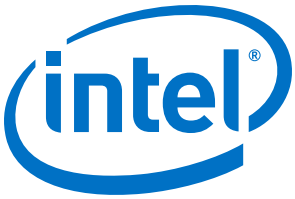
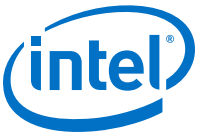


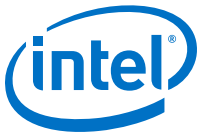
Intel® Omni-Path Fabric Suite FastFabric

User Guide

November 2015

Order No.: H76469-1.0

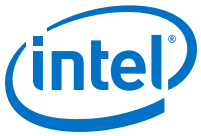




Revision History

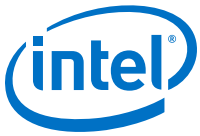
Date	Revision	Description
November 2015	1.0	Document has been updated for Revision 1.0
September 2015	0.7	Document has been updated for Revision 0.7.
April 2015	0.5	Alpha release of document.

Contents



Figures

Tables



Preface

This manual is part of the documentation set for the Intel® Omni-Path Fabric (Intel® OP Fabric), which is an end-to-end solution consisting of adapters, edge switches, director switches and fabric management and development tools.

The Intel® OP Fabric delivers a platform for the next generation of High-Performance Computing (HPC) systems that is designed to cost-effectively meet the scale, density, and reliability requirements of large-scale HPC clusters.

Both the Intel® OP Fabric and standard InfiniBand* are able to send Internet Protocol (IP) traffic over the fabric, or *IPoFabric*. In this document, however, it is referred to as *IP over IB* or *IPoIB*. From a software point of view, IPoFabric and IPoIB behave the same way and, in fact, use the same `ib_ipoib` driver to send IP traffic over the ib0 and/or ib1 ports.

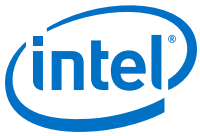
Intended Audience

The intended audience for the Intel® Omni-Path (Intel® OP) document set is network administrators and other qualified personnel.

Documentation Set

The following are the list of the complete end-user publications set for the Intel® Omni-Path product. These documents can be downloaded from https://downloadcenter.intel.com/.

- Hardware Documents:
  - *Intel® Omni-Path Fabric Switches Hardware Installation Guide*
  - *Intel® Omni-Path Fabric Switches GUI User Guide*
  - *Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide*
  - *Intel® Omni-Path Edge Switch Platform Configuration Reference Guide*
  - *Intel® Omni-Path Fabric Managed Switches Release Notes*
  - *Intel® Omni-Path Fabric Externally-Managed Switches Release Notes*
  - *Intel® Omni-Path Host Fabric Interface Installation Guide*
  - *Intel® Omni-Path Host Fabric Interface Release Notes*
- Software Documents:
  - *Intel® Omni-Path Fabric Software Installation Guide*
  - *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*
  - *Intel® Omni-Path Fabric Suite FastFabric User Guide*
  - *Intel® Omni-Path Fabric Host Software User Guide*
  - *Intel® Omni-Path Fabric Suite Fabric Manager GUI Online Help*



- *Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide*
- *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*
- *Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide*
- *Intel® Omni-Path Fabric Performance Tuning User Guide*
- *Intel® Omni-Path Host Fabric Interface Platform Configuration Reference Guide*
- *Intel® Omni-Path Fabric Software Release Notes*
- *Intel® Omni-Path Fabric Manager GUI Release Notes*

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, or section in this guide. Links to Web sites are also shown in blue. For example:

  See License Agreements on page 9 for more information.

  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, point to **Programs**, point to **Accessories**, and then click **Command Prompt**.
- Text in `Courier` font indicates a file name, directory path, or command line text. For example:

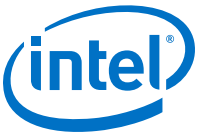
  Enter the following command: `sh ./install.bin`
- Key names and key strokes are shown in underlined bold uppercase letters. For example:

  Press **CTRL+P** and then 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*.

License Agreements

This software is provided under one or more license agreements. Please refer to the license agreement(s) provided with the software for specific detail. Do not install or use the software until you have carefully read and agree to the terms and conditions of the license agreement(s). By loading or using the software, you agree to the terms of the license agreement(s). If you do not wish to so agree, do not install or use the software.



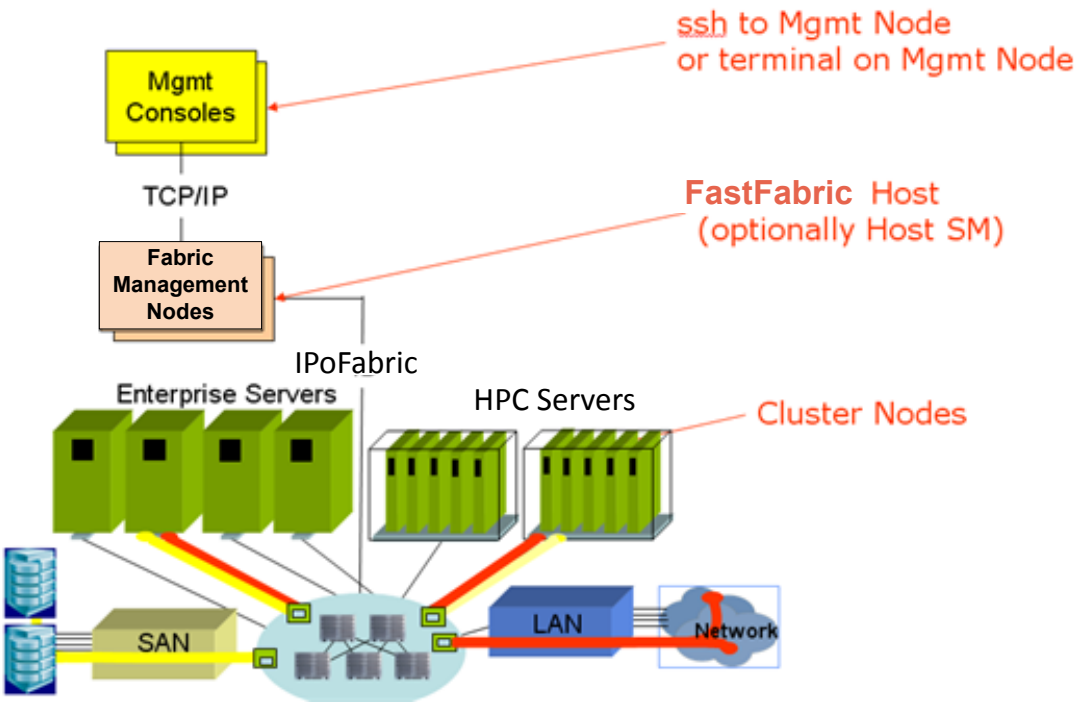
Technical Support

Technical support for Intel® Omni-Path products is available 24 hours a day, 365 days a year. Please contact Intel Customer Support or visit www.intel.com for additional detail.

1.0 FastFabric Architecture

FastFabric is typically installed on one or more Fabric Management Nodes. The Fabric Management Node must be connected to the rest of the cluster through the Intel® Omni-Path Fabric and a management network. The management network may be the primary Internet Protocol over InfiniBand* (IPoIB) network or Ethernet*. The management network is used for FastFabric host setup and administration tasks. It may also be used for other aspects of server administration or operation. Refer to the following figure for a high-level block diagram of the FastFabric architecture.

Figure 1. FastFabric Architecture



Depending on cluster size and design, the Fabric Management node may also be used as the master node for starting Message Passing Interface (MPI) jobs. It may also be used to run an Intel® Omni-Path Fabric Suite Fabric Manager and other management software. Refer to the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide* for details and what combinations are valid.

Note: When IPoIB is used as the management network, FastFabric is not able to install host software or configure IPoIB. However in this configuration, FastFabric is able to support host software upgrades, verification, and all its other features.

If remote access to FastFabric is desired, set up remote access to the Fabric Management Node using ssh, Telnet, X-Windows, VNC or any other mechanism that will allow the remote user to access a Linux* Command Line shell. Typically FastFabric is used only by cluster administrators.

1.1 How FastFabric Works

FastFabric manages two types of switching devices. These are reachable via the "Chassis Setup/Admin" and "Externally Managed Switch Setup/Admin" menus, respectively.

The Chassis menu allows management of switching devices that are termed "internally managed". These include both edge and director class switching devices that have one or more management cards in place. The management card provides an environment that exposes various TCP/IP services. This includes a command line interpreter login shell environment, with which FastFabric communicates. The device has an active Ethernet connection for LAN connectivity. The user is instructed to build a list of chassis in a "chassis" file, listing either the IP addresses or host names of the chassis to be managed; FastFabric provides tools to help discover such devices in the fabric and construct such a file. Communication with these devices is primarily out-of-band. <perhaps a reference to the Embedded CLI UG here? >

The Externally Managed Switch menu allows management of switching devices that are edge switches without management cards. Consequently, there is no environment present to provide TCP/IP services. There is no active Ethernet connection on the device. Therefore, communication to these devices must be accomplished in-band, via OPA management protocols designed specifically for this purpose. The user is instructed to build a list of switches in a "switches" file, listing the GUIDs of the switches to be managed; FastFabric provides tools to help discover such devices in the fabric and construct such a file.

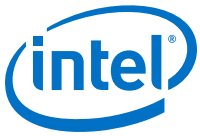
FastFabric consists of a variety of tools to administrate hosts, chassis and externally managed switches. Depending on the tool, the method of accessing and administering the target devices may differ.

The following table describes the access methods that FastFabric uses:

Table 1. FastFabric Methods

Method	Examples
Inband access	Fabric performance, error and congestion monitoring. Fabric topology reports, SA database queries, fabric error and link speed analysis, tools for externally managed switches, etc.
Log in through a management network	Host setup and installation, tools for internally managed chassis, etc.
MPI job startup (can be inband or through a management network)	Verify MPI performance, running sample MPI benchmarks, host-to-switch cable test.

Tools that log into other hosts will do so in a password-less manner using ssh. Tools that log into internally managed chassis can also use ssh. Chassis tools can prompt for a single password for all chassis, use password-less ssh, or can be pre-configured with the password. These approaches permit the tools to operate with minimal user interaction, and for this reason reduce the time to perform operations against many hosts or chassis.



After initial installation, FastFabric can be configured to use IPoIB instead of the management network.

Note: IPoIB cannot be used to reconfigure IPoIB or install new hosts.

2.0 FastFabric TUI Menu

The following sections describe the menu and related functions for the Intel® Omni-Path Fabric Suite FastFabric textual user interface (TUI).

2.1 FastFabric TUI Menu Overview

FastFabric is easiest to use from the textual user interface (TUI) menu system. The menu system provides a way to perform all common tasks and presents common options. Additional less common options are available directly, using the Command Line Tools, documented in the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*.

The following sections discuss the menu system. The majority of menu items directly invoke various FastFabric command tools. As such, the section on each menu item indicates what command tool it invokes and a summary of the operation performed. For further details about the given command tool, refer to the relevant section in the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*.

Menu items that are marked with **(Linux)** apply only when Linux* is being used. Similarly, some of the menu items are only applicable when Intel® Omni-Path Fabric Host Software is being used on the hosts, and will be marked with **(Host)**. All menu items that are applicable only when Intel Switches or Chassis are being used are marked with **(Switch)**. All remaining menu items are generally applicable to all environments and are not marked or may be marked with **(All)**.

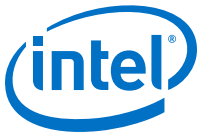
2.1.1 TUI Menu Usage

The textual user interface (TUI) menus are set up for ease of use. The submenus are designed to present operations in the order they would typically be used during an installation. Typing the keys corresponding to menu items (`0-9`, `a-d`) toggles the `Skip/Perform` selection for the given item. More than one item may be selected. Once the desired set of items has been selected, typing `P` performs the operations that were selected. To unselect all items, type `N`. Typing `X` or pressing `Esc` exits this menu and returns to the Main Menu.

If more than one item is selected, the items are performed in the order shown in the menu. This is the typical order desired during fabric setup. If you want to perform items in a different order, select a single item and type `P` to perform the operation by itself. Then repeat for the next operation to be performed. An opportunity is presented to abort the process after each item is selected, as follows:

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

If you press `Esc`, the sequence of operations ends and you return to the previous menu; pressing any other key results in the next selected menu item being performed. This prompt is also shown after the last selected item completes, providing an opportunity to review the results before the screen is cleared to display the menu.



At the top of each FastFabric menu, the directory and file listing the components to operate on is shown. For example:

```
Host File: /etc/sysconfig/opa/hosts
```

On each FastFabric menu, item `0` permits a different file to be selected and permits the editing of the file (using the editor selected by the EDITOR environment variable). In addition, it also permits review and editing of the `opafastfabric.conf` file. The `opafastfabric.conf` file guides the overall configuration of FastFabric and describes cluster-specific attributes of how FastFabric will operate. It is discussed in greater detail in Table 2 on page 53.

During the execution of each menu selection, the actual FastFabric command line tool being used is shown. This can be used as an educational aid to learn the command line tools.

2.2 Intel OPA Software Main Menu

The `Intel OPA Software` menu is the top level menu for the Intel OPA Software. It can be activated using the `opaconfig` command. This menu is not part of the FastFabric TUI. However, since it is one way of getting to the FastFabric Main Menu it is summarized here. The following is a example of the `Intel OPA Software` main menu.

Figure 2. Intel OPA Software Main Menu (Example)

```
Intel OPA X.X.X.X.X Software

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

   X) Exit
```

2.2.1 OPA Software Main Menu Items Description

Selecting items `1` through `7` will display the given submenu. Typing `X` will exit the menu system. The submenus are described below.

Show Installed Software

Menu item `Show Installed Software`, when selected, displays the `Intel OPA Installed Software` list and shows what is `Installed` and `Not Installed`.

Reconfigure OFED IP over IB

Menu item `Reconfigure OFED IP over IB`, when selected, proceeds through the reconfiguration of the OFED IPoIB configuration.

Reconfigure Driver Autostart

Menu item `Reconfigure Driver Autostart`, when selected, proceeds through the reconfiguration of the drivers autostart configuration.

Generate Supporting Information for Problem Report

Menu item `Generate Supporting Information for Problem Report`, when selected, proceeds through the process of generating a report and saving it to a user-specified file.

FastFabric (Host/Chassis/Switch Setup/Admin)

Menu item `FastFabric (Host/Chassis/Switch Setup/Admin)`, when selected, displays the `Intel FastFabric OPA Tools` menu. Refer to FastFabric Main Menu.

Uninstall Software

Menu item `Uninstall Software`, when selected, proceeds to the `Intel OPA Install Menu`.

2.3 FastFabric Main Menu

The `Intel FastFabric OPA Tools` menu is the starting point to manage the fabric using the textual user interface. Selecting 5 from the `Intel OPA Software` menu, or executing the `opafastfabric` command at a command prompt, displays the `Intel FastFabric OPA Tools` menu.

Note: Throughout the FastFabric Textual User Interface (TUI) and the following sections, "chassis" refers to internally managed switches, and "switches" refers to externally managed switches

Figure 3. Intel FastFabric OPA Tools Menu (Example)

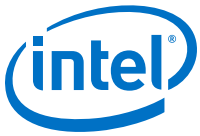
```
Intel FastFabric OPA Tools
Version: X.X.X.X.X

   1) Chassis Setup/Admin
   2) Externally Managed Switch Setup/Admin
   3) Host Setup
   4) Host Verification/Admin
   5) Fabric Monitoring

   X) Exit
```

2.3.1 FastFabric Main Menu Items Description

Selecting items `1` through `5` will display the given submenu. Typing `X` will exit the menu system. The submenus are described in the following sections.

Chassis Setup/Admin

Menu item `Chassis Setup/Admin`, when selected, displays the `FastFabric OPA Chassis Setup/Admin Menu`. Refer to FastFabric OPA Chassis Setup/Admin Menu for detailed information.



Externally Managed Switch Setup/Admin

Menu item `Externally Managed Switch Setup/Admin`, when selected, displays the `FastFabric OPA Switch Setup/Admin Menu`. Refer to FastFabric OPA Switch Setup/Admin Menu on page 24 for detailed information.

Host Setup

Menu item `Host Setup`, when selected, displays the `FastFabric OPA Host Setup Menu`. Refer to FastFabric OPA Host Setup Menu on page 29 for detailed information.

Host Verification/Admin

Menu item `Host Verification/Admin`, when selected, displays the `FastFabric OPA Host Verification/Admin Menu`. Refer to FastFabric OPA Host Verification/Admin Menu on page 33 for detailed information.

Fabric Monitoring

Menu item `Fabric Monitoring`, when selected, displays the `FastFabric OPA Fabric Monitoring Menu`. Refer to Fabric Monitoring Menu on page 38 for detailed information.

2.4 FastFabric OPA Chassis Setup/Admin Menu

This menu is focused on the initial setup and administration of Intel® Omni-Path architecture internally-managed switches. Press the keys corresponding to menu items (`0-9`) to toggle the `Skip/Perform` selection for the given item. You may select more than one item. After you select the desired set of items, type `P` to perform the operation(s). To unselect all items, type `N`. To exit this menu and return to the Main Menu, type `X` or press `Esc`.

Note: The alpha option selection parameters (a-e, n, p, and x) are all case *insensitive*.

To display the `FastFabric OPA Chassis Setup/Admin Menu`, select `1` from the `Intel FastFabric OPA Tools` menu (Figure 3 on page 16).

Figure 4. FastFabric OPA Chassis Setup/Admin Menu

```
FastFabric OPA Chassis Setup/Admin Menu
Chassis File: /etc/sysconfig/opa/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) Get basic Chassis configuration           [ Skip  ]
7) Configure Chassis Fabric Manager (FM)     [ Skip  ]
8) Update Chassis FM security files          [ Skip  ]
9) Get Chassis FM security files             [ Skip  ]
Admin:
a) Check OPA Fabric status                   [ Skip  ]
b) Control Chassis Fabric Manager (FM)       [ Skip  ]
c) Generate all Chassis Problem Report Info  [ Skip  ]
d) Run a command on all chassis              [ Skip  ]
Review:
e) View opachassisadmin result files         [ Skip  ]

P) Perform the selected actions              N) Select None
X) Return to Previous Menu (or ESC)
```

2.4.1 OPA Chassis Setup/Admin Menu Items Description

Select items `0` through `c` to change the item from `Skip` to `Perform`. To unselect all items, type `N`. To exit this menu, type `X` or press `Esc`. The items are described below.

Edit the Configuration and Select/Edit Chassis File

(Switch) The `Edit config and select/edit Chassis file` selection permits the `chassis`, `ports`, and `opafastfabric.conf` files to be edited. The chassis file selected and created by the menu lists the internally managed Intel switching chassis that are to be operated on. After editing the files, an opportunity is given to edit them again or to continue forward. The first file to review and edit is the FastFabric configuration file, `/etc/sysconfig/opa/opafastfabric.conf`:

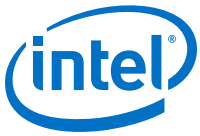
```
About to: vi /etc/sysconfig/opa/opafastfabric.conf
Hit any key to continue (or ESC to abort)...
```

The next file to review is the FastFabric ports file, `/etc/sysconfig/opa/ports`:

```
About to: vi /etc/sysconfig/opa/ports
Hit any key to continue (or ESC to abort)...
```

The next file to review is the chassis file. You are first asked to select the chassis file to use and offers a default. If a different chassis file needs to be edited, you need to type in the file location and name. Pressing `Enter` selects the specified chassis file.

```
Select Chassis File to Use/Edit [/etc/sysconfig/opa/chassis]:
About to: vi /etc/sysconfig/opa/chassis
Hit any key to continue (or ESC to abort)...
```



Commands such as opagenchassis can be useful while editing this file. For example in vi you could execute

```
:r! opagenchassis
```

which will execute the opagenchassis command and put its output directly at the current location in the file.

After exiting the vi editor, the TUI asks if you want to edit the chassis file again. Answer `n` to continue and return to the `FastFabric OPA Chassis Setup/Admin Menu`.

Refer to FastFabric Configuration File for more details about the format of the FastFabric configuration file. Refer to Port List File for more details about the format of the FastFabric ports file. Refer to Chassis List Files on page 57 for more details about the format of the chassis list file and about the `opagenchassis` command, which can help generate the chassis file.

Verify Chassis via Ethernet Ping

(Switch) The `Verify Chassis via Ethernet ping` selection runs the `opapingall -C -p -F` command to verify the existence of each selected chassis listed in the `chassis` file, using a ping over the management network.

Update Chassis Firmware

(Switch) The `Update Chassis firmware` selection runs the `opachassisadmin update` command to 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 using FastFabric.

Prompts will guide you through the options:

- `run` – Ensures given firmware is in the primary image and is running. As needed, push firmware to each chassis, select it for use and/or if it is not the presently running firmware, reboot the chassis
- `select` – Ensures given firmware is in the primary image. As needed, push firmware to each chassis and/or select it for use on next reboot
- `push` – Ensures given firmware is in the primary or alternate image. As needed, push firmware to each chassis but do not change selected nor running firmware

Additional options prompted for:

- selection of firmware files or directory containing `pkg` files
- parallel vs serial update
- chassis password (default is to have password in `opafastfabric.conf` or to use password-less ssh)

If any chassis fails to update, use the `View opachassisadmin result files` option to review the result files from the update. Refer to View opachassisadmin Results Files on page 24.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details on the `opachassisadmin update` command.

Set Up Chassis Basic Configuration

(Switch) The `Setup Chassis basic configuration` selection runs the `opachassisadmin configure` command, which prompts for chassis configuration settings and then configures all the selected chassis accordingly. The following aspects of chassis configuration may be set:

- Syslog Server IP Address, and Facility code
- NTP Server IP Address
- Time zone and Daylight Savings Time (DST)
- Link Width Supported
- OPA Node Description (configured to match chassis Ethernet* name)

  The OPA node description must be a string consisting of the characters A–Z, a–z, 0–9, and underscore. No spaces are allowed in the node description string, and it may not begin with a digit.
- Link CRC mode
- OPA Node Description format (concise format or verbose format)

Note: The chassis IP address must be set using the chassis serial port command line interface (CLI) during initial chassis installation and setup.

Set up Password-less ssh/scp

(Switch) The `Setup password-less ssh/scp` selection runs the `opasetup_ssh -p -S -C -F chassisfile` command. This sets up secure password-less SSH such that the Fabric Management Node can securely log into all the other chassis as admin through the management network without requiring a password.

Password-less SSH avoids the need to enter the chassis password for each FastFabric chassis operation. It also avoids the need to put the chassis password in `/etc/sysconfig/opa/opafastfabric.conf`. Once password-less ssh is set up, the password in the chassis may be changed without impacting the ability to use password-less ssh.

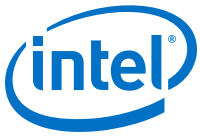
As part of this operation, the you will receive the following prompt:

```
Would you like to override the default Chassis Password? [n]:
```

If your response is `y`, you are prompted for the current password to use for logging into each chassis. Otherwise the password specified in `/etc/sysconfig/opa/opafastfabric.conf` is used.

Reboot Chassis

(Switch) The `Reboot Chassis` selection runs the `opachassisadmin -S -F chassisfile reboot` command to reboot each chassis listed in the `/etc/sysconfig/opa/chassis` file that was created in an earlier step. It also ensures that each chassis reboot is successful (as verified using ping over the management network).



Get Basic Chassis Configuration

(Switch) `Get basic Chassis configuration` retrieves basic information from the chassis, such as syslog, NTP configuration, time zone information, Link Width, Link CRC Mode, and node description. The following is an example of the information retrieved:

```
Performing Chassis Admin: Get basic Chassis configuration
Executing: /usr/sbin/opachassisadmin  -F /etc/sysconfig/opa/chassis getconfig
Executing getconfig Test Suite (getconfig) day mmm dd hh:mm:ss timezone yyyy ...
Executing TEST SUITE getconfig CASE (getconfig.xx.xx.xx.xx.getconfig) get
xx.xx.xx.xx ...
TEST SUITE getconfig CASE (getconfig.xx.xx.xx.xx.getconfig) get xx.xx.xx.xx
xx.xx.xx.xx:
    Firmware Active         : xx.xx.xx.xx
    Firmware Primary        : xx.xx.xx.xx
    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
    LinkWidth Support       : 4X
    Node Description        : Node_Name
    Link CRC Mode           : 48b_or_14b_or_16b
PASSED
TEST SUITE getconfig: 1 Cases; 1 PASSED
TEST SUITE getconfig PASSED
Done getconfig Test Suite day mmm dd hh:mm:ss timezone yyyy

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

Configure Chassis Fabric Manager (FM)

(Switch) The `Configure Chassis Fabric Manager(FM)` selection assists in configuring the Intel® Omni-Path Fabric Suite Fabric Manager (FM) for any member of the Intel® Omni-Path Chassis 100 Family.

Prompts first guide you through selection or generation of a `opafm.xml` file. When `generate` is selected, the `config_generate` command is used to guide you through selecting FM configuration options. Refer to the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide* for more information about `config_generate`.

Prompts guide you through selecting the configuration update options:

- `run` – Ensures the given FM configuration is on the chassis and running. As needed, pushes FM configuration to each chassis. It unconditionally restarts the FM on master Management Modules (MM) and makes sure it is not running on secondary MMs.
- `runall` – Ensures the given FM configuration is on the chassis and running. As needed, pushes the FM configuration to each chassis. It unconditionally restarts the FM on master and secondary MMs.
- `push` – Ensures the given FM configuration is on the chassis. As needed, pushes the FM configuration to each chassis.

Prompts also guide you through selecting the FM autostart options:

- `enable` – Enables FM start on the master MM upon chassis boot/reboot, disables FM autostart on any secondary MMs in selected chassis.
- `enableall` – Enables FM start on the master and any secondary MMs in the selected chassis upon boot/reboot.

- `disable` – Disables FM start on the master and any secondary MMs in the selected chassis upon boot/reboot.

Additional options prompted for:

- parallel vs. serial update
- chassis password (default is to have password in `opafastfabric.conf` or to use password-less ssh)

If any chassis fails update, use the `View opachassisadmin Results Files` option to review the result files from the update. Refer to View opachassisadmin Results Files on page 24

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Update Chassis FM Security Files

(Switch) The `Update Chassis FM security files` selection runs the `opachassisadmin fmsecurityfiles` command to permit the chassis security files to be verified and updated as needed.

Note: The FM security files are the private key, public key, and certificate files required by the FM to support secure socket connections to the Embedded Fabric Manager Fabric Executive (FE) by the Intel® Omni-Path Fabric Suite Fabric Manager GUI, and tools such as opafequery. Refer to the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide* for instructions regarding the administration tasks required to support these files.

Prompts will guide you through the options:

- `push` - Ensures given security files are pushed to each chassis

Additional options prompted for:

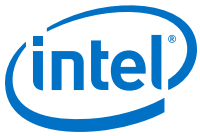
- selection of security files or directory containing pem files
- parallel vs serial update
- chassis password (default is to have password in opafastfabric.conf or to use password-less ssh)

If any chassis fails to be updated, use the `View opachassisadmin results files` option to review the result files from the update. Refer to View opachassisadmin Results Files on page 24 for more information.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details on the opachassisadmin update command.

Get Chassis FM Security Files

(Switch) The `Get Chassis FM security files` selection runs the `opachassisadmin fmgetsecurityfiles` command to permit the chassis FM security files to be retrieved from the chassis.



Check OPA Fabric Status

(Switch or All) `Check OPA Fabric status` allows the state and error counts of all ports to be checked and reviewed.

Once the prompts shown below have been answered, the `/sbin/opalinkanalysis` command is used.

```
Would you like to perform fabric error analysis? [y]:
Clear error counters after generating report? [n]:
Would you like to perform fabric link speed error analysis? [y]:
Check for links configured to run slower than supported? [n]:
Check for links connected with mismatched speed potential? [n]:
```

(All) The answer to `Would you like to perform fabric error analysis` selects whether `opareport -o errors` should be run. If you answer `y`, the `Clear error counters after processing report` question is asked. If you answer `y` to that question, the `-C` option is also used on `opareport` to clear the port error counters after doing the error analysis.

(All) The answer to `Would you like to perform fabric link speed error analysis` indicates whether `opareport -o slowlinks` should be run. If you answer `y` to this question, the `Check for links configured to run slower than supported` question is asked. If you answer `y`, the `-o misconfiglinks` option is also used for `opareport`. Additionally, if you answer `y` to `Would you like to perform fabric link speed error analysis`, the `Check for links connected with mismatched speed potential` question is also asked. If you answer `y`, the `-o misconnlinks` option is also used for `opareport`.

Intel recommends answering `y` to the first and third questions (these are the defaults for each prompt). This checks all the ports in the fabric for any links that have high error rates or are running at a lower speed than expected. Any identified links should be diagnosed and corrected.

Note: If the fabric is homogeneous and all links are expected to be running at full speed, answer `y` to the last two questions as well.

(Switch) If you respond `n` to all of the prompts, the `opashowallports -C` command is run to allow the state of all chassis ports to be manually reviewed.

Control Chassis Fabric Manager (FM)

(Switch) The `Control Chassis Fabric Manager (FM)` selection assists in controlling the FM for any Intel® Omni-Path Chassis 100 Family chassis. This operation is skipped for other chassis models.

Prompts guide you through selecting the control options:

- `restart` – Unconditionally restarts the FM on master Management Modules (MMs). It makes sure it is not running on secondary MMs.
- `restartall` – Unconditionally restarts the FM on master and secondary MMs.
- `run` – Makes sure FM is running on master MMs and ensures it is not running on secondary MMs.
- `runall` – Makes sure FM is running on master and secondary MMs.

- `stop` – Stops FM on master and secondary MMs.

Additional options prompted for:

- parallel vs. serial operation
- chassis password (default is to have password in `opafastfabric.conf` configuration file or to use password-less ssh)

If any chassis fails the operation, use the `View opachassisadmin result files` option to review the result files from the update. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Generate all Chassis Problem Report Information

(Switch) The `Generate all Chassis Problem Report Info` selection runs the `opacaptureall -C` command to collect configuration and status information from all selected chassis and generates a single `*.tgz` file that can be sent to a support representative.

Run a Command on all Chassis

(Switch) The `Run a command on all chassis` selection runs the `opacmdall -C` command. A Chassis CLI command may be set to execute against all selected chassis.

View opachassisadmin Results Files

(All) The `View opachassisadmin result files` selection permits viewing of the `test.log` and `test.res` files, which reflect the results from `opachassisadmin` runs (such as for updating Chassis Firmware or rebooting all chassis per menu items above). You are also given the option to remove these files after viewing them.

If not removed, subsequent runs of `opachassisadmin`, `opahostadmin`, or `opaswitchadmin` from within the current directory will continue to append to these files.

2.5 FastFabric OPA Switch Setup/Admin Menu

This menu is focused on administration of Intel® Omni-Path Edge Switches. Press the keys corresponding to menu items (`0-9, a-b`) to toggle the `Skip/Perform` selection for the given item. You may select more than one item. After you select the desired set of items, type `P` to perform the operation(s). To unselect all items, type `N`. To exit this menu and return to the Main Menu, type `X` or press `Esc`.

Note: The alpha option selection parameters (a, n, p, and x) are all case *insensitive*.

Select **2** from the Intel **FastFabric OPA Tools** menu (Figure 3 on page 16) to display the **FastFabric OPA Switch Setup/Admin Menu**.

Figure 5. FastFabric OPA Switch Setup/Admin Menu

```
FastFabric OPA Switch Setup/Admin Menu
Externally Managed Switch File: /etc/sysconfig/opa/switches

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  ]
Review:
a) View opaswitchadmin result files          [ Skip  ]

P) Perform the selected actions              N) Select None
X) Return to Previous Menu (or ESC)
```

2.5.1 OPA Switch Setup/Admin Menu Items Description

Pressing the keys corresponding to menu items (`0-9`) toggles the `Skip/Perform` selection for the given item. You may select more than one item. Typing `N` unselects all items. Type `X`, or press `Esc` to exit this menu and return to the Main Menu. The items are described in the following sections.

Edit Config and Select/Edit Switch File

(Switch) The `Edit config and select/edit Switch file` selection permits the `opafastfabric.conf`, `ports`, and `switches` files to be edited. The `opafastfabric.conf` file controls the default operation of the FastFabric Tools. The values specified in the `opafastfabric.conf` file specify the defaults used for all FastFabric operations. Existing environment variables override the values in this `ports` file. The `switches` file selected and created using this menu should list externally managed Intel® Omni-Path Switch 100 Family switches that are to be used. After editing the three files, you can edit them again or to continue.

```
Do you want to edit/review/change the files? [y]:
```

The default will repeat the editing process. Respond with `n` to continue.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details about the format of the `switches` and `ports` file, and about the `opagenswitches` command, which will help generate the switches file.

Generate or Update Switch File

(Switch) The `Generate or update Switch file` selection will generate or update the `/etc/sysconfig/opa/switches` file that can optionally be based on the `/etc/sysconfig/opa/topology.%P.xml` file(s) if the following prompt is answered `y`.

```
Do you want to update switch names based on
/etc/sysconfig/opa/topology.%P.xml file(s)? [y]
```

Test for Switch Presence

(Switch) The `Test for Switch Presence` selection runs the `opaswitchadmin ping` command to test for the presence of the selected switches in the fabric.

Verify Switch Firmware

(Switch) The `Verify Switch firmware` selection runs the `opaswitchadmin fwverify` command to verify the integrity of the present firmware in the switch. If this operation fails, prior to any switch reboots or power-offs of the switch, perform `Update Switch Firmware` to correct the firmware in the switch.

Update Switch Firmware

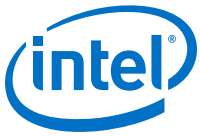
(Switch) The `Update Switch firmware` selection runs the `opaswitchadmin upgrade` command to permit the switch firmware version to be updated and the switch node name to be set.

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

Prompts guide you through the following options:

- Selection of firmware files or directory containing `.emfw` files

  Note: Ensure that the only `.emfw` file that is in the directory is the one you are using for the update.
- Reboot switch after update (needed to run new firmware)
- Parallel versus serial update
- Prompt for switch password



Note: Because the fabric itself is used to update externally managed switches, updating multiple switches with the reboot option may disrupt parallel update operations. If the selected switches file has the distance field properly filled in (`opagenswitches` will compute this field relative to the FastFabric node from which `opagenswitches` was run), then the update and reboot will be performed in a non-disruptive manner starting with the switches furthest from the FastFabric node and working back toward the FastFabric node. This will permit safe use of the parallel update option. Alternatively, if there are no selected externally managed switches in the path from the Fabric Management Node to any other externally managed switch, parallel operations may be used, for example, if a Fabric Management node is connected directly to a core switch and the externally managed switches are only at the edges. If in doubt, do not use parallel update. Be aware that non-parallel operation for a fabric with many externally managed switches could take a significant amount of time. Another alternative is to perform the update in parallel without a reboot and then perform the reboot separately using the `Reboot Switch` menu selection. To control the order of the rebooting of externally managed switches by FastFabric, refer to the discussion of the `distance` value in Externally Managed Switch List File on page 58.

If information fails to be reported for any switches, use the `View opaswitchadmin Result Files` option to review the result files from the update. Refer to View opaswitchadmin Result Files on page 29.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Set Up Switch Basic Configuration

(Switch) The `Setup Switch basic configuration` selection runs the `opaswitchadmin configure` command, which prompts for switch configuration settings and then configures all the selected Intel® Omni-Path Edge Switch 100 Family externally managed switches accordingly. The following aspects of switch configuration may be set:

- Link Width Supported
- Switch Description

  Note: Normally, the Switch Description is updated as part of a firmware upgrade; however, you are given the option to update the node description outside of an upgrade procedure.
- FM Enabled option

  Note: This can enable a host Fabric Manager to be connected to any port on the given switch. See the *Intel® Omni-Path Fabric Switches Hardware Installation Guide* for alternate ways to enable a host FM for individual ports
- Link CRC Mode

Note: This only operates on Intel® Omni-Path Edge Switch 100 Family externally managed switches.

Reboot Switch

(Switch) The `Reboot Switch` selection runs the `opaswitchadmin reboot` command to reboot all the switches listed in the `/etc/sysconfig/opa/switches` file that was created in a previous step.

Note: Because the fabric itself is used to reboot the externally managed switches, rebooting multiple switches with the reboot option may disrupt parallel operations. If the selected switches file has the distance field properly filled in (`opagenswitches` will compute this field relative to the FastFabric node from which `opagenswitches` was run), then the reboot will be performed in a non-disruptive manner starting with the switches furthest from the FastFabric node and working back toward the FastFabric node. This will permit safe use of the reboot option. Alternatively, if there are no selected externally managed switches in the path from the Fabric Management Node to any other externally managed switch, parallel operations may be used, for example, if a Fabric Management node is connected directly to a core switch and the externally managed switches are only at the edges. Be aware that non-parallel operation for a fabric with many externally managed switches could take a significant amount of time. To control the order of the rebooting of externally managed switches by FastFabric, refer to the discussion of the `distance` value in Externally Managed Switch List File on page 58.

Report Switch Firmware & Hardware Information

(Switch) The `Report Switch firmware & hardware info` selection will run the `opaswitchadmin info` command to provide a summary of the present state for all the selected switches.

The information reported by this option includes:

- Firmware Version
- Hardware Version
- Hardware Part Number
- Switch Capability
- Present Fan Status
- Present status of power supplies

If information fails to be reported for any switches, use the `View opaswitchadmin result files` option to review the result files from the update. Refer to View opaswitchadmin Result Files on page 29.

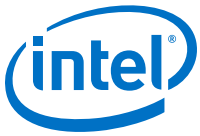
Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Get Basic Switch Configuration

(Switch) The `Get basic Switch configuration` selection runs the `opaswitchadmin -S -L /etc/sysconfig/opa/switches getconfig` command to retrieve basic information from an externally managed switch such as: MTU, VL Cap, Credit Distribution, Link Width, Link Speed, and node description.

The following files are produced from this selection:

- test.res – appended with summary results of run
- test.log – appended with detailed results of run
- save_tmp/ – contains a directory per failed operation with detailed logs
- test_tmp*/ – intermediate result files while operation is running



Report Switch VPD Information

(Switch) The `Report Switch VPD information` selection runs the `opaswitchadmin hwvpd` command to provide the Virtual Product Data (VPD) for all the selected switches. This information can be useful for inventory and asset control as well as to provide details about the product to customer support.

The information reported by this option includes:

- Serial Number
- Part Number
- Model Number
- Hardware Version
- Manufacturer
- Product description
- Manufacturer ID code
- Manufacture date
- Manufacture time of day

If information fails to be reported for any switches, use the `View opaswitchadmin Result Files` option to review the result files from the update. Refer to View opaswitchadmin Result Files on page 29.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

View opaswitchadmin Result Files

(All) The `View opaswitchadmin result files` selection permits viewing of the `test.log` and `test.res` files that reflect the results from `opaswitchadmin` runs (such as those for updating switch firmware, or for rebooting all switches per menu items above). You are also given the option to remove these files after viewing them.

If not removed, subsequent runs of `opachassisadmin`, `opahostadmin`, or `opaswitchadmin` from within the current directory continues to append to these files.

2.6 FastFabric OPA Host Setup Menu

This menu is focused on initial host setup and installation of Fabric software on all the hosts. Press the keys corresponding to menu items (`0-9`) to toggle the `Skip/ Perform` selection for the given item. You may select more than one item. After you select the desired set of items, type `P` to perform the operation(s). To unselect all items, type `N`. To exit this menu and return to the Main Menu, type `X` or press `Esc`.

Note: The alpha option selection parameters (a-d, n, p, and x) are all case *insensitive*.

Select `3` from the **Intel FastFabric OPA Tools** menu (Figure 3 on page 16) to display the **FastFabric OPA Host Setup Menu** (Figure 6 on page 30).

Figure 6. FastFabric OPA Host Setup Menu

```
FastFabric OPA Host Setup Menu
Host File: /etc/sysconfig/opa/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 OPA 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 opahostadmin result files            [ Skip  ]

P) Perform the selected actions              N) Select None
X) Return to Previous Menu (or ESC)
```

2.6.1 OPA Host Setup Menu Items Description

Select items `0` through `9` and `a` through `d` to change the item from `Skip` to `Perform`. To unselect all items, type `N`. To exit this menu, type `X` or press `Esc`. The items are described in the following sections.

Edit Configuration and Select/Edit Hosts File

(All) The `Edit config and select/edit Host file` selection permits the hosts and `opafastfabric.conf` files to be edited. The hosts file selected and created using this menu should not list the FastFabric host itself. After editing the two files, you can edit them again or continue.

```
Selected Host File: /etc/sysconfig/opa/hosts
Do you want to edit/review/change the files? [y]:
```

The default repeats the editing process; answer `n` to continue.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details about the format of the hosts file.

Verify Hosts Pingable

(All) The `Verify hosts pingable` selection runs the `opapingall` command. All the hosts listed are pinged through the Management Network.

Setup Password-less ssh/scp

(Linux) The `Setup password-less ssh/scp` selection runs the `opasetupssh -p -S -i ""` command. This will set up secure password-less SSH such that the Fabric Management Node can securely log into all the other hosts as root through the management network without requiring a password. You will be prompted for the

present password of the hosts; the same password will be used to log into all selected hosts. Once password-less ssh is set up, the password in the hosts can be changed without impacting the ability to use password-less ssh.

Password-less SSH is required by FastFabric, MPI test applications, and most versions of MPI (including QuickSilver MPI, OFED openmpi, and OFED mvapich2).

Copy /etc/hosts to all Hosts

(Linux) The `Copy /etc/hosts to all hosts` selection runs the `scpall /etc/hosts /etc/hosts` command to copy the `/etc/hosts` file on this host to all the other selected hosts. This is not necessary when using a DNS server to resolve hostnames for the cluster.

Show uname -a for all Hosts

(Linux) The `Show uname -a for all hosts` selection runs the `opacmdall "uname -a"` command to show the OS version on all the hosts. Review the results carefully to verify that all the hosts have the expected OS version. In typical clusters, all hosts are running the same OS and kernel version.

Install/Upgrade OPA Software

(Host) The `Install/Upgrade OPA Software` selection runs the `opahostadmin load` or `opahostadmin update` command to install the Intel® OPA software on all the hosts. By default, it looks in the current directory for *`FF_PRODUCT`*`.`*`VERSION`*`.tgz` file. If it is not found in the current directory, it prompts for input of a directory name where this file can be found.

Prompts will guide you through options:

- `upgrade` – updates all servers with new release. Only components previously installed are upgraded. Updates will fail for any hosts that have no Intel® OPA software currently installed
- `initial install/load` – uninstalls any existing Intel® OPA software, and installs the given release based on `opafastfabric.conf` installation options specified.

After the install is completed, the hosts must be rebooted to bring up the new drivers. This can be performed using the `Reboot Hosts` option (refer to Reboot Hosts on page 32).

If any hosts fail to be updated, use the `View opahostadmin result files` option (refer to View opahostadmin Result Files on page 33) to review the result files from the update. For more details, refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*.

Note: When using the Intel® OPA software packaging of the Open Fabrics Alliance (OFA) software, referred to as OFED Delta, the entire Intel® Omni-Path Fabric stack may be installed using FastFabric, which is the recommended approach. When using other packagings of OFA software, FastFabric may be used to install the Intel® Omni-Path Fabric Host Software (`IntelOPA-Basic.`*`DISTRO`*`.`*`VERSION`*`.tgz`) on the remaining hosts. To do so, the *`DISTRO`* is the OS on the node and *`VERSION`* must be the desired OFED Delta release level.

Note: The Intel® Omni-Path Fabric Host Software selected for installation must be appropriate for the OS version and distribution installed on the destination hosts.

Configure IPoIB IP Address

(Host) The `Configure IPoIB IP address` selection runs the `opahostadmin configipoib` command to create the `ifcfg-ib0` files on each host. The file will be created with a statically assigned IPv4 address. The IPoIB IP address for each host is determined by the resolver (Linux host command). If not found using the resolver, `/etc/hosts` on the given host is checked.

Build Test Applications and Copy to Hosts

(Host) The `Build Test Apps and copy to Hosts` selection builds the MPI sample benchmarks on the Fabric Management Node and copies the resulting object files to all the hosts. This is in preparation for execution of MPI performance tests and benchmarks in a later step.

Note: This option is available for the Intel® OPA software packaging of OFA software, but is not presently available for other packaging of OFA software.

Reboot Hosts

(Linux) The `Reboot Hosts` selection runs the `opahostadmin reboot` command to reboot all the selected hosts and to ensure they reboot fully (as verified using ping over the management network). When the hosts come back up, they will be running the software installed.

Refresh ssh Known Hosts

(Linux) The `Refresh ssh Known Hosts` selection will run the `opasetupssh -p -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.

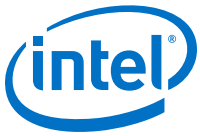
Rebuild MPI Library and Tools

(Host) The `Rebuild MPI library and tools` selection rebuilds the MPI Library and related tools (such as mpirun). This is performed using the `do_build` tool supplied with the MPI Source. When rebuilding MPI, `do_build` prompts you to select which MPI (openmpi or mvapich2) to rebuild, and provides choices as to which available compiler to use. Refer to the *Intel® Omni-Path Fabric Host Software User Guide* for more information.

Note: This option is available for the Intel®OPA software packaging of OFA software, but is not presently available for other packagings of OFA software.

Run a Command on all Hosts

(Linux) The `Run a command on all hosts` selection runs the `opacmdall` command. A Linux shell command (or sequence of commands separated by semicolons) may be specified to be executed against all selected hosts.



Copy a File to all Hosts

(Linux) The `Copy a file to all hosts` selection runs the `opascpall` command. A file on the local host may be specified to be copied to all selected hosts.

View opahostadmin Result Files

(All) The `View opahostadmin result files` selection permits viewing of the `test.log` and `test.res` files that reflect the results from `opahostadmin` runs (such as for installing software or rebooting all hosts per menu items Install/Upgrade OPA Software on page 31 and Reboot Hosts on page 32). You can also remove these files after viewing them.

If not removed, subsequent runs of `opachassisadmin`, `opahostadmin`, or `opaswitchadmin` from within the current directory continues to append to these files.

2.7 FastFabric OPA Host Verification/Admin Menu

The `FastFabric OPA Host Verification/Admin Menu` is focused on verifying hosts and the fabric, as well as administration of all the hosts. Press the keys corresponding to menu items (`0-9`, `a-d`) to toggle the `Skip/Perform` selection for the given item. You may select more than one item. After you select the desired set of items, type `P` to perform the operation(s). To unselect all items, type `N`. To exit this menu and return to the Main Menu, type `X` or press `Esc`.

Note: The alpha option selection parameters (a-d, n, p, and x) are all case *insensitive*.

Select `4` from the **Intel FastFabric OPA Tools** menu (Figure 3 on page 16) to display the **FastFabric OPA Host Verification/Admin Menu** (Figure 7 on page 33).

Figure 7. FastFabric OPA Host Verification/Admin Menu

```
FastFabric OPA Host Verification/Admin Menu
Host File: /etc/sysconfig/opa/allhosts

Validation:
0) Edit config and select/edit Host file            [ Skip  ]
1) Summary of Fabric components                     [ Skip  ]
2) Verify hosts are pingable, sshable, and active   [ Skip  ]
3) Perform single host verification                 [ Skip  ]
4) Verify OPA 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 opahostadmin result files                   [ Skip  ]

P) Perform the selected actions             N) Select None
X) Return to Previous Menu (or ESC)
```

2.7.1 OPA Host Verification/Admin Menu Items Description

Press the keys corresponding to menu items (0-9, a-d) to toggle the Skip/Perform selection for the given item. To unselect all items, type N. To exit this menu and return to the Main Menu, type X or press Esc.The items are described below.

Edit Config and Select/Edit Hosts File

(All) The Edit config and select/edit Host file selection permits the allhosts, ports, and opafastfabric.conf files to be edited. The allhosts file selected and created using this menu lists the FastFabric host itself. After editing the three files, you have the opportunity to edit them again or continue.

The default repeats the editing process, enter n to continue.

Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details about the format of the allhosts and ports files.

Summary of Fabric Components

(All) The Summary of Fabric components selection runs the opafabricinfo command to provide a brief summary of the counts of components in the fabric including how many switch chips, HFIs, and links are in the fabric. It also indicates whether any degraded or omitted (quarantined or out of policy) links were found (that could indicate a poorly seated or bad cable). Review the results against the expected configuration of the cluster.

Note: The link count includes both external and internal links within the switch boxes. This means that the count displayed is greater than the actual number of cables.

Verify Hosts Pingable, sshable and Active

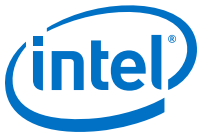
(All) The Verify hosts pingable, sshable and active selection runs the opapingall command. All the hosts listed are pinged through the Management Network.

Perform Single Host Verification

(All) The Perform single host verification selection looks at the host file selected by menu option "0" and performs a verification of node configuration, performance, and stability using a variety of tools and checks including single node HPL. The verification is performed on all nodes in the selected host file. It checks the configuration of each node for proper settings and configurations of the HFI and other system hardware. For additional information on the verification that is performed, refer to the /opt/opa/samples/hostverify.sh file..

Verify OPA Fabric Status and Topology

(Host or All): The Verify OPA Fabric status and topology selection allows the state and error counts of all ports to be checked and reviewed.



Based on the answers to the prompts shown in the following example, either the `opashowallports` or the `opareport` command is used.

```
Would you like to perform fabric error analysis? [y]:
Clear error counters after generating report? [n]:
Would you like to perform fabric link speed error analysis? [y]:
Check for links configured to run slower than supported? [n]:
Check for links connected with mismatched speed potential? [n]:
Would you like to verify fabric topology? [y]:
Verify all aspects of topology (links, nodes, SMs)? [y]:
Include unexpected devices in punchlist? [y]:
Enter filename for results [/root/linkanalysis.res]:
```

(All): The answer to `Would you like to perform fabric error analysis` selects whether `opareport -o errors` should be run. If you enter `y`, the `Clear error counters after generating report` question is asked. If you enter `y`, the `-C` option is also used on `opareport` to clear the error counters after doing the error analysis.

(All): The answer to `Would you like to perform fabric link speed error analysis` indicates when `opareport -o slowlinks` should be run. If you enter `y` to this question, the `Check for links configured to run slower than supported` question is asked. If you enter `y`, the `-o misconfiglinks` option is also used for `opareport`. Additionally, if you enter `y` to the `Would you like to perform fabric link speed error analysis` question, the `Check for links connected with mismatched speed potential` question is also asked. If you enter `y`, the `-o misconnlinks` option is also used for `opareport`.

Intel recommends entering the defaults for each prompt. This checks all the ports in the fabric for any links that have high error rates or are running at a lower speed than expected. Any identified links should be diagnosed and corrected.

(Host): If you enter `n` to all of the prompts, the `opashowallports` command is run to allow the state of all host ports to be manually reviewed. Selection of this option requires that Intel® Omni-Path Fabric Host Software be installed on all hosts being checked.

Verify Hosts See Each Other

(Host) The `Verify hosts see each other selection` runs the `opahostadmin sacache` command to verify that each host can see all the others through queries to the Subnet Administrator.

Note: This operation requires that the hosts being queried be specified by a resolvable TCP/IP host name. This operation will FAIL if the selected hosts are specified by IP address. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* "Selection of Hosts" section for more information.

Verify Hosts Ping via IPoIB

(Host) The `Verify hosts ping via IPoIB` selection runs the `opahostadmin ipoibping` command to verify that IPoIB is properly configured and running on all the hosts. This is accomplished through the Fabric management node pinging each host using IPoIB.

Refresh ssh Known Hosts

(Linux) The `Refresh ssh Known Hosts` selection runs the `opasetupssh -p -U` command to refresh the ssh known hosts list on this server for the IPoIB and Management Networks. This may be used to update security for this host if hosts are replaced, reinstalled, renamed, or repaired.

Check MPI Performance

(Host) The `Check MPI performance` selection does a quick check of PCI and MPI performance using end-to-end latency and bandwidth tests.

Note: This option is available for the Intel® OPA software packaging of OFA software, but is not presently available for other packaging of OFA software.

Based on the answer to the prompt shown in the following example either the `opahostadmin mpiperfdeviation` or the `opahostadmin mpiperf` command will be used.

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

Intel recommends answering `y`. This runs the `opahostadmin mpiperfdeviation` command to do pair-wise analysis of latency and bandwidth for the selected hosts and report 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 performance). Failing hosts are clearly indicated.

Answer `n` to run the `opahostadmin mpiperf` command. This displays the MPI latency and bandwidth between pairs of hosts (1-2, 3-4, 5-6, etc.). The numbers reported should be checked against the practical PCI speeds in the Performance Impact table in the *Intel® Omni-Path Fabric Software Installation Guide*. If any pairs are not in the expected performance range, it should be considered a failure for those pairs of hosts.
For either test, if any hosts fail, carefully check the bandwidth reported against the practical PCI speeds in the Performance Impact section. If all pairs are not in the expected performance range, carefully examine all hosts to verify the HFI models, PCI slot used, BIOS settings, and any motherboard jumpers related to devices on PCI buses or slot speeds.

The results of either test are also written to the `test.res` file, which may be viewed using the `View opahostadmin result files` selection. Refer to View opahostadmin Result Files on page 38.

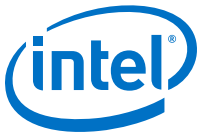
Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more details.

Check Overall Fabric Health

(Host) The `Check overall Fabric Health` selection runs the `opaallanalysis` command to check the overall fabric health.

You will be prompted:

```
Baseline present configuration? [n]:
```



If you enter `y`, a new baseline is created using the present fabric configuration. If you enter `n`, the present fabric state is checked against the baseline and the general health of the fabric is also checked.

Start or Stop Bit Error Rate Cable Test

(Host) The `Start or stop Bit Error Rate Cable Test` selection asks the following questions:

```
Stop or cleanup any already running Cable Test? [y]:
Stop HFI-Switch Cable Test? [y]:
Stop ISL Cable Test? [y]:
Start Cable Test? [y]:
Clear error counters? [y]:
Force Clear of hardware error counters too? [y]:
Start HFI-Switch Cable Test? [y]:
Number of Processes per host: [3]:
Start ISL Cable Test? [y]:
```

Answering all of the above questions using their defaults runs the following command:

```
/sbin/opacabletest  -A -n 3 -f '/etc/sysconfig/opa/allhosts'  stop_fi stop_isl
start_fi start_isl
```

Generate all Hosts Problem Report Info

(Host) The `Generate all Hosts Problem Report Info` selection runs the `opacaptureall` command to collect configuration and status information from all hosts and generates a single `*.tgz` file, which can be sent to a support representative.

Based on the answer to the prompt shown int the following example, 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 `opareport`.
- `3-Fabric+FDB` – In addition to "Fabric", also obtains all the switch forwarding tables and InfiniBand* multicast membership lists from the SM.
- `4-Analysis` – In addition to "Fabric+FDB", also obtains `opaallanalysis` results. If `opaallanalysis` 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 FM. 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 opacaptureall command. The information is gathered for every fabric specified in the /etc/sysconfig/opa/ports file.

Run a Command on all Hosts

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

View opahostadmin Result Files

(All) The View opahostadmin result files selection permits viewing of the test.log and test.res files, which reflect the results from opahostadmin runs (such as those for installing software or rebooting all hosts per menu items above). You can are also remove these files after viewing them.

If not removed, subsequent runs of opachassisadmin, opahostadmin or opaswitchadmin from within the current directory continue to append to these files.

2.8 Fabric Monitoring Menu

The Fabric Monitoring menu is focused on monitoring the performance of the fabric. Press the key corresponding to menu item (0) to toggle the Skip/Perform selection for the given item. You may select more than one item. After you select the desired set of items, type P to perform the operation(s). To unselect all items, type N. Type X or pressing ESC to exit this menu and return to the Main Menu.

Select 5 from the Intel FastFabric OPA Tools menu (Figure 3 on page 16) to display the FastFabric OPA Fabric Monitoring Menu (Figure 8 on page 38).

Figure 8. FastFabric OPA Fabric Monitoring Menu

```
FastFabric OPA Fabric Monitoring Menu

0) Fabric Performance Monitoring              [ Skip  ]

P) Perform the selected actions               N) Select None
X) Return to Previous Menu (or ESC)
```

2.8.1 Fabric Monitoring Menu Items Description

Selecting item 0 will change the item from skip to perform. Selecting N will unselect all items and X will exit the menu system. The item is described below.

Fabric Performance Monitoring

(All) The Fabric Performance Monitoring selection initiates opatop. For full details about opatop, refer to opa_top Fabric Performance Monitor.

3.0 Opatop Fabric Performance Monitor

`Opatop` is a command line tool that displays performance, congestion, and error information about a fabric. Fabric information is divided into two areas - performance and error statistics, which are the main starting points for analyzing fabric traffic. Performance (bandwidth utilization) can identify over-utilized areas (bottle necks) and under-utilized areas (potentially mis-configured); errors can identify problems in fabric hardware or configuration, as well as congestion and other performance situations.

3.1 opatop TUI

The opatop TUI screen layout and options consist of four areas in the layout. Image Identification, Screen Specific Information, Common Input Commands, and Screen Specific Input Commands. The following figure shows the top level summary screen as an example to show the screen layout. The section following the figure explains each of the areas and the common commands that are available on each screen.

Figure 9. opatop TUI Screen Layout (Example)

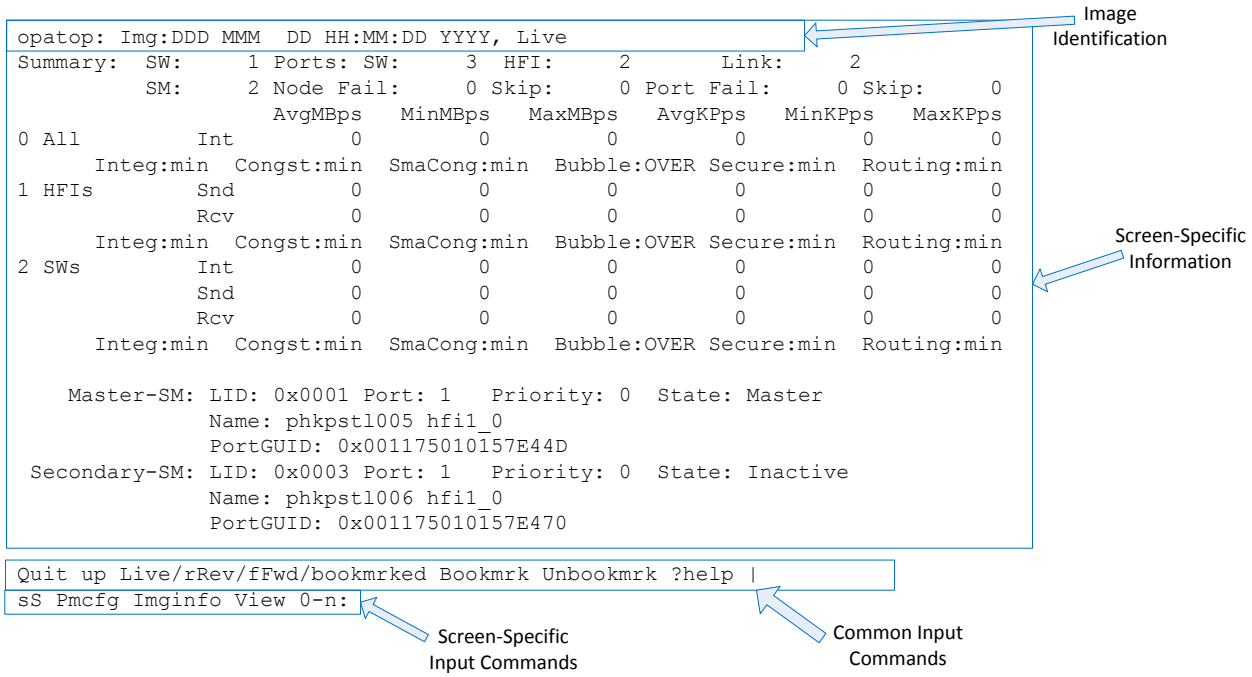


Image Identification

The first display line of opatop shows the timestamp for the PM sweep (image) being displayed and the type of image (Live, Hist, Bkmk). If a Live image is not being displayed, the current time ('Now:') is also shown.

Screen Specific Information

The information and layout of this area of the screen will vary depending on which screen is selected. The information and the layout of this section will be discussed for each specific screen in the following sections.

Command Entry

The last display line of opatop is a prompt showing available input commands. The left section of commands are available in every screen and perform the same action in each screen. The right section are screen-specific. Commands are case insensitive except as noted by `*`. The ENTER key must be pressed after multi-character commands and for `Quit`. Note that a help command, (`?`), is available at every screen, and provides information about the screen contents and input commands.

Common Input Commands

The following input commands are available in every screen:

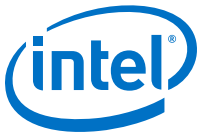
- `Q/q` – Quit program;
- `u*` – Up to previous screen;
- `L` – Select Live image;
- `R` – Navigate reverse 1 (`r*`) or 5 (`R*`) sweeps;
- `F` – Navigate forward 1 (`f*`) or 5 (`F*`) sweeps;
- `b*` – Select (previously) Bookmarked image;
- `B*` – Bookmark currently selected image;
- `U*` – Unbookmark Bookmarked image;
- `?` – Help

Screen-Specific Input Commands

The screen-specific input commands will be discussed with each screen description in the following sections.

Access to Live and Recent PM Historical Data

`opatop` allows you to access statistics from sequential PM sweeps (the PM keeps a history of previous sweep images) and queries the PM at a user-specified interval (10 seconds by default). When `opatop` queries for statistics for the most recent PM sweep, it is in "Live" mode. In Live mode the data will change, at the `opatop` interval rate, as `opatop` queries new PM sweeps. At each screen (summary or detail) the data being displayed is refreshed for the current PM sweep. A PM sweep can be in "frozen" mode. The data in a frozen sweep will not change, allowing the statistics to be examined in summary and detail screens.



opatop can access sweeps from the short term history database being recorded by the PM. This allows access to statistics from up to 24 hours in the past.

Two user actions result in a sweep being frozen. The first is when you "Bookmark" a sweep. A bookmarked sweep will remain frozen until you explicitly "Unbookmark" it; opatop allows one sweep at a time to be bookmarked. The second action is when you move (navigate) opatop's focus to another sweep within the history of sweeps maintained by the PM. For the duration of opatop's focus on such a sweep it will remain frozen. Navigation can occur from Live mode or when displaying a Bookmarked image; during navigation opatop is in "Historic" mode. Navigation can be performed backward or forward, 1 or 5 sweeps at a time.

3.2 opatop TUI Screens

Additional screens, described in the following paragraphs, are available to display detailed information about: PM configuration, PM sweep (image) configuration, performance statistics, error statistics, port group and/or virtual fabric configuration, and port statistics (port counters). The screens can be navigated in a hierarchal manner to examine the state of a fabric. The following figure illustrates the opatop TUI screen hierarchy.

Figure 10. opatop TUI Screen Hierarchy

```
* LEVEL:  SCREEN_NAME:
*   0     SCREEN_SUMMARY
*   1         SCREEN_PM_CONFIG
*   1         SCREEN_IMAGE_INFO
*   1         SCREEN_GROUP_INFO_SELECT
*   2             SCREEN_GROUP_BW_STATS
*   3                 SCREEN_GROUP_FOCUS
*   4                     SCREEN_PORT_STATS
*   2             SCREEN_GROUP_ERR_STATS
*   3                 SCREEN_GROUP_FOCUS
*   4                     SCREEN_PORT_STATS
*   2             SCREEN_GROUP_CONFIG
*   3                 SCREEN_PORT_STATS
*   0     SCREEN_VF_SUMMARY
*   1         SCREEN_PM_CONFIG
*   1         SCREEN_IMAGE_INFO
*   1         SCREEN_VF_INFO_SELECT
*   2             SCREEN_VF_BW_STATS
*   3                 SCREEN_VF_FOCUS
*   4                     SCREEN_PORT_STATS
*   2             SCREEN_VF_CONFIG
*   3                 SCREEN_PORT_STATS
*   2             SCREEN_VF_ERR_STATS
*   3                 SCREEN_VF_FOCUS
*   4                     SCREEN_PORT_STATS
```

3.2.1 opatop Summary Screen

The top level (summary) screen of opatop shows basic fabric configuration information as well as performance and error information. An example of the opatop summary screen is shown in the following figure.

Figure 11. opatop Summary Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Summary:  SW:     1 Ports: SW:     3  HFI:     2       Link:     2
          SM:     1 Node Fail:     0 Skip:     0 Port Fail:     0 Skip:     0
                    AvgMBps   MinMBps   MaxMBps   AvgKPps   MinKPps   MaxKPps
0 All         Int         24         0        31        24         0        30
      Integ:min  Congst:min  SmaCong:min  Bubble:OVER Secure:min  Routing:min
1 HFIs        Snd         31        31        31        30        30        30
              Rcv         31        31        31        30        30        30
      Integ:min  Congst:min  SmaCong:min  Bubble:OVER Secure:min  Routing:min
2 SWs         Int          0         0         0         0         0         0
              Snd         31        31        31        30        30        30
              Rcv         31        31        31        30        30        30
      Integ:min  Congst:min  SmaCong:min  Bubble:OVER Secure:min  Routing:min

    Master-SM: LID: 0x0001 Port: 1   Priority: 0  State: Master
               Name: phkpstl005 hfi1_0
               PortGUID: 0x001175010157E44D
 Secondary-SM: none



Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help |
sS Pmcfg Imginfo View 0-n:
```

Fabric Configuration Information

Fabric configuration information includes numbers of links, switches, SMs, and ports, as well as details about the master and secondary (if present) SMs.

Performance and Error Statistics for Each Port Group

Fabric performance and error statistics are presented based on three groupings of ports: All (all ports in the fabric), HFIs, and SWs. These groups provide a natural subdivision of the ports in a fabric for analysis. For more information about Groups and the operation of the PM, refer to the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*.

For each port group, average, minimum and maximum MBps (megabytes per second), and KPps (kilopackets per second) are shown, as well as a status indicator for each of five error categories.

Performance Statistics

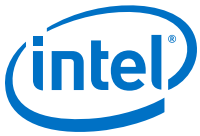
Performance statistics for each port group are further divided into up to three subgroups - Internal, Send, and Receive - based on whether a port's neighbor port is in its group. If a port's neighbor port is in its group, all performance statistics are contained in the Internal subgroup. If a port's neighbor is not in its group, statistics for data leaving the port (group) are contained in the Send subgroup and statistics for data entering the port are contained in the Receive subgroup.

All Group

In the All group, all ports are Internal because, by definition, the neighbor port must be in the All group.

HFIs Groups

In the HFIs groups, all neighbor ports are outside the group, so statistics are contained in the Send and Receive subgroups.



SWs Group

In the SWs group, neighbor ports are either outside the group (HFI) or inside the group (another switch), so statistics are contained in all three subgroups. A special case for a switch port is the special switch port 0, which is always considered internal to the SWs group

Error Categories

The error categories are:

- `Integ` – Integrity
- `Congst` – Congestion
- `Bubble` - Idles due to congestion
- `SmaCong` – SMA Congestion
- `Secure` – Security
- `Routing` – Routing

These error categories are each based on one or more port error counters. Each error category's status indicator is shown at one of five values/colors: minimum/green, Low/blue, Moderate/cyan, Warning/yellow or OVER/red based on the error value as compared to a threshold value.

Screen-Specific Input Commands

The summary screen accepts the following input commands:

- `P` – PM Configuration screen;
- `I` – Image Information screen
- `0-2` – Select Port Group - All (0), HFIs (1), SWs (2);

Additional Screens

After looking at the summary screen you can decide which area of the fabric (performance or error) and which port group or virtual fabric most warrants investigation, and can then drill down into that area.

3.2.2 PM Configuration Screen

The PM Configuration screen (Figure 12) displays information as provided by the PM (refer to the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*). The Sweep Interval parameter is separate from the opatop interval. Normally the opatop interval should be set to a value greater than or equal to Sweep Interval. The PM configuration screen shows the results for image information (total images, freeze images, freeze lease time), error thresholds, integrity weights, PM memory footprint, PMA MADs retry/timeout, and sweep information. The PM Configuration screen has no screen-specific input commands.

Figure 12. PM Configuration Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
PM Config:
 Sweep Interval: 10 sec  PM Flags(0x33):
   ProcessHFICntrs=On ProcessVLCntrs=On ClrDataCntrs=Off Clr64bitErrCntrs=Off
   Clr32bitErrCntrs=On Clr8bitErrCntrs=On
 Max Clients:        3
 Total Images:      10   Freeze Images: 5         Freeze Lease: 60 seconds
 Err Thresholds: Integrity:       100  Congestion:     100
                 SmaCongest:      100  Bubble:         100
                 Security:         10  Routing:        100
 Integrity Wts:  Loc Link Integ:    0  Rcv Errors:     100
                 Link Err Reco:     0  Link Downed:     25
                 Uncorrectable:   100  FM Config Err:  100
                 Link Qual:        40  Lnk Wdth Dngd:  100
                 Excs Bfr Ovrn:   100
 Congest Wts:    Tx Wait:          10  Cong Discards:  100
                 Rcv FECN:          5  Rcv BECN:         1
                 Tx Time Cong:     25  Mark FECN:       25
 PM Memory Size: 169 MB (169203442 bytes)
 PMA MADs: MaxAttempts:     3 MinRespTimeout:  35 RespTimeout:  250
 Sweep: MaxParallelNodes:  10 PmaBatchSize:     2 ErrorClear:     7

Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help |
```

3.2.3 Image Information Screen

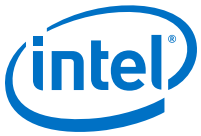
The Image Information screen (Figure 13) displays image information as provided by the PM. Sweep start and duration, numbers of ports in each group, node and port information for the sweep, and SM information is shown. The Image Information screen has no screen-specific input commands.

Figure 13. Image Information Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Image Info:
 Sweep Start: DDD MMM  D D HR:MM:SS YYY
 Sweep Duration: 0.013 Seconds

 Num SW-Ports:       3  HFI-Ports:       2
 Num SWs:            1  Num Links:       2  Num SMs:         1

 Num Fail Nodes:       0  Ports:       0  Unexpected Clear Ports: 0
 Num Skip Nodes:       0  Ports:       0

    Master-SM: LID: 0x0001 Port: 1   Priority: 0  State: Master
               Name: phkpstl005 hfi1_0
               PortGUID: 0x001175010157E44D
 Secondary-SM: none




Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help |
```

3.2.4 Group Information Select Screen

The Group Information Select screen (Figure 14) allows you to select the type of group information to display for the group selected in the summary screen. The following input commands are accepted and lead to the corresponding screen.



- `W` – Performance (Bandwidth Utilization) statistics
- `E` – Error statistics
- `C` – Group configuration (port list)

Figure 14. Group Information Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Group Info Sel: All
Int NumPorts: 4680  Rate Min:  any  Max: 100g
Ext NumPorts: 0
  Group BW Summary (W)
  Group Err Summary (E)
  Group Config (C)



Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | W E C:
```

3.2.5 Bandwidth Statistics Screen

The Bandwidth Statistics screen (Figure 15) displays, for each valid performance data subgroup (Internal, Send, Receive), the total, average, minimum and maximum MBps and KPps. For each subgroup, ten performance 'buckets', from 0+% to 90+% in 10% increments, count the number of ports whose 'MBps compared to link rate' value corresponds to that bucket. This provides an indication of how the data rate of the group compares to its potential.

Figure 15. Bandwidth Statistics Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Group BW Stats: HFIs   Criteria: Util-High  Number: 10

Snd:  TotMBps  AvgMBps  MinMBps  MaxMBps      TotKPps  AvgKPps  MinKPps  MaxKPps
           18        9        9        9           60       30       30       30

 Buckt 0+%   10+%   20+%   30+%   40+%   50+%   60+%   70+%   80+%   90+%
         2      0      0      0      0      0      0      0      0      0

Rcv:  TotMBps  AvgMBps  MinMBps  MaxMBps      TotKPps  AvgKPps  MinKPps  MaxKPps
           18        9        9        9           60       30       30       30

 Buckt 0+%   10+%   20+%   30+%   40+%   50+%   60+%   70+%   80+%   90+%
         2      0      0      0      0      0      0      0      0      0



Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | cC N0-n Detail:
```

The Bandwidth Statistics screen accepts input commands that specify parameters to be used in a group focus query, which will provide a list of ports (in the port group) sorted according to a specified performance criterion. The second line of the Bandwidth Statistics screen displays the group name, and the currently selected focus criterion and number of ports for a group focus query. The `D` command causes the group focus query to be performed and displayed in a Group Focus screen. The following input commands are accepted in the Bandwidth Statistics screen:

- `C` – Select group focus criterion forward (c*) or reverse (C*):

  - Util-High – Bandwidth Utilization (highest first)
  - UtlPkt-Hi – Packet Utilization (highest first)
  - Util-Low – Bandwidth Utilization (lowest first)
- Nn – Number of entries n in group focus list
- D – Display detail group focus list

3.2.6 Error Statistics Screen

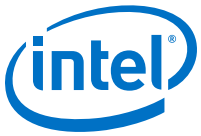
The Error Statistics screen (Figure 16) displays error statistics for a port group, divided into up to two subgroups, Internal or External, based on whether a port's neighbor port is in its group (Internal) or not (External). In the All group, all ports are Internal. In the HFIs group, all ports are External. In the SWs group, ports are Internal and External.

Figure 16. Error Statistics Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Group Err Stats: HFIs  Criteria: Integ  Number: 10

Ext                  Max       0+%      25+%      50+%      75+%     100+%
    Integrity          0         2         0         0         0         0
    Congestion         0         2         0         0         0         0
    SmaCongest         0         2         0         0         0         0
    Bubble          8022         0         0         0         0         2
    Security           0         2         0         0         0         0
    Routing            0         2         0         0         0         0
    Utilization:    0.4%  Discards:   0.0%




Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | cC N0-n Detail:
```

The six error categories are each based on one or more port error counters. The integrity and congestion error values are calculated by using a weighted sum. The weights for each and the threshold value for each error category can be seen in the PM Configuration screen (Refer to Figure 12). For more details about how the values for each error category is composed, refer to *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*.

- Integrity:
  - Link Quality Indicator
  - Link Width Downgrade
  - Local Link Integrity Errors
  - Port Receive Erros
  - Excessive Buffer Overrun Errors (neighbor port)
  - Link Error Recovery
  - Link Downed
  - Uncorrectable Errors



  - FM Config Errors

- Congestion:
  - Port Transmit Wait
  - Switch Port Congestion
  - Port Receive FECN (neighbor port)
  - Port Receive BECN (only from FIs)
  - Port Transmit Time Congestion
  - Port Mark FECN
- SmaCongestion:
  - The counters included in the SMA Congestion category are the VL 15 counters equivalent to the port counters in the Congestion category.
- Bubble:
  - Port Transmit Wasted Bandwidth
  - Port Transmit Wait Data
  - Port Receive Bubble (neighbor port)

- Security:
  - Port Receive Constraint Errors (neighbor port)
  - Port Transmit Constraint Errors

- Routing:
  - Port Receive Switch Relay Errors

For each error subgroup, five error "buckets" from 0+% to 100+% in 25% increments count the number of ports whose "error compared to error threshold" value corresponds to that bucket. This provides an indication of how error rates compare to their thresholds.

In addition, to aid analysis of congestion, Inefficiency, Discard percentage, and the percentage of congestion related discards to the total amount of discards are shown.

The Error Statistics screen accepts input commands that specify parameters to be used in a group focus query, which will provide a list of ports sorted according to a specified error criterion. The second line of the Error Statistics screen displays the group name and the currently selected focus criterion and number of ports for a group focus query. The `D` command causes the group focus query to be performed and displayed in a Group Focus screen. The following input commands are accepted in the Error Statistics screen:

- `C` – Select group focus criterion forward (c*) or reverse (C*):
  - `Integrity errors` (highest first)

  - — `Congestion errors` (highest first)
  - — `SmaCongestion errors` (highest first)
  - — `Bubble errors` (highest first)
  - — `Security errors` (highest first)
  - — `Routing routing` (highest first)
- `Nn` – Number of entries n in group focus list
- `D` – Display detail group focus list

3.2.7 Group Configuration Screen

The Group Configuration screen (Figure 17) displays a list of the ports in a group, including the LID, port number, port GUID, and NodeDesc of each. The second line of the screen displays the group name and the number of ports returned in the group configuration query. If more ports exist than will fit on a screen, the list can be scrolled forward and backward. An index value, shown with each port, can be used to select a port and show that port's counters in a Port Stats screen. The Group Configuration screen accepts the following input commands.

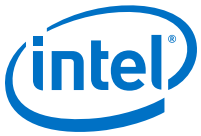
- `S` – Scroll forward (s*) or backward (S*) through port list
- `Pn` – Select port index value n

Figure 17. Group Configuration Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Group Config: All  NumPorts: 4680
  Ix  LIDx Port   Node GUID 0x   NodeDesc
    0 0001   1  001175010070BD58 <host_name> HFI-1
    1 0002   0  0002B30102000000 OPA-SWITCH0
    2 0002   1  0002B30102000000 OPA-SWITCH0
    3 0002   2  0002B30102000000 OPA-SWITCH0
    4 0002   3  0002B30102000000 OPA-SWITCH0
    5 0002   4  0002B30102000000 OPA-SWITCH0
    6 0002   5  0002B30102000000 OPA-SWITCH0
    7 0002   6  0002B30102000000 OPA-SWITCH0
    8 0002   7  0002B30102000000 OPA-SWITCH0
    9 0002   8  0002B30102000000 OPA-SWITCH0
   10 0002   9  0002B30102000000 OPA-SWITCH0
   11 0002  10  0002B30102000000 OPA-SWITCH0
   12 0002  11  0002B30102000000 OPA-SWITCH0
   13 0002  12  0002B30102000000 OPA-SWITCH0
   14 0002  13  0002B30102000000 OPA-SWITCH0
   15 0002  14  0002B30102000000 OPA-SWITCH0
   16 0002  15  0002B30102000000 OPA-SWITCH0
   17 0002  16  0002B30102000000 OPA-SWITCH0
   18 0002  17  0002B30102000000 OPA-SWITCH0
Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | sS P0-n:
```

3.2.8 Group Focus Screen

The Group Focus screen (Figure 18) displays a list of the ports you have selected to focus on within a group, including the LID, port number, focus criterion, port GUID and NodeDesc of each. If the port has a neighbor port, the same information is displayed for the neighbor. The second line of the screen displays the group name, the number of ports selected by in the combination of group, criteria, and requested ports, and the number of ports requested in the group focus query. If more ports exist



than will fit on a screen, the list can be scrolled forward and backward. Like the Bandwidth Statistics and Error Statistics screens that precede this screen, the focus criterion and number of requested focus ports can be changed to modify the focus port list. An index value, shown with each port, can be used to select a port and show that port's counters in a Port Stats screen. The Group Focus screen accepts the following input commands:

- `S` – Scroll forward (s*) or backward (S*) through port list
- `C` – Select group focus criteria forward (c*) or reverse (C*):
  - `Bandwidth Utilization` (highest first)
  - `Packet Utilization` (highest first)
  - `Bandwidth Utilization` (lowest first)
  - `Integrity errors` (highest first)
  - `Congestion errors` (highest first)
  - `SmaCongestion errors` (highest first)
  - `Bubble errors` (highest first)
  - `Security errors` (highest first)
  - `Routing errors` (highest first)
- `Nn` – Number of entries n in group focus list
- `Pn` – Select port index value n

Figure 18. Group Focus Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Group Focus: All   GrpNumPorts: 4680  NumPorts: 10  Number: 10
  Ix  Util-High LIDx Port   Node GUID 0x   NodeDesc
    0       0.0 0001   1  001175010070BD58 phgppriv18 HFI-1
  <->       0.0 0002  25  0002B30102000000 MOOSE_STL_SWITCH0
    1       0.0 0002   0  0002B30102000000 MOOSE_STL_SWITCH0
  <-> none
    2       0.0 0002   1  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0003   1  0002B30102000018 MOOSE_STL_SWITCH24
    3       0.0 0002   2  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0004   1  0002B30102000019 MOOSE_STL_SWITCH25
    4       0.0 0002   3  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0005   1  0002B3010200001A MOOSE_STL_SWITCH26
    5       0.0 0002   4  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0006   1  0002B3010200001B MOOSE_STL_SWITCH27
    6       0.0 0002   5  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0007   1  0002B3010200001C MOOSE_STL_SWITCH28
    7       0.0 0002   6  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0008   1  0002B3010200001D MOOSE_STL_SWITCH29
    8       0.0 0002   7  0002B30102000000 MOOSE_STL_SWITCH0
  <->       0.0 0009   1  0002B3010200001E MOOSE_STL_SWITCH30




Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help | sS cC N0-n P0-n:
```

3.2.9 Port Statistics Screen

The Port Statistics screen (Figure 19) displays a port's counters (performance and error). Error counters are grouped according to the error category to which they belong. A trailing asterisk ('*') on the counter name indicates the count will be used in computing Error Category information for the neighbor port. When the Port Statistics screen is entered from the Group Focus screen, port neighbor and link information is available. When the Port Statistics screen is entered from the Group Configuration screen this information is not available. The second line of the Port Statistics screen displays the group name, and LID and port number of the port, as well as link rate and MTU (if available). The third line of the screen displays the NodeDesc and Node GUID of the port. The fourth line of the screen displays the NodeDesc, LID and port number of the neighbor port (if available). The Port Statistics screen accepts the following input command (when neighbor information is available):

- `N` – Switch between statistics for port and port's neighbor

Figure 19. Port Statistics Screen (Example)

```
opatop: Img:Day Month Date HR:MIN:SEC YYYY, Live
Port Stats: All  LID: 0x1 PortNum: 1
NodeDesc: phgppriv18 HFI-1  NodeGUID: 0x001175010070BD58

 Xmit: Data:          7 MB (    972992 Flits) Pkts:       1872
 Recv: Data:          7 MB (    972992 Flits) Pkts:       1872
 Multicast: Xmit Pkts: 0          Recv Pkts: 0
 Integrity:                    | Congestion:
  Link Quality:             3 |  Cong Discards:          0
  Loc Lnk Integ:            0 |  Tx Wait:                0
  Rcv Errors:               0 |  Tx Time Cong:           0
  Excs Bfr Ovrn*:           0 |  Mark FECN:              0
  Lnk Err Recov:            0 |  Rcv FECN*:              0
  Link Downed:              0 |  Rcv BECN:               0
  Uncorrectable:            0 | Discards:
  FM Conf Err:              0 |  Tx Discards:            0
 Routing:                      | Bubble:
  Rcv Sw Relay:             0 |  Tx Wasted BW:           0
 Security:                     |  Tx Wait Data:           0
  Tx Constrain:             0 |  Rcv Bubble*:            0
  Rcv Constrain*:           0 |
 SmaCongestion (VL15):         |
  Tx Wait:                  0 |  Cong Discards:          0
  Mark FECN:                0 |  Rcv FECN*:              0
  Rcv BECN:                 0 |  Tx Time Cong:           0
Quit up Live/rRev/fFwd/bookmrked Bookmrk Unbookmrk ?help |
```

3.3 Command Line Options

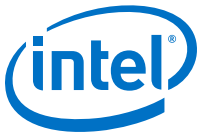
The following command line options are available for `opatop`:

Usage:

```
opatop [-v][-q] [-h hfi] [-p port] [-i seconds]
```

or

```
opatop --help
```



Options

- `--help` – Produce full help text
- `-v/--verbose` *`level`* – Verbose output level (additive):
  - `1` – Screen
  - `4` – STDERR (opatop)
  - `16` – STDERR PaClient
- `-q/--quiet` – Disable progress reports
- `-h/--hfi` *`hfi`* – HFI to send by, default is 1st HFI
- `-p/--port` *`port`* – Port to send by, default is 1st active port
- `-i/--interval` *`seconds`* – Obtain performance stats over interval seconds

4.0 Configuration of IPoIB Name Mapping

The FastFabric tools support the concept of a management network and an IPoIB network. For some clusters, the management network will be a low speed network such as 1gb or 10gb Ethernet. For other clusters, IPoIB may serve double duty as the host management network.

Note: When using IPoIB as the management network, the initial installation of Fabric software cannot be done using FastFabric.

The various FastFabric tools will translate from host names provided to and from IPoIB names as needed. This permits the host names given to be either management network or IPoIB network names. The default configuration file assumes that IPoIB host names are formed by adding a `-opa` suffix to the management network name. If a different suffix is desired, `FF_IPOIB_SUFFIX` can be changed. If IPoIB is also being used as the management network, `FF_IPOIB_SUFFIX` can be set to an empty string "".

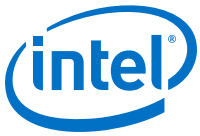
The translation is driven by the following functions within `opafastfabric.conf`:

`ff_host_basename` - given a management network or IPoIB hostname, translate to management network name, should match hostname -s

`ff_host_basename_to_ipoib` - given a management network name, translate to IPoIB hostname

More complex mappings can be specified by implementing alternate algorithms for these functions.

Note: When managing a cluster where the IPoIB settings on the compute nodes are incompatible with the Fabric Management node, it is recommended that you do not run IPoIB on the Fabric management nodes.



5.0 Configuration Files for FastFabric

Table 2 on page 53 lists the configuration files that are used by FastFabric. The description in the table also list the following sections that have detailed descriptions of each file. For a given release, refer to the files with `-sample` at the end of the file name for a sample file with the defaults of the given release. The sample files are installed into `/opt/opa/samples`.

Table 2. FastFabric Configuration Files

Configuration File	Description
`/etc/sysconfig/opa/opafastfabric.conf`	Overall configuration file. Refer to FastFabric Configuration File on page 53.
`/etc/sysconfig/opa/opamon.conf`	Error thresholds. Refer to Port Statistics Thresholds Configuration File on page 54.
`/etc/sysconfig/opa/opamon.si.conf`	Error thresholds related to Signal Integrity. Refer to Signal Integrity Thresholds Configuration File on page 55.
`/etc/sysconfig/opa/allhosts`	List of all hosts managed by FastFabric including the localhost. Refer to Host List Files on page 56.
`/etc/sysconfig/opa/hosts`	List of all hosts managed by FastFabric except the localhost. Refer to Host List Files on page 56.
`/etc/sysconfig/opa/chassis`	List of all chassis managed by FastFabric. Refer to Chassis List Files on page 57.
`/etc/sysconfig/opa/esm_chassis`	List of all chassis running an embedded SM that are to be monitored using esm_analysis. Refer to Chassis List Files on page 57.
`/etc/sysconfig/opa/switches`	List of all externally managed switches managed by FastFabric. Refer to Externally Managed Switch List File on page 58.
`/etc/sysconfig/opa/ports`	List of local HFI ports (for example subnets) to be used for fabric health analysis. Refer to Port List File on page 60.
`/etc/sysconfig/opa/topology.0:0.xml`	Fabric topology input file used by opareports and fabric health tools. Refer to Fabric Topology Input File on page 61.

5.1 FastFabric Configuration File

The FastFabric tools support a configuration file, `/etc/sysconfig/opa/opafastfabric.conf`. This file can be used to provide default settings for most of the FastFabric command line options. The configuration file is a bash shell script that will be included by each tool. As such, the file should be implemented such that environment variables defined, before the configuration file is executed, will not be altered. The sample displayed below makes use of the bash syntax such that only uninitialized variables are overwritten by the configuration file:

```
var= "${var:-value}"
```

An example of a sample file is provided, and matches the internal defaults of the FastFabric tools. For a given release refer to /opt/opa/samples/opafastfabric.conf-sample for a sample file with the defaults of the given release. If opafastfabric.conf does not assign a value to a given configuration variable, the default value will be used.

Note: Do not edit /opt/opa/samples/opafastfabric.conf-sample.

Other sample files are located in /opt/opa/samples.

The use of various configuration variables are discussed in the Environment Variables section for each command.

Note: Configuration files are self documented. Please see the particular configuration file of interest for information regarding its components and functions.

5.2 Port Statistics Thresholds Configuration File

The /etc/sysconfig/opa/opamon.conf configuration file defines port statistics thresholds for use by opareport, opafabricanalysis, opalinkanalysis, opaextractbadlinks, opaextractstat, opaextractstat2, and opaallanalysis.

This file lists a threshold for each port statistic. If the threshold for a given statistic is not defined or is set to 0, the given statistic will not be checked.

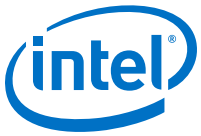
Note: When used by opareport or fabric health tools, the counts are absolute values and are applied against the counters as found in the system.

An example of a sample file is provided, and matches the internal defaults of the FastFabric tools. For a given release refer to /etc/sysconfig/opa/opamon.conf-sample for a sample file with the defaults of the given release.

Sample files are in /opt/opa/samples/*-sample, where "*" is a file such as opamon.conf , or opamon.conf .

Note: Do not edit any sample files.

```
# This file controls the opareport Port Counter Thresholds.
# [ICS VERSION STRING: @(#) ./fastfabric/samples/opamon.si.conf-sample
10_0_0_991_42 [10/01/15 03:28]

# This is a variation of the default opamon.conf file.  This file only
# checks error counters related to Signal Integrity.  Thresholds are set
# such that any and all non-zero counters will be visible.  This can be
# useful when using opareport -o errors, opaextracterror, and other
# related tools.  For many FastFabric tools this filename can be specified by
# the -c option.

#
# Error Counters are specified in absolute number of errors since last cleared.
# All Data Movement thresholds are specified in terms of absolute data
# over the monitoring interval.
#
# Setting a threshold to 0 disables monitoring of the given counter
#
# Output is generated when a threshold is exceeded.
#
```



```
# Counters for which a non-zero threshold is specified will be checked
# and potentially cleared by opareport and related tools using opareport.

Threshold                   Equal    # how to compare counter to threshold
                                     # Greater - reports values > threshold
                                     # Equal - reports values >= threshold
                                     # Does not apply to Link Quality Indicator

# Normal Data Movement
# --------------------
XmitData                          0  # as MB
RcvData                           0  # as MB
XmitPkts                          0  # as packets
RcvPkts                           0  # as packets
MulticastXmitPkts                 0  # as packets
MulticastRcvPkts                  0  # as packets

# Signal Integrity and Node/Link Stability
# ----------------------------------------
LinkQualityIndicator              4  # (range 0-5) higher number indicates
                                     # better quality.
                                     # Unlike other thesholds, links with a
                                     # value less (worse quality) than this
                                     # threshold are flagged
UncorrectableErrors               1  # indicate device internal instability
LinkDowned                        1
RcvErrors                         1
ExcessiveBufferOverruns           1  # can be side effect of SI
FMConfigErrors                    1  # can be a side effect of SI
LinkErrorRecovery                 1
LocalLinkIntegrityErrors          1
RcvRemotePhysicalErrors           0  # side effect of errors elsewhere, ignore

# Security
# --------
XmitConstraintErrors              0
RcvConstraintErrors               0

# Routing or Down nodes still being sent to
# ------------------------------------------
RcvSwitchRelayErrors              0
XmitDiscards                      0  # superset of CongDiscards

# Congestion
# ----------
CongDiscards                      0
RcvFECN                           0
RcvBECN                           0
MarkFECN                          0
XmitTimeCong                      0
XmitWait                          0

# Bubbles
# -------
XmitWastedBW                      0  # as MB
XmitWaitData                      0  # as MB
RcvBubble                         0
```

5.3 Signal Integrity Thresholds Configuration File

The `/etc/sysconfig/opa/opamon.si.conf` configuration file defines port counter signal integrity thresholds.

This file allows analysis for any non-zero error counters related to Signal Integrity (bad cables, etc) and can be used by adding the `-c` option to `opareport`, `opaextractbadlinks`, `opaextractstat`, `opaextractstat2`, `opalinkanalysis` and other related fastfabric tools.

An example of a sample file is provided at `/opt/opa/samples/opamon.si.conf`, and matches the internal defaults of the FastFabric tools.

Note: Do not edit `/opt/opa/samples/opamon.si.conf`.

5.4 Host List Files

The `/etc/sysconfig/opa/hosts` and `/etc/sysconfig/opa/allhosts` files are used to specify the hosts that FastFabric will operate against for many operations.

Alternate filenames may be specified in `opafastfabric.conf`, using environment variables or on the command line. Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

Below is a sample allhost list file:

```
# [ICS VERSION STRING: @(#) ./fastfabric/samples/allhosts-sample x_x_x_x_x [MM/DD/
YY
hh:mm]
# This file lists the TCP/IP names of ALL the hosts in the cluster.
# THIS SHOULD INCLUDE THE NODE RUNNING FASTFABRIC
#
# If Ethernet is being used for the management network, specify
# the hostname corresponding to the ethernet IP address.
# This file will be used by FastFabric to indicate which hosts should be
# operated on by various fastfabric menus and CLI commands.

include /etc/sysconfig/opa/hosts
# add line below with TCP/IP name of FastFabric host (eg. this host)
myadminhost
```

Each line of the host list file may specify a single host, a comment or another host list file to include.

Hosts may be specified by IP address or a resolvable TCP/IP hostname. Typically, hostnames are used for readability. Also, some FastFabric tools will translate the supplied host names to IPoIB hostnames, in which case names are generally easier to translate than numeric IP addresses. Typically, management network host names are specified. However, if desired, IPoIB hostnames or IP addresses may be used. This can accelerate large file transfers and other operations.

Files to be included may be specified using an `include` directive followed by a file name. File names specified should generally be absolute path names. If relative path names are used, they will be searched for within the current directory, then `/etc/sysconfig/opa` directory.

Comments may be placed on any line by using a `#` to precede the comment. On lines with hosts or include directives, the `#` must be white-space separated from any preceding host name, IP address, or included file name.

5.5 Chassis List Files

The `/etc/sysconfig/opa/chassis` and `/etc/sysconfig/opa/esm_chassis` files are used to specify the Intel chassis that FastFabric will operate against for many operations.

Alternate filenames may be specified in `opafastfabric.conf`, using environment variables or on the command line. Refer to *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

The following is a sample chassis file:

```
# [ICS VERSION STRING: @(#) ./fastfabric/samples/chassis-sample x_x_x_x_x
[MM/DD/YY hh:mm]
# This file lists the TCP/IP names of the Intel Internally Managed
# Switches in the cluster.
#
# If Ethernet is being used for the management network, specify
# the name corresponding to the ethernet IP address of the chassis.
# This file will be used by FastFabric to indicate which chassis should be
# operated on by various fastfabric menus and CLI commands.
```

Each line of the chassis list file may specify a single chassis, a comment, or another chassis list file to include.

Chassis may be specified by chassis management network IP address or a resolvable TCP/IP name. Typically, names are used for readability.

Files to be included may be specified using an `include` directive followed by a file name. File names specified should generally be absolute path names. If relative path names are used, they will be searched for within the current directory then `/etc/sysconfig/opa` directory.

Comments may be placed on any line by using a `#` to precede the comment. On lines with chassis or include directives, the `#` must be white-space separated from any preceding name, IP address or included filename.

The `opagenchassis` command can be used to help locate chassis in the fabric and generate a chassis file.

5.5.1 Selection of Slots Within a Chassis

Normally, operations are performed against the management card in the chassis. For operations such as `opacmdall`, the command is executed against the management interface for the given chassis. For more sophisticated operations, such as firmware update, a directory with firmware for each chassis card type can be supplied and all cards in the chassis will be updated with the appropriate firmware from that directory. However, in some cases it may be desirable to perform operations against a specific subset of cards within the chassis. In this case the chassis IP address, a name within a chassis list, or a chassis file can be augmented with a list of slot numbers to operate on. This is done in the form:

```
chassis:slot1,slot2,…
```

Note: There must be no spaces within the chassis name and/or slot list.

This format is used by `opacmdall` and chassis firmware update. This format may be used anywhere a chassis name or IP address is valid, such as the `-H` option, the `CHASSIS` environment variable, or chassis list files. The slot number specified is ignored on some operations (such as `opapingall`). Only slots containing management cards, may be specified with this format. For all Intel® Omni-Path Chassis 100 Family chassis, slot 0 is always an alias for the presently active management card for the chassis. For the remainder of slot usages in the chassis, the `chassisQuery` command can be executed against a given chassis to identify which slots have management cards.

Note: For any operation, care should be taken that a given chassis is listed only once with all relevant slots as part of that single specification. This is important so that parallel operations do not cause conflicting concurrent operations against a given chassis.

5.6 Externally Managed Switch List File

The `/etc/sysconfig/opa/switches` file is used to specify the externally-managed Intel switches that FastFabric will operate against for many operations.

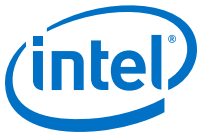
Alternate file names may be specified in `opafastfabric.conf`, using environment variables or on the command line. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

The following is a sample `switches` file:

```
# [ICS VERSION STRING: @(#) ./fastfabric/samples/switches-sample x_x_x_x_x
[MM/DD/YY hh:mm]
# This file lists all the Intel Externally Managed Switches
#
# specify one line per switch of the form guid,nodeDesc,distance
#  guid - node guid of the switch
#  nodeDesc - optional node description which should be programmed into
#      the switch by FastFabric.  It is recommended to supply a unique
#      nodeDesc for each switch to simplify management of the cluster.
#  distance - optional relative distance of the switch from the FastFabric node
#      this is used by reboot operations to first operate on switches furthest
#      from the FastFabric node.
#      Nodes without a distance specified will be treated as furthest.
# For fabrics with multiple IB subnets, the local hfi and port to use may be
# specified as: guid:hfi:port,nodeDesc,distance.
# See the FastFabric Manual for more info
#
# The opagenswitches tool can be used to query the SM and generate a list
# Externally Managed switches in the proper form for this file.
#
# for example:
# 0x00066a00e300299f,SwitchA1,2
```

Each line of the switch list file may specify a single switch, a comment, or another switch list file to include.

Switches can be specified by node GUID optionally followed by a colon and the hfi:port, optionally followed by a comma and the OPA Node Description (nodename) to be assigned to the switch, and optionally followed by the distance value indicating the relative distance from the FastFabric node for each switch.



The `opagenswitches` command can be used to help locate externally managed switches in the fabric and generate a `switches` file. The `opagenswitches` tool will by default provide the proper distance value relative to the FastFabric node from which it was run. This capability requires use of IBTA standard TraceRecord queries that are not supported by openSM, but can be supplied by the Intel® Omni-Path Fabric Suite Fabric Manager (FM). Alternatively the `opagenswitches -R` option can suppress generation of this field. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

In a typical pure fat tree topology, with externally managed switches as edge switches and internally managed switches as core switches, you can also easily manually specify proper distance by simply specifying 1 for the distance value of the switch next to the FastFabric node. Note that in such a topology, all other switches are an equal length from the FastFabric node, and a missing distance value will cause them to be treated as having a distance value that is larger than any other found in the file. Therefore the other switches would be rebooted first and the FastFabric node's switch would be rebooted last.

The GUID will be used to select the switch and on firmware update operations, the node description will be written to the switch such that other FastFabric tools (such as `opasaquery` and `opareport`) can provide a more easily readable name for the switch. The node description can also be updated as part of switch basic configuration.

The `hfi:port` may be used to specify which local port (subnet) to use to access the switch. If this is omitted, all local ports specified will be checked for the switch and the first port found to be able to access the switch will be used to access it. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information about how to specify an `hfi:port` value.

Files to be included may be specified using an `include` directive followed by a file name. File names specified should generally be absolute path names. If relative path names are used, they will be searched for within the current directory then `/etc/sysconfig/opa`.

Comments may be placed on any line by using a `#` to precede the comment. On lines with chassis or include directives, the `#` must be white-space separated from any preceding GUID, name, or included file name.

Intel recommends that a unique node description be specified for each switch. This name should follow typical naming rules and use the characters a-z, A-Z, 0-9, and underscore. No spaces are allowed in the node description. Additionally, names should not start with a digit.

For externally-managed switches, the node GUID can be found on a label on the bottom of the switch. Alternately the node GUIDs for switches in the fabric can be found using a command such as:

```
opasaquery -t sw -o nodeguid
```

Note: The preceding command will report all switch node GUIDs, including those of internally-managed chassis such as the Intel® Omni-Path Switch 100 Family switches. GUIDs for internally-managed chassis cannot be specified for use in the `switches` file.

FastFabric is topology-aware when updating externally managed switch firmware or resetting the switches. Switches furthest from the FastFabric node are updated or reset first, and then each switch, working toward the FastFabric node. This way switches that are rebooted are not in the path between the FastFabric node and others that are being rebooted.

The ordering is controlled by an optional `distance` field in the `switches` file or the `switches` provided on the command line. The `distance` field indicates the relative distance from the FastFabric node for each switch. Any `switches` file entries that do not specify a distance value are treated as having a value larger than any others in the file. The `switches` file contains any one of the following formats per line:

- `nodeguid`
- `nodeguid,,distance`
- `nodeguid:hfi:port`
- `nodeguid:hfi:port,,distance`
- `nodeguid,nodename`
- `nodeguid,nodename,distance`
- `nodeguid:hfi:port,nodename`
- `nodeguid:hfi:port,nodename,distance`

By default, the `opagenswitches` tool provides the proper distance value relative to the FastFabric node on which it ran. This capability requires the use of IBTA standard TraceRecord queries, which are not supported by openSM but can be supplied by FM. Alternatively, the `opagenswitches -R` option can suppress generation of this field.

In a typical pure fat tree topology with externally managed switches as edge switches and internally managed switches as core switches, you can also manually specify proper distance by simply specifying 1 for the distance value of the switch next to the FastFabric node. Note that in such a topology, all other switches are an equal length from the FastFabric node and a missing hops value will cause them to be treated as having a distance value that is larger than any other found in the file. Therefore, the other switches would be rebooted first and the FastFabric node's switch would be rebooted last.

5.7 Port List File

The `/etc/sysconfig/opa/ports` file is used to specify the local HFI ports (i.e., subnets) that FastFabric will use for assorted commands (such as `opareports`, `opafabricinfo`, `opaswitchadmin`, `opafabricanalysis`, `opaallanalysis`) for fabric access.

Alternate filenames may be specified in `opafastfabric.conf` using environment variables, or on the command line. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

The following is a sample port list file:

```
# [ICS VERSION STRING: @(#) ./fastfabric/opatools/ports x_x_x_x_x
[MM/DD/YY hh:mm]
# This file defines the local HFI ports to use to access the fabric(s)
```

```
#
# specify one line per HFI port of the form hfi:port such as:
#       0:0 = 1st active port in system
#       0:y = port y within system
#       x:0 = 1st active port on HFI x
#       x:y = HFI x, port y
# The first HFI in the system is 1.  The first port on an HFI is 1.
0:0
```

Each line of the port list file may specify a single port, a comment, or another port list file to include.

Ports are specified as `hfi:port`. No spaces are permitted. The first Host Fabric Interface Adapter is 1, and the first port is 1. The special value 0 for Host Fabric Interface or port has special meaning. The allowed formats are shown in the previous sample.

Files to be included may be specified using an `include` directive followed by a file name. File names specified should generally be absolute path names. If relative path names are used, they will be searched for within the current directory then `/etc/sysconfig/opa`.

Comments may be placed on any line by using a `#` to precede the comment. On lines with a port or include directive, the `#` must be white-space separated from any preceding port or included file name.

5.8 Fabric Topology Input File

The `/etc/sysconfig/opa/topology.0:0.xml` file is used to specify the expected fabric topology and augmented fabric information (such as cable labels, types, lengths, SM details, node details, link details, etc.). If present, this file will be used by assorted FastFabric commands (such as `opareports`, `opafabricanalysis`, `opaallanalysis`). Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information on how to create a topology file describing the fabric.

If desired, alternate filenames may be specified in `opafastfabric.conf`, using environment variables or on the command line. Refer to the *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* for more information.

The XML format of topology input can appear as in the following sample:

```
<?xml version="1.0" encoding="utf-8" ?>
<Report date="day mmm  dd hh:mm:ss yyyy" unixtime="1446650124" options="-o
topology" >
<Nodes>
    <FIs>
        <ConnectedFICount>2</ConnectedFICount>
        <Node id="0x00117501007067a2">
            <NodeGUID>0x00117501007067a2</NodeGUID>
            <NodeType>FI</NodeType>
            <NodeType_Int>1</NodeType_Int>
            <NodeDesc>mindy2 HFI-1</NodeDesc>
            <Port id="0x00117501007067a2:1">
                <PortNum>1</PortNum>
                <LID>0x0001</LID>
                <PortGUID>0x00117501007067a2</PortGUID>
                <LinkWidthActive>4</LinkWidthActive>
                <LinkWidthActive_Int>8</LinkWidthActive_Int>
```

```
                <LinkSpeedActive>25Gb</LinkSpeedActive>
                <LinkSpeedActive_Int>2</LinkSpeedActive_Int>
            </Port>
        </Node>
        <Node id="0x00117501007067e6">
            <NodeGUID>0x00117501007067e6</NodeGUID>
            <NodeType>FI</NodeType>
            <NodeType_Int>1</NodeType_Int>
            <NodeDesc>mindy2 HFI-1</NodeDesc>
            <Port id="0x00117501007067e6:1">
                <PortNum>1</PortNum>
                <LID>0x0002</LID>
                <PortGUID>0x00117501007067e6</PortGUID>
                <LinkWidthActive>4</LinkWidthActive>
                <LinkWidthActive_Int>8</LinkWidthActive_Int>
                <LinkSpeedActive>25Gb</LinkSpeedActive>
                <LinkSpeedActive_Int>2</LinkSpeedActive_Int>
            </Port>
        </Node>
    </FIs>
    <Switches>
        <ConnectedSwitchCount>0</ConnectedSwitchCount>
    </Switches>
    <SMs>
        <ConnectedSMCount>1</ConnectedSMCount>
        <SM id="0x00117501007067a2:1">
            <SMState>Master</SMState>
            <SMState_Int>3</SMState_Int>
            <NodeGUID>0x00117501007067a2</NodeGUID>
            <NodeDesc>mindy2 HFI-1</NodeDesc>
            <PortNum>1</PortNum>
            <PortGUID>0x00117501007067a2</PortGUID>
            <NodeType>FI</NodeType>
            <NodeType_Int>1</NodeType_Int>
        </SM>
    </SMs>
</Nodes>
<LinkSummary>
    <LinkCount>1</LinkCount>
    <Link id="0x00117501007067a2:1">
        <Rate>100g</Rate>
        <Rate_Int>16</Rate_Int>
        <Internal>0</Internal>
        <Port id="0x00117501007067a2:1">
            <NodeGUID>0x00117501007067a2</NodeGUID>
            <PortGUID>0x00117501007067a2</PortGUID>
            <PortNum>1</PortNum>
            <NodeType>FI</NodeType>
            <NodeType_Int>1</NodeType_Int>
            <NodeDesc>mindy2 HFI-1</NodeDesc>
        </Port>
        <Port id="0x00117501007067e6:1">
            <NodeGUID>0x00117501007067e6</NodeGUID>
            <PortGUID>0x00117501007067e6</PortGUID>
            <PortNum>1</PortNum>
            <NodeType>FI</NodeType>
            <NodeType_Int>1</NodeType_Int>
            <NodeDesc>mindy3 HFI-1</NodeDesc>
        </Port>
    </Link>
</LinkSummary>
</Report>
```