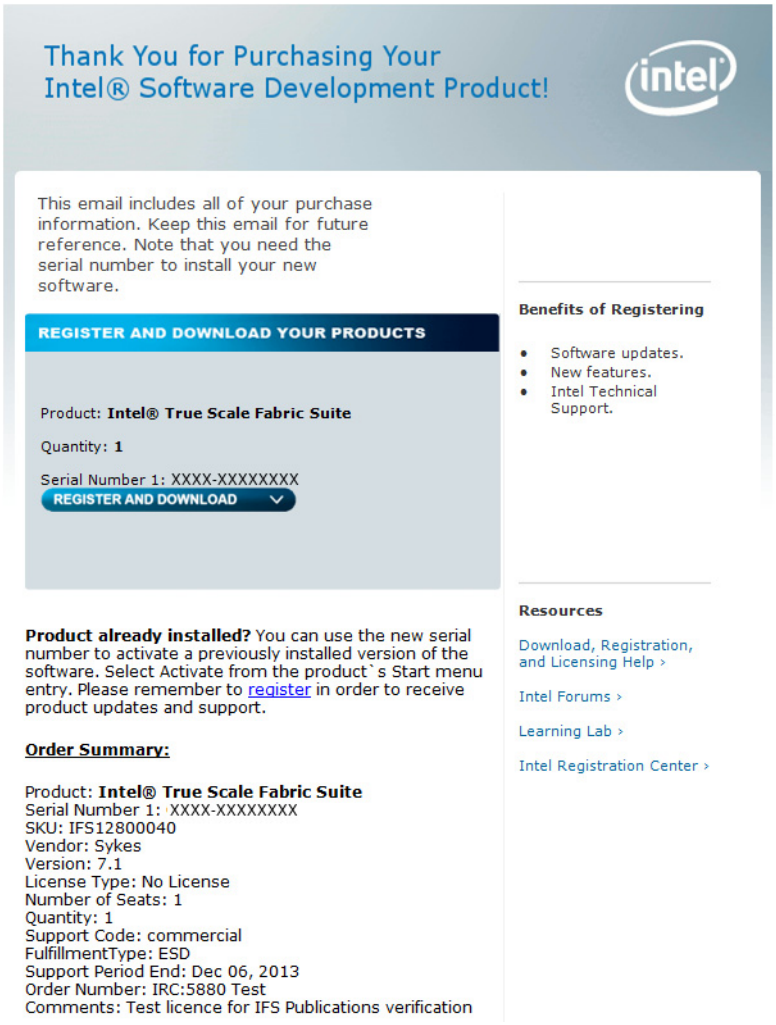
Use the package file, `IntelIB-IFS.DISTRO.VERSION.tgz`. Using the menus, select to install the required components (True Scale Fabric Suite FastFabric and True Scale Fabric Suite Fabric Manager) as described in the following procedures.

11.1 Register and Download the True Scale Fabric Suite Software

Use the following procedure to register and download the True Scale Fabric Suite Software. When you purchased the True Scale Fabric Suite Software an e-mail was sent to the e-mail address provided during the purchase. Refer to that e-mail in the following procedure.

1. Select the **REGISTER AND DOWNLOAD** button in the e-mail received when the True Scale Fabric Suite Software was purchased. Figure 42 shows an example of the e-mail body.

Figure 42. Intel® Registration and Download E-Mail (Example)



The **True Scale Fabric Suite Software, Product Registration** web page will open.

2. Follow the instructions on the web pages to register and download the product.

11.2 Unpack the Tar File

Use the following procedure to unpack the `IntelIB-IFS.`*DISTRO.VERSION*`.tgz`ar file.

1. Copy the tar file to the /root directory.
2. Change directory to /root.

```
cd /root
```

3. Unpack the `IntelIB-IFS.`*DISTRO*`.`*VERSION* tar file to the `IntelIB-IFS.`*DISTRO*`.`*VERSION* directory using the following command:

```
tar xvfz IntelIB-IFS.DISTRO.version.tgz
```

11.3 Install IntelIB-IFS

1. Type `cd IntelIB-IFS.DISTRO.VERSION` and press **Enter**
2. Type `./INSTALL` and press **Enter**.

   Displays the **Intel IB Software** main menu (Figure 43).

Figure 43. Intel IB Main Menu

```
Intel IB VERSION Software


   1) Install/Uninstall Software

   2) Reconfigure OFED IP over IB

   3) Reconfigure Driver Autostart

   4) Update HCA Firmware

   5) Generate Supporting Information for Problem Report

   6) FastFabric (Host/Chassis/Switch Setup/Admin)


   X) Exit
```

3. Press `1`

   Displays screen 1 of 3 of the **Intel IB Install Menu** (Figure 44). The FastFabric and IFS FM selections are showing `Install` while the other selections show `Up To Date`.

Figure 44. Intel IB Install Menu (Screen 1 of 3) (Example)

```
Intel IB Install (VERSION release) Menu


Please Select Install Action (screen 1 of 3):

0) OFED IB Stack       [ Up To Date  ][Available] VERSION

1) True Scale HCA Libs [ Up To Date  ][Available] VERSION

2) OFED mlx4 Driver    [ Up To Date  ][Available] VERSION

3) IB Tools            [ Up To Date  ][Available] VERSION

4) OFED IB Development [ Up To Date  ][Available] VERSION

5) FastFabric          [   Install   ][Available] VERSION

6) OFED IP over IB     [ Up To Date  ][Available] VERSION

8) OFED IB Bonding     [ Up To Date  ][Available] VERSION

9) OFED SDP            [ Up To Date  ][Available] VERSION

a) IFS FM              [   Install   ][Available] VERSION

b) MVAPICH (gcc)       [ Up To Date  ][Available] VERSION

c) MVAPICH2 (gcc)      [ Up To Date  ][Available] VERSION

d) OpenMPI (gcc)       [ Up To Date  ][Available] VERSION


N) Next Screen

P) Perform the selected actions       I) Install All

R) Re-Install All                     U) Uninstall All

X) Return to Previous Menu (or ESC)
```

4. Press **P**.

   Installs the FastFabric and Fabric Manager software selected.

   During the installation, the following prompt will be displayed. Select the default by pressing enter.

```
Do you want to keep //etc/sysconfig/iba/ibnodes? [y]:
```

5. Press **Enter** to select default (n).

   The **Intel IB Autostart Menu** appears (Refer to Figure 45).

Figure 45. Intel IB Autostart Menu

```
Intel IB Autostart (VERSION release) Menu


Please Select Autostart Option:

0) OFED IB Stack (openibd)            [Enable ]

1) OFED mlx4 Driver (openibd)         [Enable ]

2) IB Port Monitor (iba_mon)          [Disable]

3) S20 Port Tuner (s20tune)           [Disable]

4) Distributed SA (dist_sa)           [Disable]

5) OFED IP over IB (openibd)          [Enable ]

6) OFED SDP (openibd)                 [Enable ]

7) IFS FM (ifs_fm)                    [Enable ]

8) OFED RDS (openibd)                 [Enable ]

a) OFED SRP (openibd)                 [Enable ]


P) Perform the autostart changes

S) Autostart All                      R) Autostart None

X) Return to Previous Menu (or ESC)
```

6. Review the items to be autostarted; the default value is in brackets (Enable or Disable). To change a value, type the alphanumeric character associated with the item.

   Intel recommends leaving all of the autostart selections as default, unless one of the following scenarios apply:

- If FastFabric will not monitor the fabric health, performance, and/or check the fabric for errors, change `IB Port Monitor (iba_mon)` to `Enable`.
- Intel recommends changing `Distributed SA (dist_sa)` to `Enable` when installing software in mesh/torus fabrics, or when using Virtual Fabrics with Intel HCAs, `dist_sa` must be enabled on management nodes only. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information.

7. Press **P** to perform the selected actions from the screen.

   The system prompts:

```
Hit any key to continue...
```

8. Press any key.

   The installation completes and displays the main menu on the screen (Figure 43)

9. Press **X** to exit.
10. Reboot the server.

§ §

12.0 Install a Previous Version of Software

If the need exists to install a previous version of the Intel IFS use the following procedure.

1. Uninstall all existing software using the following command:

```
iba_config -u
```

2. Install the older version of the software using the installation procedures provided in the documentation that was released for that specific version of software.
3. Carefully review all configuration files for information that may need to be discarded or edited which are specific to features in the newer release which were not available in the older release
4. Reboot server.

§ §

13.0 Installation Verification and Additional Settings

This section provides instructions for verifying that the software has been properly installed, the Intel True Scale Fabric drivers are loaded, and that the fabric is active and ready to use. Information on the Intel HCAs and Performance tuning is also provided.

13.1 LED Link and Data Indicators

The LEDs function as link and data indicators once the Intel OFED+ software has been installed, the driver has been loaded, and the fabric is being actively managed by a subnet manager.

Table 4 describes the LED states. The green LED indicates the physical link signal; the amber LED indicates the link. The green LED normally illuminates first. The normal state is *Green On*, *Amber On*. The QLE7240 and QLE7280 have an additional state, as shown in Table 4.

Table 4. LED Link and Data Indicators

LED States	Indication
Green OFF Amber OFF	The switch is not powered up. The software is neither installed nor started. Loss of signal. Verify that the software is installed and configured with `ipath_control -i`. If correct, check both cable connectors.
Green ON Amber OFF	Signal detected and the physical link is up. Ready to talk to SM to bring the link fully up. If this state persists, the SM may be missing or the link may not be configured. Use `ipath_control -i` to verify the software state. If all HCAs are in this state, then the SM is not running. Check the SM configuration, or install and run `opensmd`.
Green ON Amber ON	The link is configured, properly connected, and ready. Signal detected. Ready to talk to an SM to bring the link fully up. The link is configured. Properly connected and ready to receive data and link packets.
Note: Green BLINKING (quickly) Amber ON	Indicates traffic
Note: Green BLINKING† Amber BLINKING	Locates the adapter This feature is controlled by `ipath_control -b [On | Off]`

†. This feature is available only on the QLE7200 series and QLE7300 series adapters.

13.2 Adapter and Other Settings

The following settings can be adjusted for better performance.

Note: This section is only applicable to clusters using Intel HCAs.

- **Use `taskset` to tune CPU affinity on Opteron systems with the QLE7240, QLE7280, and QLE7140.** Latency will be slightly lower for the Opteron socket that is closest to the PCI Express bridge. On some chipsets, bandwidth may be higher on this socket. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information on using `taskset`. Also see the `taskset(1)` man page.
- **On the switch, use an IB MTU of 4096 bytes instead of 2048 bytes, if available, with the QLE7240, QLE7280, and QLE7140.** 4K MTU is enabled in the InfiniPath driver by default. To change this setting for the driver, Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

- **Use a PCIe Max Read Request size of at least 512 bytes with the QLE7240 and QLE7280.** QLE7240 and QLE7280 adapters can support sizes from 128 bytes to 4096 bytes in powers of two. This value is typically set in the BIOS.
- **Use a PCIe MaxPayload size of 256, where available, with the QLE7240 and QLE7280.** The QLE7240 and QLE7280 adapters can support 128, 256, or 512 bytes. This value is typically set by the BIOS as the minimum value supported both by the PCIe card and the PCIe root complex.
- **Make sure that write combining is enabled.** The x86 Page Attribute Table (PAT) mechanism that allocates Write Combining (WC) mappings for the PIO buffers has been added and is now the default. If PAT is unavailable or PAT initialization fails for some reason, the code will generate a message in the log and fall back to the MTRR mechanism. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*.
- **Check the PCIe bus width.** If slots have a smaller electrical width than mechanical width, lower than expected performance may occur. Use the following command to check PCIe Bus width:

```
$ ipath_control -iv
```

This command also shows the link speed.

13.3 ARP Neighbor Table Setup for Large Clusters

The ARP neighbor table may overflow and produce a neighbor table overflow message to `/var/log/messages` along with other effects such as ping failing. To resolve this issue increase the threshold level for the network devices as follows:

5. Check the present threshold level 1.

```
cat /proc/sys/net/ipv4/neigh/default/gc_thresh1
```

It will give some value as 128, 256, or 512.

6. Increase the value to the next level. For example, if the value is 128 then make the thresh1 value as 256 and thresh2 as 512 and thresh3 as 1024. The following is an example of the syntax to be used if the value is 128:

```
echo 256 > /proc/sys/net/ipv4/neigh/default/gc_thresh1

echo 512 > /proc/sys/net/ipv4/neigh/default/gc_thresh2

echo 1024 > /proc/sys/net/ipv4/neigh/default/gc_thresh3
```

This will stop the error messages that were received in the log file.

13.4 Customer Acceptance Utility

`ipath_checkout` is a `bash` script that verifies that the installation is correct and that all the nodes of the network are functioning and mutually connected by the InfiniPath fabric. It must be run on a front end node, and requires specification of a nodefile. For example:

```
$ ipath_checkout [options] nodefile
```

The *`nodefile`* lists the hostnames of the nodes of the cluster, one hostname per line. The format of *`nodefile`* is as follows:

```
hostname1

hostname2
```

...

Note: The hostnames in the nodefile are Ethernet hostnames, not IPv4 addresses.

Note: To create a *nodefile*, use the `ibhosts` program. It will generate a list of available nodes that are already connected to the switch. The `ibhosts` program is described in more detail in the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

`ipath_checkout` performs the following seven tests on the cluster:

1. Executes the `ping` command to all nodes to verify that they all are reachable from the front end.
2. Executes the `ssh` command to each node to verify correct configuration of `ssh`.
3. Gathers and analyzes system configuration from the nodes.
4. Gathers and analyzes RPMs installed on the nodes. Missing RPMs can be found this way.
5. Verifies Intel hardware and software status and configuration. Includes tests for link speed, PIO bandwidth (incorrect MTRR settings), and MTU size.
6. Verifies the ability to `mpirun` jobs on the nodes.
7. Runs a bandwidth and latency test on every pair of nodes and analyzes the results.

The options available with `ipath_checkout` are shown in Table 5.

Table 5. `ipath_checkout` Options

Command	Meaning
-h, --help	These options display help messages describing how a command is used.
-v, --verbose -vv, --vverbose -vvv, --vvverbose	These options specify three successively higher levels of detail in reporting test results. There are four levels of detail in all, including the case where none of these options are given.
-c, --continue	When this option is not specified, the test terminates when any test fails. When specified, the tests continue after a failure, with failing nodes excluded from subsequent tests.
-k, --keep	This option keeps intermediate files that were created while performing tests and compiling reports. Results will be saved in a directory created by `mktemp` and named `infinipath_XXXXXX` or in the directory name given to `--workdir`.
--workdir=DIR	Use `DIR` to hold intermediate files created while running tests. `DIR` must not already exist.
`--run=LIST`	This option runs only the tests in `LIST`. See the seven tests listed previously. For example, `--run=123` will run only tests 1, 2, and 3.
--skip=LIST	This option skips the tests in `LIST`. See the seven tests listed previously. For example, `--skip=2457` will skip tests 2, 4, 5, and 7.
-d, --debug	This option turns on the `-x` and `-v` flags in `bash(1)`.

In most cases of failure, the script suggests recommended actions. Please see the `ipath_checkout` man page for more information and updates.

Also refer to the Troubleshooting appendix in the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

13.5 SM Loop Test

The SM Looptest is a diagnostic test facility in the Fabric Manager. As part of this test, the SM stress tests inter-switch links (ISLs) by continuously passing traffic through them. Other tools like FastFabric can be used to monitor the links for signal integrity

issues or other errors. The advantage of the Looptest, is that it provides a guaranteed way to test all of the ISLs in the fabric, without the need for a large number of end hosts or applications.For information on the SM Loop Test and how to use the test refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide*.

§ §

14.0 Configuration

Proper use of some advanced fabric features require consistent configuration of multiple components in the fabric as well as proper execution of jobs and applications.

This chapter summarizes the interdependencies of some of the advanced features and can serve as a reminder and checklist such that the configuration and operation allows the user to take full advantage of the required features.

The configuration subjects discussed in this section are:

- Intel Interface for NVIDIA GPUS
- Virtual Fabrics
- Congestion Analysis
- Mesh/Torus
- Adaptive Routing
- Adaptive Routing, Switch Configuration
- Dispersive Routing
- Distributed SA

14.1 Intel Interface for NVIDIA GPUS

The Linux* distribution software and the NVIDIA CUDA software supported are listed in the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for this release.

To get optimal performance with the ib_qib driver in the applications which use CUDA, set the following environment variable:

```
CUDA_NIC_INTEROP=1
```

Propagate this variable to the compute nodes running the application. Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for information on using MPI to propagate the variable. You can set this variable in the `.bash.rc` file of the shell you are using on each node.

14.2 Virtual Fabrics

Virtual Fabrics (vFabrics™) includes both security and quality of service (QOS) capabilities. vFabrics are configured within the FM, either by directly editing the `ifs_fm.xml` configuration file or through the Fabric Viewer GUI. The configuration of the FM must be consistent with the capabilities of the hardware and the usage of the fabric by the applications.

14.2.1 Virtual Fabrics, Switch Configuration

On the switches, the number of virtual lanes (VLs) and the maximum MTU must be configured. The recommended way to accomplish this is through FastFabric during initial installation or reconfiguration of the fabric. See "Configure Intel Chassis" on page 31. If required the VL and MTU configuration can also be directly configured through the switch CLI. See `ismSetChassisMtu` CLI command in the *Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide*.

For Fat Tree topologies, 1 VL is used per QOS level. 1 QOS level is used per vFabric with QOS enabled (plus 1 more VL if there are additional vFabrics without QOS enabled). See "Mesh/Torus" on page 142 for more information about Mesh/Torus topologies.

To optimize performance Intel recommends to configure the MTU using Table 6.

Table 6. Maximum Recommended MTUs

Number of VLs	Maximum Recommended MTU
1	4K
2	4K
3-4	4K
5-8	2K

Note: The MTU shown is the maximum recommended. Use of a larger MTU than shown can result in reduced performance in some cases. Use of a smaller MTU may provide useful performance trade-offs.

14.2.2 Virtual Fabrics, Fabric Manager Configuration

The Fabric Manager sample configuration has a variety of sample Virtual Fabrics which can be easily enabled. See the Virtual Fabrics section of the *Intel® True Scale Fabric Suite Fabric Manager User Guide*. Table 7 displays a few possibilities of the useful combinations.

Table 7. Sample Fabric Manager vFabric Combinations

Number of vFabrics	Typical Combinations	Result
1	Default	All traffic together.
2	Compute, AllOthersWithSA	Separate MPI from everything else.
2	Admin, AllOthers	Separate FM/FastFabric from everything else.
2	PSM_Compute_Control, AllOthersWithSA	Separate PSM control messages from everything else.
3	Networking, Compute, AllOthersWithSA	Separate Networking (and IPoIB based file systems), MPI, and everything else.
3	Networking, Compute, Admin	Separate Networking, MPI, and Admin. Assumes no other applications use fabric.
3	PSM_Compute_Control, PSM_Compute_Data, AllOthersWithSA	Separate PSM control messages from data and everything else.
4	Networking, Compute, AllOthers, Admin	Separate Networking, MPI, Fabric Manager/FastFabric, and everything else.
4	Networking, PSM_Compute_Control, AllOthers, Admin	Separate Networking, PSM Control, Fabric Manager/FastFabric, and everything else.
4	Networking, PSM_Compute_Control, PSM_Compute_Data, AllOthersWithSA	Separate PSM control and data, networking, and everything else.
4	Networking, PSM_Compute_Control, PSM_Compute_Data, Admin	Separate PSM control and data, networking, and FM/FastFabric. Assumes no other applications use the fabric.

Note: Only one of Default, Admin or AllOthersWithSA can be enabled at a time.

Note: It is not recommended to use Default when configuring more than 1 Virtual Fabric. Instead use AllOthersWithSA or Admin in this case.

14.2.3 Virtual Fabrics, OFED+ Configuration

When using Virtual Fabrics in conjunction with Intel HCAs and Performance Scaled Messaging (PSM) with PathRecord query enabled ("Using PathRecord Query" on page 145), Intel recommends to enable the Distributed SA (dist_sa) for autostart on all the compute nodes. This will simplify the operation of MPI jobs using PSM.

14.2.3.1 Enabling Distributed SA

Enable the Distributed SA (`dist_sa`) for autostart on all the compute nodes. It can be enabled in any of the following manners:

1. Interactively responding to the prompts for `dist_sa` autostart when installing, upgrading, or running `iba_config`
2. Adding `-E dist_sa` to the FF_INSTALL_OPTIONS and FF_UPGRADE_OPTIONS in `fastfabric.conf`, and using FastFabric to install or upgrade all the compute nodes
3. Using the `cmdall` command or any other distributed shell to perform a `iba_config -E dist_sa` operation on all nodes.

Refer to Distributed SA section in the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information on Distributed SA.

14.2.4 Virtual Fabrics, Application and ULP Configuration

Virtual Fabrics operates using IBTA compliant mechanisms. One of the key requirements of Virtual Fabrics is that applications make SA PathRecord queries to obtain critical address information such as PKey and SL.

When using Virtual Fabrics, applications must be properly assigned to the right vFabric. For applications which use PathRecord queries or use IB Multicast Groups, the FM will use the ServiceID or MGID respectively to lookup the application and select an appropriate vFabric to use in response to the query.

14.2.4.1 MPI over PSM Configuration

Unlike Verbs MPIs, PSM provides a simple and consolidated way to assign jobs to Virtual Fabrics. For MPI jobs using PSM, this is most easily accomplished by using Path Record queries and the Distributed SA.

14.2.4.1.1 Using PathRecord Query

Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for how to specify `PSM_PATH_REC=opp` and how to configure that variable globally on the compute nodes. If multiple vFabrics are configured for PSM with different sets of ServiceIDs in each, MPI jobs can be assigned to a specific ServiceID through `PSM_IB_SERVICE_ID`. For more information on this parameter refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

When using FastFabric to perform the `iba_host_admin mpiperf` or `iba_host_admin mpiperdeviation` verification steps, the `PSM_PATH_REC` option must be specified in `fastfabric.conf`, in the `FF_MPI_ENV` setting. See Appendix B, "Configuration Files" for the `fastfabric.conf` file. Alternatively, it can be specified in the `/opt/iba/src/mpi_apps/ofed.*.param` files

When using `/opt/iba/src/mpi_apps/run_*` scripts to run sample MPI applications and benchmarks, the `/opt/iba/src/mpi_apps/ofed.*.param` files must be edited to uncomment the `PSM_PATH_REC` setting.

14.2.4.1.2 Directly Specifying vFabric or PKey and SL

When not using PathRecord query, vFabric addressing information must be directly supplied to the MPI job using `mpirun` command line options or environment variables. To use `mpirun` command refer to the "Using mpirun" section of the *Intel® True Scale Fabric OFED+ Host Software User Guide*. To use environment variables refer to the "Using SL and PKeys" section of the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

When using FastFabric or the `/opt/iba/src/mpi_apps/run_*` scripts to run sample MPI applications and benchmarks, the `/opt/iba/src/mpi_apps/ofed.*.param` files must be edited.

14.2.4.2 MPI over Verbs Configuration

For historical reasons most MPI implementations over Verbs do not interact with the SA and instead hand construct address information using hardcoded PKey, SL, and other information. When using vFabrics with such MPIs, the vFabric addressing information must be directly supplied to the MPI job using `mpirun` command line options or environment variables. To use the `mpirun` command refer to the "Using mpirun" section of the *Intel® True Scale Fabric OFED+ Host Software User Guide*. To use environment variables refer to the "Using SL and PKeys" section of the *Intel® True Scale Fabric OFED+ Host Software User Guide*

When using FastFabric or the `/opt/iba/src/mpi_apps/run_*` scripts to run sample MPI applications and benchmarks, the `/opt/iba/src/mpi_apps/ofed.*.param` files must be edited.

14.2.4.3 IPoIB Configuration

IPoIB will automatically use the 1st enabled vFabric in the FM configuration file. Typically this will be Networking, Default, AllOthers or AllOthersWithSA. The sample FM configuration file is constructed such that one of these will typically be first so that IPoIB operation will be as is typically required.

If its desirable to create more than 1 IPoIB vFabric, then additional pKeys must be configured into IPoIB. Refer to "Using SL and PKeys" section in the *Intel® True Scale Fabric OFED+ Host Software User Guide.*

14.2.4.4 Other Applications and ULPs

Most applications will not require any special operational changes. This is because, with the exception of MPI, all other applications tend to use IBTA standard approaches for connection establishment. The FM sample configuration file has application definitions for many popular applications and ULPs. Additional applications can be added as needed or the Default, AllOthers, or AllOthersWithSA vFabric definitions can be used and/or the AllOthers Application.

Note: Some applications make use of IPoIB for address resolutions, such applications may end up using the same vFabric as IPoIB.

14.2.5 Virtual Fabrics, Moab Scheduler Configuration

The sole intent of the Moab scripts is to act as templates for the administrator, on how to enable Moab to take advantage of Virtual Fabric (vFabric) functionality. The Moab scripts are located in the `/opt/iba/Moab_scripts/` directory. In order to make these scripts accessible to common users, the administrator must do the following:

1. When necessary, edit the scripts to meet the configuration and usage criteria requirements specific to your Moab environment.

2. Copy the scripts to a shared file system service that provides the appropriate access rights for regular users.

In order for Moab to take advantage of the vFabric functionality, vFabric must be integrated with Moab. There are several integration methods that can be used, the scripts used here use the method of Moab based queues. This method requires that the administrator perform the following steps (for detailed information on vFabric, refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide*):

3. For both Moab and the Resource Manager(s), create a queue for each vFabric defined within the FM configuration file (`/etc/sysconfig/ifs_fm.xml`).
4. Configure each vFabric queue defined within Moab with the compute nodes assigned to it within the FM configuration file.
5. (Optional) Configure each vFabric queue defined within Moab with the priorities, policies, QoS, and attributes that are specific to the needs of the organization and fabric environment.

Note: Depending on the queue support requirements of a Resource Manager, steps 2 and 3 may also need to be performed on that Resource Manager (TORQUE does not require these two steps).

14.2.5.1 Moab Submit Scripts

Moab submit scripts utilize the Moab primitive `msub`, encapsulating the arguments and call structure for ease-of-use. The scripts query the SA for information regarding virtual fabrics, or in the case of PSM with dist_sa, pass the appropriate arguments to the PSM for SA lookup.

The submit scripts default the MPICH_PREFIX if it is not set in the environment. The openmpi, mvapich and mvapich2 scripts default to the -qlc versions of MPI.

The submit scripts utilize a helper script called `moab.mpi.job.wrapper`. This helper script is used to dynamically generate a temporary host file called `mpi_hosts_file.PID` that is required for MPI. This host file is created in the work directory on the primary node executing the job and is a list of hosts that need to run MPI. The Moab scripts do not delete the temporary host file; therefore the user will need to delete them at the end of the configuration.

The following are some examples of how to call the scripts and pass vFabric parameters (these examples utilize the proprietary MPI applications of Intel, which are located in the `/opt/iba/src/mpi_apps/` directory. These MPI applications are provided specifically for administrative use only):

14.2.5.1.1 Example 1:

```
moab.submit.openmpi.job -p 2 -V vf_name -n Compute -d 3 osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem, For example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under `gcc`. The number of processes is given as 2. The virtual fabric to use is identified by name, and that name is Compute. Option 3 is given to denote a dispersive routing option of static_dest.

14.2.5.1.2 Example 2:

```
moab.submit.openmpi.job -p 2 -V sid -s 0x0000000000000022 -d 1 osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem, For example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under gcc. The number of processes is given as 2. The virtual fabric to use is identified by service id, and that id is `0x0000000000000022`. Option 1 is given to denote a dispersive routing option of adaptive.

14.2.5.1.3 Example 3:

```
moab.submit.openmpi.job -p 2 -V sid_qlc -s 0x1000117500000000 -q night
osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem, For example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under gcc. The number of processes is given as 2. The virtual fabric to use is identified by service id, and that id is `0x1000117500000000`. The service id will be passed to the QIB driver which will perform the vFabric lookup on behalf of the job. No dispersive routing instructions are given. The job is being requested for submission through the night Moab queue.

14.2.5.2 Moab Script Administration

The scripts are intended to be tailored by the Moab administrator in order to structure the msub parameters appropriately. Any parameters, such as queue name, application profile, work directory, and so on, may be inserted into the script either using script options, accessing environment variables, or by hard-coding.

The scripts use `mpirun_rsh` or `mpirun` to start MPI jobs. This method enables the vFabric parameters to be passed to the MPI jobs so that the appropriate lower level actions can be taken based on those settings (for example, inserting Service Level information into data messages). Any MPI job command parameters should be used on the script command line and will be passed to the MPI run job.

14.2.6 Virtual Fabrics, LSF Scheduler Configuration

The sole intent of the Load Sharing Facility (LSF) scripts is to act as templates for the administrator, on how to enable LSF to take advantage of Virtual Fabric (vFabric) functionality. The LSF scripts are located in the `/opt/iba/LSF_scripts/` directory. In order to make these scripts accessible to common users, the administrator must do the following:

1. When necessary, edit the scripts to meet the configuration and usage criteria requirements specific to your LSF environment.
2. Copy the scripts to a shared file system service that provides the appropriate access rights for regular users.

In order for LSF to take advantage of vFabric functionality, vFabric must be integrated with LSF. There are several integration methods that could be used, the scripts used here use the method of LSF based queues. This method requires that the administrator perform the following steps (for detailed information on vFabric, refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide*):

3. For LSF, create a queue for each vFabric defined within the FM configuration file (`/etc/sysconfig/ifs_fm.xml`).
4. Configure each vFabric queue defined within LSF with the compute nodes assigned to it within the FM configuration file.
5. (Optional) Configure each vFabric queue defined within LSF with the priorities, policies, QoS, and attributes that are specific to the needs of the organization and fabric environment.

14.2.6.1 Configuring Nodes for LSF

It is recommended that the changes described in this section be performed on or propagated to all nodes in the fabric. This will ensure proper job submission and operation. FastFabric may be used to set up password-less ssh. Once password-less ssh is set up, it can be used to copy the changed files to all nodes in the fabric.

14.2.6.2 LSF Submit Scripts

LSF submit scripts utilize the LSF primitive "bsub", encapsulating the arguments and call structure for ease-of-use. They query the SA for information regarding virtual fabrics, or in the case of PSM with dist_sa, pass the appropriate arguments to PSM for SA lookup.

The submit scripts default the MPICH_PREFIX if it is not set in the environment, although Intel recommends it to be set explicitly in the environment to avoid any confusion. The mvapich and mvapich2 script default to the -qlc versions of MPI. There are two openmpi scripts; one for standard mpi and one for the -qlc version. The `bsub` is different for standard openmpi, so it is in its own script.

The following are some examples of how to call the scripts and pass vFabric parameters (these examples utilize the proprietary MPI applications of Intel, which are located in the `/opt/iba/src/mpi_apps/` directory. These MPI applications are provided specifically for administrative use only):

14.2.6.2.1 Example 1:

```
lsf.submit.openmpi-q.job -p 2 -V vf_name -n Compute -d 3 osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem. For example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under `gcc`. The number of processes is given as 2. The virtual fabric to use is identified by name, and that name is Compute. Option 3 is given to denote a dispersive routing option of static_dest.

14.2.6.2.2 Example 2:

```
lsf.submit.openmpi-q.job -p 2 -V sid -s 0x0000000000000022 -d 1 osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem. For example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under gcc. The number of processes is given as 2. The virtual fabric to use is identified by service id, and that id is `0x0000000000000022`. Option 1 is given to denote a dispersive routing option of adaptive.

14.2.6.2.3 Example 3:

```
lsf.submit.openmpi-q.job -p 2 -V sid_qlc -s 0x1000117500000000 -q night
osu2/osu_bibw
```

This submits a job with the openmpi-qlc subsystem, Foe example, the run file `osu_bibw` is expected to have been compiled with the openmpi-qlc compiler under gcc. The number of processes is given as 2. The virtual fabric to use is identified by service id, and that service id is `0x1000117500000000`. The service id will be passed to the QIB driver which will perform the vFabric lookup on behalf of the job. No dispersive routing instructions are given. The job is being requested for submission through the night LSF queue.

To view the vFabric parameters associated with the job, use the `bjobs -l` *`job number`* command in LSF for the long format of output.

14.2.6.3 LSF Script Administration

The scripts are intended to be tailored by the LSF administrator in order to structure the bsub parameters appropriately. Any parameters, such as queue name, application profile, and so on, may be inserted into the script either using script options, accessing environment variables, or by hard-coding.

The scripts use various forms of mpirun to start jobs. For mvapich and mvapich2, `mpirun_rsh` is used. For Platform HPC 3.2 and Platform HPC 4.1.1, `openmpi-q` and the standard `openmpi`, both use mpirun. Note that older releases of Platform HPC (i.e., Platform HPC 3.1 and earlier) will continue to use `mpirun.lsf` with the standard `openmpi` script. This method enables the vFabric parameters to be passed to the MPI jobs so that the appropriate lower level actions can be taken based on those settings (for example, inserting Service Level information into data messages). Any MPI job command parameters should be used on the script command line and will be passed to the MPI run job.

14.2.6.4 SSH Script

In accordance with the instructions in the Platform document *Integrating LSF's blaunch with MPI Applications*, a script for ssh is provided so that LSF can intercept MPI jobs and convey them to the blaunch facility. The script is constructed in such a way so that normal users of ssh are not affected, while LSF usage is bypassed to blaunch.

While the *Integrating LSF's blaunch with MPI Applications* document illustrates how to perform this with rsh. ssh is the preferred mechanism with OpenMPI.

14.2.6.5 Instructions to install the ssh script:

The following procedures provides the instruction for installing the `ssh` scripts:

1. Log in as root.
2. Change directory to `/usr/bin`.

```
cd /usr/bin
```

3. Move `ssh` to `ssh.bin`.

```
mv ssh ssh.bin
```

4. Copy `ssh.script` as `/usr/bin/ssh`.

```
cp /opt/iba/LSF_scripts/ssh.script /usr/bin/ssh
```

5. Change the permissions on the ssh script:

```
chmod 755 /usr/bin/ssh
```

14.2.6.6 Modifying openmpi_wrapper

The openmpi_wrapper script provided with LSF, located in `LSF_BINDIR`, contains a hard-coded path to the OpenMPI directory as installed by Platform's HPC kit. Since this installation supersedes the Platform's HPC kit with the Intel kit, the location of the OpenMPI files is different. Therefore, the openmpi_wrapper script needs to be modified.

The ex script, `edit.openmpi.wrapper`, may be used to modify the openmpi_wrapper. It uses an ex editor session to change the path from `MPIHOME` to `MPICH_PREFIX`. The file may be modified by using the script while logged in as root, as follows:

6. Change directory to `/opt/iba/lsf_scripts`.

```
cd /opt/iba/lsf_scripts
```

7. Run the ex script to modify the `openmpi.wrapper`.

```
ex /opt/lsf/7.0/linux2.6-glibc2.3-x86_64/bin/openmpi_wrapper <
edit.openmpi.wrapper
```

If editing with the script fails, `openmpi_wrapper` may be edited manually to ensure that the path to MPIRUN_CMD is correct.

14.2.6.7 Configuration changes in lsf.conf

Changes to the LSF configuration file are required in order to ensure proper operation: The `lsf.conf` file is located in the `/opt/lsf/conf` directory. Edit the file to change the LSF_ROOT_REX parameter from local to all.

```
LSF_ROOT_REX=all
```

The LSF daemons need to be restarted for the configuration change to take effect. Refer to the *Administering Platform LSF* document information on restarting the LSF daemons.

14.2.7 Virtual Fabrics, Fabric Viewer Configuration

The Fabric Viewer may be used to configure the Virtual Fabrics for the FM. Refer to the *Intel® True Scale Fabric Suite Fabric Viewer Online Help* for more information.

14.3 Congestion Analysis

The FastFabric Congestion Analysis capabilities leverage a new paradigm for statistics gathering. These tools centralize the statistics monitoring into the PM. For proper operation of these tools, no other applications can be directly clearing the PMA counters in the fabric.

14.3.1 Congestion Analysis, Switch Configuration

Intel recommends using FastFabric during initial installation or reconfiguration of the fabric to configure the switches in the fabric. Refer to "Configure Intel Chassis" on page 31. If required the ismAutoClearConf environment variable can also be directly configured using the switch CLI.

Ensure the ismAutoClearConf environment variable is disabled. Refer to the *Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide*.

Ensure the `ismSetPStatThresh` environment variables are not enabled. Refer to the *Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide*.

14.3.2 Congestion Analysis, Fabric Manager Configuration

The PM monitoring should be enabled by configuring the Pm.SweepInterval. Refer to "Install and Configure the Fabric Manager" on page 43 and the *Intel® True Scale Fabric Suite Fabric Manager User Guide*.

14.3.3 Congestion Analysis, OFED+ Configuration

Ensure that `iba_mon` and `s20tune` are disabled. Refer to "Install OFED+ Host Software" on page 67, Step 13. Ensure that `[ disable ]` is selected (the default) for the following items:

```
IB Port Monitor (iba_mon)
```

S20 Port Tuner (s20tune)

Avoid using tools which clear the counters, such as `clear_p1stats`, and `ibcheckerrors`. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

14.3.4 Congestion Analysis, Management Node Configuration

The Management Node configuration is accomplished by installing the IFS software and the FastFabric Toolkit. The `iba_top`, `iba_rfm`, `iba_paquery`, and `iba_report` tools will all detect and access the PM within the Fabric Manager. These tools, and related tools such as `all_analysis`, `fabric_analysis`, `iba_extract_perf`, etc, are all safe to use. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

Avoid using tools that directly clear the counters, such as `iba_report --pmadirect`. Refer to the *Intel® True Scale Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

14.4 Mesh/Torus

Proper operation of a Mesh/Torus fabric requires a complete system solution. Unlike other fabric topologies, the use of LIDs and SLs can vary between different pairs of source and destination nodes even within a single vFabric. As such all applications should use SA PathRecord queries to obtain path information.

By design Mesh/Torus topologies can have more potential for congestion. It is recommended to also use other advanced features when configuring a Mesh/Torus:

- Adaptive Routing - Refer to "Adaptive Routing" on page 144
- Dispersive Routing - Refer to "Dispersive Routing" on page 144
- Congestion Analysis - Refer to "Congestion Analysis" on page 141

14.4.1 Mesh/Torus Fabric, Switch Configuration

On the switches, the number of VLs and the maximum MTU must be configured. Intel recommends using FastFabric during initial installation or reconfiguration of the fabric. Refer to "Configure Intel Chassis" on page 31. If required the VL and MTU configuration can also be directly configured through the switch CLI. See `ismSetChassisMtu` in the *Intel® True Scale Fabric Switches 12000 Series CLI Reference Guide*.

For Mesh/Torus topologies, multiple VLs and SLs can be used per QOS level. Refer to the table of QoS capabilities for Mesh/Torus in the *Intel® True Scale Fabric Suite Fabric Manager User Guide*, for more information about the SL and VL requirements for various Mesh/Torus topologies.

The MTU recommendations for a given number of VLs are the same as for other topologies. See Table 6 for more information.

14.4.2 Mesh/Torus Fabric, Fabric Manager Configuration

The Fabric Manager must have the dor-updown RoutingAlgorithm selected. Refer to the "Routing Algorithm" section of the *Intel® True Scale Fabric Suite Fabric Manager User Guide.* In addition the FM configuration file must have the MeshTorusToplology specified, Refer to the "Mesh/Torus Topology" section of the *Intel® True Scale Fabric Suite Fabric Manager User Guide*.

This routing algorithm will use 1 LID for the secondary routing algorithm and 1 or more LIDs for the primary routing algorithm. The FM will automatically use an LMC>=1. If LMC is specified as 0 or 1, a value of 1 is used. If an LMC > 1 is used then 2^{LMC}-1 LIDs will be available for the primary routing algorithm. Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide*, on LMC for Mesh/Torus.

14.4.3 Mesh/Torus Fabric, OFED+ Configuration

The Distributed SA (`dist_sa`) must be enabled for autostart on all the compute nodes. Refer to "Enabling Distributed SA" on page 135. For more information on Distributed SA refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide*.

14.4.4 Mesh/Torus Fabric, Application and ULP Configuration

Mesh/Torus operates using IBTA compliant mechanisms. One of the key requirements of Virtual Fabrics is that applications make SA PathRecord queries to obtain critical address information such as Source Lid, Destination LID, PKey and SL.

Mesh/Torus has many of the same requirements as Virtual Fabrics as discussed previously, but the requirements are stricter in this case.

The FM will configure the Base LIDs to use the secondary routing algorithm. Therefore applications which do not perform SA PathRecord queries will not be able to take full advantage of the performance potential of the cluster.

14.4.4.1 MPI over PSM Configuration

For high performance MPI jobs on a Mesh/Torus, use of PSM is highly recommended. PSM must be configured to use Path Record queries and the Distributed SA.

Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for how to specify PSM_PATH_REC=opp and how to configure that variable globally on the compute nodes.

When using FastFabric to perform the `iba_host_admin`, `mpiperf`, or `iba_host_admin mpiperdeviation` verification steps, the `PSM_PATH_REC` option must be specified in `fastfabric.conf` file in the FF_MPI_ENV setting. Refer to Appendix B. Alternatively it can be specified in the `/opt/iba/src/mpi_apps/ofed.*.param` files,.

When using `/opt/iba/src/mpi_apps/run_*` scripts to run sample MPI applications and benchmarks, the `/opt/iba/src/mpi_apps/ofed.*.param` files must be edited to uncomment the PSM_PATH_REC setting.

14.4.4.2 MPI over Verbs Configuration

For historical reasons most MPI implementations over Verbs do not interact with the SA and instead hand construct address information using hardcoded PKey, SL, and other information.

Use of MPI over verbs is not recommended for Mesh/Torus. If MPI jobs are run using Verbs, they will end up using the Base LID (secondary route). Such routes will have higher latency and more congestion that the optimized primary routes which MPI over PSM will benefit from.

14.4.4.3 IPoIB Configuration

IPoIB performs PathRecord queries and will automatically use the proper Mesh/Torus routes.

14.4.4.4 Other Applications and ULPs

Most applications will not require any special operational changes. This is because, with the exception of MPI, all other applications tend to use IBTA standard approaches for connection establishment.

14.5 Adaptive Routing

Adaptive Routing allows the Intel switches to dynamically adjust the fabric routes in real time. The adjustments occur in response to observed traffic patterns and fabric congestion. This permits more efficient fabric operation.

14.6 Adaptive Routing, Switch Configuration

Make sure the version of switch firmware being used supports Adaptive Routing. Refer to *Intel® True Scale Fabric Switches 12000 Series Release Notes* for more information.

14.6.1 Adaptive Routing, Fabric Manager Configuration

Adaptive routing is configured completely in the Fabric Manager. Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for information on configuring adaptive routing.

14.7 Dispersive Routing

Dispersive Routing allows MPIs using Intel PSM technology to spread their traffic across multiple routes. This allows more of the fabric routes to be used by each HCA and also statistically improves the efficiency of the fabric and reduces potential congestion and bottlenecks.

14.7.1 Dispersive Routing, Fabric Manager Configuration

Use of Dispersive Routing requires the Fabric Manager to configure more than 1 route to each HCA. This is accomplished by enabling the LMC capabilities of the InfiniBand* architecture. To do this, set the Sm.Lmc configuration to 1 or more. This will cause up to 2^{LMC} alternate routes to be available per HCA port. Refer to the "LMC and Fabric Resiliency" section in the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for more information.

Within the Fabric Manager the PathSelection option can affect how many paths are reported. The default of Minimal is recommended. Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for more information.

Note: For Mesh/Torus fabrics, the number of alternate routes may be reduced. See "Mesh/Torus" on page 142 for more information.

14.7.2 Dispersive Routing, PSM Configuration

By default PSM will detect and use the multiple LIDs for dispersive routing. Additional controls over the routes used are possible through:

- Using Distributed SA and PathRecord queries in PSM - Refer to "Distributed SA" on page 145
- Configuring PSM_PATH_SELECTION - Refer to *Intel® True Scale Fabric OFED+ Host Software User Guide*.

14.8 Distributed SA

The Distributed SA allows for a highly scalable way for applications to query the FM's SA. This can help to improve flexibility of the fabric operation and configuration, especially when using Virtual Fabrics or Mesh/Torus.

The implementation involves dist_sa processes running on each compute node, these processes synchronize key PathRecord information with the FM such that is it always up to date and immediately available for rapid application startup.

14.8.1 Distributed SA, Fabric Manager Configuration

Refer to the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for information on configuring Distributed SA in the FM.

14.8.2 Distributed SA, OFED+ Configuration

Enable the Distributed SA (dist_sa) for autostart on all the compute nodes. Refer to "Enabling Distributed SA" on page 135 for information on enabling Distributed SA.

For more information on Distributed SA refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide.*

14.8.3 Distributed SA, Application and ULP Configuration

At this time, the Distributed SA is only integrated with Intel PSM for use by MPI jobs. The default `/etc/sysconfig/iba/dist_sa.conf` file matches the default `ifs_fm.xml` configuration with regard to IB ServiceIDs used for PSM.

14.8.3.1 MPI over PSM Configuration

14.8.3.1.1 Using PathRecord Query

Refer to the *Intel® True Scale Fabric OFED+ Host Software User Guide* for how to specify PSM_PATH_REC=opp and how to configure that variable globally on the compute nodes.

When using FastFabric to perform the `iba_host_admin`, `mpiperf`, or `iba_host_admin mpiperdeviation` verification steps, the PSM_PATH_REC option must be specified in `fastfabric.conf` file in the FF_MPI_ENV setting. Refer to Appendix B, "Configuration Files". Alternatively it can be specified in the `/opt/iba/src/mpi_apps/ofed.*.param` files.

When using `/opt/iba/src/mpi_apps/run_*` scripts to run sample MPI applications and benchmarks, the `/opt/iba/src/mpi_apps/ofed.*.param` files must be edited to uncomment the PSM_PATH_REC setting.

14.8.3.2 Other Applications and ULPs

At this time, no other applications will take advantage of the Distributed SA. Other applications will continue to operate as they have in the past. Those which are IBTA compliant will interact directly with the centralized SA.

§ §

Appendix A IFS Software Installation Checklist

The following sections provide a checklist to aid in tracking the steps as they are completed for Fabric Setup, Installation and verification. Check off each step as its performed. Refer to "Install the True Scale Fabric Suite Software" on page 21 for a more detailed explanation of each step.

A.1 Pre-Installation

Table 8. Pre-Installation

Step	Description	Complete
1.	Ensure hardware is installed, cabled, and powered. Refer to the *Intel® True Scale Fabric Switches 12000 Series Hardware Installation Guide.*	
2.	Ensure an HCA is installed in each server. Refer to the *Intel® True Scale Fabric Adapter Hardware Installation Guide*.	
3.	The hardware configuration should be reviewed to ensure everything was installed according to plan. Refer to the local hardware configuration plan.	
4.	Ensure the required Operating System is installed on each server with the following options: • Root user command prompt ends in "# " or "$ ". ***Note:*** There must be a space after # or $. • TCL and Expect packages installed on all Fabric Management Nodes. Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* for supported Operating Systems.	
5.	Ensure capability of remote login as root enabled. • SSH server enabled • All servers configured with the same root password	
6.	Ensure there is a TCP/IP Host Name Resolution • If using `/etc/hosts` update the `/etc/hosts` file on Fabric Management Node • If using DNS • All Management Network and IPoIB hostnames added to DNS • `/etc/resolv.conf` file configured on Fabric Management Node.	
7.	Ensure a NTP server is setup.	
8.	Proceed to the installation of the software using the installation checklist for your type of Installation: • "Install OFED+ Host Software on a Server" on page 147 • "Install Intel IFS on Management Node" on page 148 • "Install OFED+ Host Software on a Server" on page 147 • "Install OFED+ Host Software using Rocks" on page 148 • "Install OFED+ Host Software using a Platform HPC Kit" on page 148	

A.2 Install OFED+ Host Software on a Server

Table 9. Install OFED+ Host Software on a Server

Step	Description	Complete
1.	Complete the Pre-Installation Requirements. Refer to Table 8	
2.	Unpack Tar file on the Management Node. Refer to "Unpack the Tar File" on page 67	
3.	Install OFED+ Host Software on a Node. Refer to "Install OFED+ Host Software" on page 67	

A.3 Install OFED+ Host Software using Rocks

Table 10. Install OFED+ Host Software using Rocks

Step	Description	Complete
1.	Build the Frontend Node and Install the compute nodes with the Rocks Roll for OFED+ software. Refer to "Install Front-end and Compute Nodes" on page 79	
	OR	
	Add the Rocks Roll for OFED+ software on an existing frontend node. Refer to "Rocks Installation on an Existing Frontend Node" on page 80	

A.4 Install OFED+ Host Software using a Platform HPC Kit

Table 11. Install OFED+ Host Software using a Platform HPC Kit

Step	Description	Complete
1.	OFED+ Host Software using a Platform HPC Kit. Refer to "Install Intel Software Using the Platform Cluster Manager Kit" on page 83	

§ §

Appendix B Configuration Files

B.1 InfiniBand* and OpenFabrics Configuration Files

Table 12 contains descriptions of the configuration and configuration template files used by the InfiniPath and OFED software.

Table 12. InfiniBand* and OpenFabrics Configuration Files

`/etc/modprobe.conf`	Specifies options for modules when added or removed by the `modprobe` command. Also used for creating aliases. For Red Hat* systems only.
`/etc/modprobe.conf.local`	Specifies options for modules when added or removed by the `modprobe` command. Also used for creating aliases. For SLES systems only.

`/etc/infiniband/openib.conf`
 	
`/etc/sysconfig/infinipath`
	The primary configuration file for InfiniPath, OFED modules, and other modules and associated daemons. Automatically loads additional modules or changes IPoIB transport type.
	
	Contains settings, including the one that sets the `ipath_mtrr` script to run on reboot.

`/etc/sysconfig/network/ifcfg-`*NAME*	Network configuration file for network interfaces For SLES systems only.
`/etc/sysconfig/network-scripts/ifcfg-`*NAME*	Network configuration file for network interfaces For Red Hat* systems only.
`/usr/share/doc/initscripts-*/sysconfig.txt`	File that explains many of the entries in the configuration files For Red Hat* systems only

B.2 FastFabric Configuration Files

For a list and the description of the configuration files that are used by FastFabric refer to Appendix A of the *Intel® True Scale Fabric Suite FastFabric User Guide*.

§ §

Appendix C Multi-Subnet Fabrics

FastFabric supports management of both single-subnet fabric and multi-subnet fabrics.

When operating a multi-subnet fabric, a subnet manager (SM) is required for each subnet. A SM may be run within switches within each subnet, or a host-based SM may be run. A host-based SM can manage multiple subnets (assuming the host server is connected to more than one subnet).

For multi-subnet fabrics a number of combinations are possible:

1. **All subnets are completely independent (except for any interconnecting routers)**: If a separate FastFabric node is being used per subnet and servers are not installed in more than one subnet, the individual subnets can be treated completely separately. In this case, follow all the previous FastFabric instructions for each fabric.
2. **The subnets are primarily independent**: If the only components common to more than one subnet are the FastFabric node (and possibly SM nodes) and no servers are installed in more than one subnet, refer to the following instructions for "Primarily Independent Subnets".
3. **The subnets are overlapping**: If multiple components are common to more than one subnet, such as FastFabric node(s), servers, etc., refer to the following instructions for "Overlapping Subnets" on page 153.

C.1 Primarily Independent Subnets

If the FastFabric node (and possible SM nodes) is the only common server between subnets, FastFabric may be used to assist in server installation and fabric operation. Follow the installation instructions outlined in Section 3.0, "Install the True Scale Fabric Suite Software" on page 21 with the following adjustments:

From "Design of the Fabric" on page 17, design the cabling such that the FastFabric node will be connected to each subnet it will manage. The FastFabric node must also have a management network path to all the nodes in all the subnets that it will manage. As part of the design consider where routes between subnets are wanted between routers, IPoIB routers, etc.

"Design of the Fabric" on page 17 can be performed as per the instructions. When installing the IFS Software on the Fabric management node, IPoIB will need to be configured such that each subnet is an independent IPoIB network interface, typically with different IP subnets. Refer To the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information on configuring IPoIB.

Note: When managing a cluster where the IPoIB settings on the compute nodes are incompatible with the Fabric Management node (e.g., when a 4K MTU is used on the compute nodes and a 2K MTU is used on the Fabric Management Node), it is recommended not to run IPoIB on the Fabric management node(s).

"Configure Intel Chassis" on page 31 can be performed as per the instructions. When creating the chassis file, list all Intel 12000 series internally-managed switches in all subnets. If required, additional files may also be created per subnet that list only the Intel chassis in each subnet. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric.

"Install and Configure the Fabric Manager" on page 43 can be performed as per the instructions. At least one subnet manager is required per subnet. Refer To the *Intel®*

True Scale Fabric Suite Fabric Manager User Guide for more information on how to configure a host SM node to manage more than one subnet.

"Configure Intel Chassis" on page 31 can be performed as per the instructions. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric. If required the ibnodes file may specify a hca:port per switch. However, if hca:port is not specified, all the hca:port specified in the ports file will be searched to locate the given IB Switch's Node Guid.

"Install OFED+ Host Software on the Remaining Servers" on page 52 can be performed as per the instructions. When creating the `hosts` file, list the hosts in all subnets except the Fabric management node where FastFabric is being run. If required, additional files may also be created per subnet that list the hosts in each subnet (except the Fabric management node).

"Verify OFED+ Host Software on the Remaining Servers" on page 57 has the following adjustments from the instructions.

- **(All)**: Create the `allhosts` file as per the instructions. Next, create additional files per subnet that list all the hosts in each subnet including the Fabric management node. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric.
- **(All)**: "Verify hosts pingable" on page 54 can be performed as per the instructions.
- **(All)**: "Summary of Fabric Components" on page 59 can be performed as per the instructions.
- **(All)**: "Verify IB Fabric status and topology" on page 59 can be performed as per the instructions.
- **(Host)**: "Verify Hosts see each other" on page 60 can be run for each subnet by using the `allhosts` files specific to each subnet (i.e., those listing only hosts in a single subnet).
- **(Host)**: "Verify Hosts ping via IPoIB" on page 60may be run per the instructions.
- **(Host)**: "Check MPI Performance" on page 60 can be run for each subnet by using the `allhosts` files specific to each subnet (i.e., those listing only the hosts in a single subnet). This is currently not available on OFED.

"Installation of additional Fabric Management Nodes" on page 62 can be performed as per the instructions. When copying FastFabric configuration files to the additional Fabric management nodes, be sure to also copy the additional hosts, chassis and `allhosts` files that were created per subnet.

Note: In asymmetrical configurations where the Fabric management nodes are not all connected to the same set of subnets, the files copied to each management node may need to be slightly different. For example configuration files for `fabric_analysis` may indicate different port numbers or host files used for FastFabric and MPI may need to list different hosts.

"Configure and Initialize Health Check Tools" on page 63 can be performed as per the instructions. Additionally, make sure the `/etc/sysconfig/iba/ports` file lists each of the Fabric management node local HCAs and ports that are connected to a unique subnet. When running `iba_reports`, `fabric_info`, `fabric_analysis`, or `all_analysis`, the default will be to use the `ports` file. If required, the `-p` and `-t` options or the `PORTS/PORTS_FILE` environment variables may be used to specify all the HCAs and ports on the Fabric management node such that all subnets are checked.

Similarly, the `esm_chasssis` and `chassis` files used should list all relevant Intel chassis in all subnets.

"Running High Performance Linpack" on page 64 can be run for each subnet by creating `mpi_hosts` files specific to each subnet (i.e., only listing hosts in a single subnet).

"Upgrade the Management Node" on page 107 can be performed as per the instructions.

C.2 Overlapping Subnets

If multiple components are common between subnets (in addition to the Fabric management nodes), FastFabric may be used to assist in server installation and fabric operation. Follow the installation instructions outlined in "Install the True Scale Fabric Suite Software" on page 21 with the following adjustments:

From "Design of the Fabric" on page 17, design the cabling such that the FastFabric node will be connected to each subnet it will manage. The FastFabric node must also have a management network path to all the nodes in all the subnets it will manage. As part of the design consider where routes between subnets are required, between routers, IPoIB routers, etc.

"Set Up the Fabric" on page 18 can be performed as per the instructions. When installing the IFS software on the Fabric Management node, IPoIB will need to be configured such that each subnet is an independent IPoIB network interface, typically with different IP subnets. Refer to Section 3.5.7, "Configure IPoIB IP Address" on page 56 for more information on configuring IPoIB.

Note: When managing a cluster where the IPoIB settings on the compute nodes are incompatible with the Fabric management node (e.g., when a 4K MTU is used on the compute nodes and a 2K MTU is used on the management nodes), it is recommended not to run IPoIB on the Fabric management node(s).

"Configure Intel Chassis" on page 31 can be performed as per the instructions. When creating the chassis file, list all Intel 12000 series internally-managed switches in all subnets. If required, additional files may also be created per subnet that list only the Intel chassis in each subnet. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric.

"Install and Configure the Fabric Manager" on page 43 can be performed as per the instructions. At least one subnet manager is required per subnet. Refer To the *Intel® True Scale Fabric Suite Fabric Manager User Guide* for more information on how to configure a host-based SM node to manage more than one subnet.

"Configure Intel Chassis" on page 31 can be performed as per the instructions. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric. If required the ibnodes file may specify a `hca:port` per switch. However, if `hca:port` is not specified, all the hca:port specified in the ports file will be searched to locate the given Switch's Node Guid.

"Install OFED+ Host Software on the Remaining Servers" on page 52 can be performed as per the instructions. When creating the hosts file, list all the hosts in all subnets except the Fabric management node where FastFabric is being run. If required, additional files may also be created per subnet that list the hosts in each subnet (except the Fabric management node).

For hosts that are connected to more than one subnet, IPoIB will need to be configured such that each subnet is an independent IPoIB network interface, typically with different IP subnets. Refer To the *Intel® True Scale Fabric OFED+ Host Software User Guide* for more information on configuring IPoIB.

"Verify OFED+ Host Software on the Remaining Servers" on page 57 has the following adjustments from the instructions.

- **(All)**: Create the `allhosts` file per the instructions. Next, create additional files per subnet that list all the hosts in each subnet including the Fabric management node. When editing the ports file, list all the Fabric Management Node ports which access the managed fabrics. If required, additional files may also be created per subnet that list only the Fabric Management Node port connected to the given managed fabric.
- **(All)**: "Verify hosts pingable" on page 54 can be performed per the instructions.
- **(All)**: "Summary of Fabric Components" on page 59 can be performed as per the instructions.
- **(All)**: "Verify IB Fabric status and topology" on page 59 can be performed as per the instructions.
- **(Host)**: "Verify Hosts see each other" on page 60 can be run for each subnet by using the `allhosts` files specific to each subnet (i.e., those only listing hosts in a single subnet).
- **(Host)**: "Verify Hosts ping via IPoIB" on page 60 may be run per the instructions.
- **(Linux)**: "Refresh ssh Known Hosts" on page 60 may be run per the instructions.
- **(Host)**: "Check MPI Performance" on page 60 can be run for each subnet by using the `allhosts` files specific to each subnet (i.e., those listing only the hosts in a single subnet). This is currently not available for OFED.

"Installation of additional Fabric Management Nodes" on page 62 can be performed as per the instructions. When copying FastFabric configuration files to the additional Fabric management nodes, be sure to also copy the additional hosts, chassis and `allhosts` files created per subnet.

Note: In asymmetrical configurations where the Fabric management nodes are not all connected to the same set of subnets, the files copied to each management node may need to be slightly different. For example, configuration files for `fabric_analysis` indicating different port numbers or host files used for FastFabric and MPI may need to list different hosts.

"Configure and Initialize Health Check Tools" on page 63 can be performed per the instructions. In addition, make sure the `/etc/sysconfig/iba/ports` file lists the Fabric management node local HCAs and ports that are connected to a unique subnet. When running `iba_reports`, `fabric_info`, `fabric_analysis`, or `all_analysis`, the default is to use the ports file. If required, the `-p` and `-t` options or the `PORTS/PORTS_FILE` environment variable may be used to specify all the HCAs and ports on the Fabric management node such that all subnets are checked. Similarly, the `esm_chasssis` and `chassis` files used should list all relevant Intel chassis in all subnets.

"Running High Performance Linpack" on page 64 can be run for each subnet by creating `mpi_hosts` files specific to each subnet (i.e., only listing hosts in a single subnet).

"Upgrade the Management Node" on page 107 can be performed per the instructions.

§ §

Appendix D Install Lustre Software

This section contains information about additional third-party software installation.

Refer to the *Intel® True Scale Fabric OFED+ Host Software Release Notes* or *Intel® True Scale Fabric Suite Software Release Notes* for the Lustre cluster file system version this release supports. Lustre is a fast, scalable Linux* cluster file system that interoperates with the InfiniBand* architecture. For general instructions on downloading, installing, and using Lustre, go to:

http://www.lustre.org

Appendix E ./INSTALL Syntax

E.1 Intel OFED+ and IFS Installation

The `./INSTALL` command for the Basic and IFS installations are issued from the following directories:

- Intel OFED+ directory:

`IntelIB-Basic.`*`DISTRO`*`.`*`VERSION`*

- Intel IFS directory:

`IntelIB-IFS.`*`DISTRO`*`.`*`VERSION`*

Note: To access help for this command type `./INSTALL -?` and press **Enter**.

E.1.1 Syntax

```
./INSTALL [-r root] [-v|-vv] [-a|-n|-U|-F|-u|-s|-i comp|-e comp]
[-E comp] [-D comp] [-f] [--fwupdate asneeded|always]
[--user_configure_options 'options'] [--kernel_configure_options
'options'] [--prefix dir] [--no32bit|--32bit] [--without-depcheck]
[--rebuild] [--force] [--answer keyword=value]

or

./INSTALL -C

or

./INSTALL -V
```

E.1.2 Options

`-a` — Installs the software with the default options.

`-n` — Installs the software with the default options, but does not change the autostart options.

`-U` — Upgrades/re-installs all presently installed software with the default options, and does not change the autostart options.

`-i` *`comp`* — Installs a given component with the default options. This option can appear multiple times on a command line.

`-f` — Skips the installation of the HCA firmware upgrade during installation.

`-F` — Upgrades the HCA firmware during installation, with the default options.

`-u` — Uninstalls all ULPs and drivers with the default options.

`-s` — Enables autostart for all installed software.

`-e` *`comp`* — Uninstalls a given component with the default options. This option can appear multiple times on a command line.

-E *comp* — Enables autostart of given component. This option can appear with -D or multiple times on a command line.

This option can be combined with -a, -n, -i, -e and -U to permit control over which installed software will be configured for autostart.

-D *comp* — Disable autostart of given component. This option can appear with -E or multiple times once on command line.

This option can be combined with -a, -n, -i, -e and -U to permit control over which installed software will be disabled for autostart.

-C — Shows the list of supported component names.

-V — Outputs the version number of the software.

Additional options:

-r *dir* — Specify an alternate root directory. The default is /.

Note: This option is to permit boot images to be constructed that include Intel® True Scale Fabric Software so that the boot images can later be used for network boot of Intel True Scale Fabric enabled nodes.

Note: FastFabric use is not permitted in this environment.

--no32bit — Disable install of 32-bit libraries on 64-bit OSs

--32bit — Enable install of 32-bit libraries on 64-bit OSs

--rebuild — Force rebuild of OFED srpms

--user_queries — Permits non-root users to query the fabric. This is the default.

--no_user_queries — Non-root users cannot query the fabric.

--user_configure_options *options* — Supply additional OFED build options for user space srpms, this also forces a rebuild of all user space OFED srpms

--kernel_configure_options *options* — Supply additional OFED build options for kernel driver srpms, this also forces a rebuild of all kernel driver OFED srpms

--prefix *dir* — Specify alternate directory prefix for OFED installation. Default is /usr. This also causes a rebuild of related srpms.

--without-depcheck — Disable check of OS dependencies.

--force — Force installation even if distributions do not match. Use of this option can result in undefined behaviors

--fwupdate asneeded|always — Select fw update auto update mode:

asneeded — Sets mode for HCA Firmware update as part of -F, -a, -n or -I operation. All HCAs will be upgraded or downgraded to the currently supported revision, if needed.

always — Sets mode for HCA Firmware update as part of -F, -a, -n or -I operation. All HCAs will be upgraded or downgraded to the currently supported revision, even if it is already loaded with that revision.

Default is to upgrade as needed but not to downgrade.
This option is ignored for interactive install.

--answer keyword=*value* — Provides an answer to a question which might occur during the operation. Answers to questions which are not asked are ignored. Invalid answers will result in prompting for interactive installs or use of the default for non-interactive.

Possible Questions:

UserQueries — Permit non-root users to query the fabric

IntelSinglePort — Enable Intel HCA Single Port Mode default options retain existing configuration files supported component names:

```
ib_stack, mpi_selector, true_scale, ofed_mlx4, oftools, ib_stack_dev, fmtools,
fastfabric, ofed_ipoib, ofed_ib_bonding, ofed_sdp, ifs_fm, mvapich, mvapich2,
openmpi, mvapich_gcc_qlc, mvapich_pgi_qlc, mvapich_intel_qlc, mvapich2_gcc_qlc,
mvapich2_pgi_qlc, mvapich2_intel_qlc, openmpi_gcc_qlc, openmpi_pgi_qlc,
openmpi_intel_qlc, intel_shmem, mvapich_pathscale_qlc, mvapich2_pathscale_qlc ,
openmpi_pathscale_qlc, ofed_mpisrc, ofed_udapl, ofed_rds, ofed_srp, ofed_srpt,
ofed_iser, ofed_iwarp, opensm, ofed_nfsrdma, ofed_debug
```

Supported component name aliases:

```
iba, ipoib, mpi, verbs,_mpi, psm_mpi, mpidev, mpisrc, sdp, udapl, rds, inic,
ifibre, ibdev
```

Additional component names allowed for `-E` and `-D` options:

```
iba_mon, s20tune, ifs_fm_snmp, dist_sa
```

-v — Verbose logging. Logs to the `/var/log/iba.log` file.

-vv — Very verbose debug logging. Logs to the `/var/log/iba.log` file.

§ §

Appendix F Installing IEEL on top of IFS

The following section details the steps to install IEEL on top of IFS

F.1 Managed Mode

The Intel® Manager for Lustre* software lets you create and manage new HA Lustre file systems from its GUI. The other mode that is not described here is Monitor-only mode, which allows you to "discover" a working Lustre file system using Intel® Manager for Lustre* software. You can then monitor the file system at the Intel® Manager for Lustre* dashboard.

F.1.1 Key terms used in this document:

IML server (Intel® Manager for Lustre): Host on which we Install and run Manager for Lustre software. This host only hosts the manager software and doesn't run neither lustre softwares nor the infiniband drivers.

Agent server: Hosts which are managed by Intel® Manager for Lustre software and have Intel® Enterprise Edition for Lustre* software installed. Do not configure anything on these nodes manually unless specifically stated. These hosts are used for running MGS, MDS and OSS. All the Lustre configuration on these hosts is done by the manager.

Lustre Client: These are the hosts used to mount and access the Lustre file system.

F.1.2 Testbed:

Number of hosts used: a minimum of 4

- Host 1: For IML(Requires default yum repository configured; does not require HCA/OFED)
- Host 2: For configuring MGS and MDS (Requires two physical volumes, 10 GB each, to be configured, or these can be mounted on two different hosts instead of one)
- Host 3: For configuring OSS/OSTs.
- Host 4: For Lustre client.

Note: Hosts 2 and 3 are Agent servers.

F.1.3 Pre-requisites on IML server:

1. Install the OS.
2. Make sure a public `yum` repository is configured and working.
3. Populate the `/etc/hosts` file with all the `hostnames` present in the `clsuter`.
4. Disable `iptables/firewalld` and `selinux`. Make sure `iptables/firewalld` does not start at boot using `chkconfig`.

Note: IEEL supports both RHEL and Cent OS.

5. If using RHEL, make sure `yum` is functional and the default `yum` repositories are configured. For this, a RedHat subscription is required.
6. If using Cent OS, make sure your server is connected to the Internet. Default `yum` repositories work without the need of any subscription.

7. Make sure IFS (or any other cluster managing softwares) are not installed on the server.
8. The server hosting the Intel manager software doesn't need IB connections; everything is managed through Ethernet.

F.1.4 Pre-requisites on Agent servers:

1. Install the OS. While installing, create LVMs. LVMs can also be created after OS installation (see "Steps to create LVM after OS installation:" on page 165 for details).
2. Connect two Ethernet interfaces to same or different Ethernet switches but make sure both the Ethernet interfaces are up and running. Do not configure IP address on second Ethernet (`eth1`) port. It will be taken care by IML.
3. Populate the `/etc/hosts` file with the name and IP of all the hosts in the cluster.
4. Configure `yum` (this can also be managed with a local repository). Agent node does not need public `yum` repositories of Cent OS.
5. Disable `iptables/firewalld` and `selinux`. Make sure `iptables/firewalld` does not start at boot using `chkconfig`.
6. Install IB-Basic and reboot
7. Configure the ipoib (`ib0`) interface.

Note: Unlike native Lustre configuration, with IEEL it is required to install/build/compile OFED (IB-Basic preferably) on the RHEL/CentOS default kernel. Once this is done, Lustre is built against OFED and the Lustre kernel is loaded by the Intel Manager.

F.1.5 Steps to install IML and configure agents through IML.

1. Download the installation IEEL archive to a directory on the IML server (e.g. `/tmp`).
2. Unpack the archive using `tar: ieel-<version>.tar.gz`

```
# cd /tmp;
# mkdir install
# tar –C install –xzf ieel-<version>.tar.gz
```

3. To install the Intel® EE for Lustre* Software, run:

```
# cd /tmp/install/ieel-<version>
# ./install
```

4. When the prompts below appear, give inputs for the first superuser of Intel® Manager for Lustre*:

```
Username: <Enter the name of the superuser>
Email: <Enter an email address for the superuser>
Password: <Enter a password>
Confirm password: <Enter the password again>
```

Note: Please remember the username and password as they are required for authentication of the Intel manager for Lustre software GUI.

5. After successful installation, launch the Intel manager software GUI on any host within the network (enter the IP address of the IML server in the address bar of the host browser).
6. Click the **Configuration** tab and select **Servers**.
7. Click the **Add Server** box. A window displays prompting the user to enter the host name and root password.

8. Before adding the servers it will validate a few parameters in the compute node.
9. Make sure the validation of all the parameters passes. If the validation fails, a red square is displayed in the window. Click on the red square to see which parameter validation failed. Configure the required steps on the agent node manually and re-validate it through IML once again. It is recommended to not continue until the validation passes.
10. Click **Proceed** to deploy the agent to the host. Close the window once the agent is successfully deployed.
11. Select the appropriate server profile (i.e., the managed storage server) and click **Proceed** to setup the server.
12. Once the server setup is successfully completed, IML reboots the agent server. You see the servers and their lnet state as **up**. If you want Lustre to use the IB network, configure it by clicking **Configure LNet** and select the **Lustre Networks for IB Interface**. You can remove the Ethernet interfaces from the Lustre network. Refer to "Trouble shooting:" on page 165 if server setup fails.
13. The IML detects the volumes that were created during OS installation on the compute nodes. To get the volumes list, at the menu bar click the **Configuration** drop-down menu, then click **Volumes** to display the Volume Configuration page. A list of available volumes is displayed (if a volume does not contain unused block devices, it will not appear on this list).

    Add additional servers as needed through IML.
14. Now create the new Lustre file system. To create the file system, at the menu bar, click the **Configuration** drop-down menu and click **File Systems** to display the File System Configuration page. Click **Create File System** to display the New File System Configuration. Fill in the required fields to create the file system. On a Lustre Client host (The client requires `lustre-client-2.5.23-bundle.tar.gz` that is part of `ieel` tar file).

F.1.6 Configuring the Client

1. Install IFS and reboot the host.
2. Download the IEEL software. The overall release tar ball of IEEL software contains three packages. Use the following steps to extract the Lustre client package IEEL tar ball.

```
# tar -xvf ieel-<version>.tar.gz

# cd ieel-<version>

# tar -xvf lustre-client-<version>-bundle.tar.gz
```

3. After untarring the client package.

```
  # rpm2cpio
lustre-client-source-<version>.el6.x86_64.x86_64.rpm | cpio
-ivd

   # cd usr/

   # cd src/

   # cd lustre-<version>/
```

```
   # ./configure --disable-maintainer-mode
--with-o2ib=/usr/src/compat-rdma

   # make rpms
```

4. Rpms will be created in /root/rpmbuild/RPMS/x86_64/ directory. Go to this directory and install the rpms.

   ```
   # yum install lustre-client-modules-<ver>.<arch>.rpm
   lustre-client-<ver>.<arch>.rpm
   ```

5. Create a file "lustre.conf" in /etc/modprobe.d/ if not present already and add the following line to the file

   ```
   options lnet networks=o2ib0
   ```

6. Now bring up the lustre module by using modprobe.

   ```
   # depmod -ae

   # modprobe lustre
   ```

7. Now go to IML.
   a. Go to **Configuration**, **File Systems**.
   b. In the table listing available file systems, click the name of the file system to be accessed by the client. A page showing file system details will be displayed.
   c. Click **View Client Mount Information**. The mount command to be used on the client host to mount the file system will be displayed as shown in this example:

      ```
      # mount -t lustre 192.168.100.40@o2ib:/filesyst /mnt/
      ```

   d. Issue the above command on the client hosts to mount the file system.

F.2 Trouble shooting:

If the server setup fails on the Agent node and the LNet state shows `undeployed`, follow the steps below to attempt to resolve the issue.

In the setup servers procedure, after configuring the **Add Server** dialogue and clicking **Proceed**, a failed status message may display if the process was not successful.

1. Close the window.
2. On the **Server Configuration** page, the failed server is listed with its `LNet State` listed as `Unconfigured`.
3. To examine the cause of the failure, open the **Status** window, which displays the command (in red) Setting up host <hostname>.
4. Click **Detail** to open the Command detail window.
5. Under **Operations**, click the failed operation (i.e., a red exclamation). This will expand the operation to reveal the stack trace for that operation. You can examine the stack trace to learn the cause of the failure.
6. After correcting the cause of the failure, return to the **Server Configuration** page.
7. For the server that failed to be added, click **Actions** and select **Setup Server**.
8. Then **Setup Server** dialogue opens and the procedure starts again. The server should now add successfully.

   Example:

   a. The issue I've seen on my server was because of missing package and the error looks like this:

   ```
   Error: Package: chrome-agent-management-2.1.1.1-3893.noarch
   (iml-agent)
   Requires: python-jinja2
   Exception: Error running command: yum install -y
   chroma-agent chroma-diagnostics chroma-agent-management
   ```

   Installing the `python-jinja2` along with its dependency package `python-babel` manually on the compute node fixed this issue.

   b. Another similar issue I have observed is failing to update the `libcom_err` package. This error can be eliminated by removing `libcom_err-devel` rpm from the host using `yum`.

   After fixing any issue go to **Configuration**, **Servers**, **Actions** (of the desired server), **Setup Server**.

F.2.1 Steps to create LVM after OS installation:

1. Issue the `fdisk -l` command to see the available disks and/or partitions.
2. Use any disk that has free space for LVM partition creation:

   ```
   # fdisk /dev/sdb
   ```

   At the Linux `fdisk` command prompt,

   press `n` to create a new disk partition,

   press `p` to create a primary disk partition,

   press `1` to denote it as 1st disk partition,

press ENTER to accept the default of 1st and enter the last cylinder you want to have.

press t (will automatically select the only partition – partition 1) to change the default Linux partition type (0x83) to LVM partition type (0x8e),

press L to list all the currently supported partition type,

press 8e (as per the L listing) to change partition 1 to 8e (i.e. Linux LVM partition type),

press p to display the secondary hard disk partition setup. Please take note that the first partition is denoted as /dev/sdb1 in Linux,

press w to write the partition table and exit fdisk upon completion.

3. Next, this LVM command creates a LVM physical volume (PV) on a regular hard disk or partition:

```
pvcreate /dev/hdb1
```

Note: The cfdisk utility can also be used to create LVM.