Cornelis® Omni-Path Express® Fabric

Setup Guide

September 2024
Doc. No. J27600, Rev. 17.0

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Cornelis Networks products described herein. You agree to grant Cornelis Networks a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

The software provided is under license agreements and may contain third-party software under separate third-party licensing. Please refer to the license files provided with the software for specific details.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All product plans and roadmaps are subject to change without notice.

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

Cornelis Networks technologies may require enabled hardware, software, or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Cornelis, Cornelis Networks, Omni-Path, Omni-Path Express, and the Cornelis Networks logo belong to Cornelis Networks, Inc. Other names and brands may be claimed as the property of others.

Copyright © 2022, 2024 Cornelis Networks, Inc. All rights reserved.

Table of Contents

Revision History

Date	Rev	Description
Sep 2024	17.0	Updated the supported processors in Section 2.1 “Configuring BIOS Settings” section.
Nov 2022	16.0	Minor updates throughout
May 2022	15.0	• Updated Section 13.6 “Decoding the Physical Configuration of an HFI”. • Updated Section 13.7 “Programming and Verifying Option ROM EEPROM Device” • Updated complete publication to Cornelis branding and naming conventions.

For previous releases, refer to Appendix B “Older Revisions”.

Preface

This guide is part of the documentation set for the Omni-Path Express Fabric, which is an end-to-end solution consisting of Cornelis Omni-Path Express Host Fabric Interface Adapters (HFIs), Cornelis Omni-Path Express Edge Switches, Cornelis Omni-Path Express Director Class Switches, and fabric management and development tools.

The Cornelis Omni-Path Express Fabric delivers the next generation, High Performance Computing (HPC) network solution that is designed to cost-effectively meet the growth, density, and reliability requirements of large-scale HPC clusters.

Both the Omni-Path Express Fabric and standard InfiniBand (IB) can send Internet Protocol (IP) traffic over the fabric, or *IPoFabric*. In this document it may also be referred to as *IP over IB* or *IPoIB*. From a software point of view, IPoFabric behaves the same way as IPoIB, and in fact uses an ib_ipoib driver to send IP traffic over the ib0/ib1 ports.

Intended Audience

This document is intended for system administrators and other personnel with similar qualifications.

Documentation Library

All Cornelis Networks product documentation may be found in the Release Library located in the Cornelis Customer Center.

Refer to the "Document Library table" in the *Cornelis Omni-Path Express Fabric Quick Start Guide*.

Document Conventions

The following conventions are standardized across all Cornelis Omni-Path Express documentation:

- **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 and underlined indicates a hyperlink to a figure, table, or section in this guide. Links to websites are also shown in blue. For example:

  See License Agreements for more information.

  For more information, visit Cornelis Networks.
- Text in **bold** indicates user interface elements such as menu items, buttons, check boxes, key names, keystrokes, or column headings. For example:

  - Click the **Start** button, point to **Programs**, point to **Accessories**, and then click **Command Prompt**.
  - Press **CTRL+P** and then press the **UP ARROW** key.
- Text in `Courier` font indicates a file name, directory path, or command line text. For example:
  - Enter the following command: `sh ./install.bin`.
- Text in *italics* indicates terms, emphasis, variables, or document titles. For example:
  - Refer to *Cornelis Omni-Path Express Fabric Software Installation Guide* for details.
  - In this document, the term *chassis* refers to a managed switch.
- Most of the acronyms in this document link to the glossary.

Procedures and information may be marked with one of the following qualifiers:

- **(Linux)** – Tasks are only applicable when Linux is being used.
- **(Host)** – Tasks are only applicable when Omni-Path Express Host Software or Omni-Path Express Fabric Suite is being used on the hosts.
- **(Switch)** – Tasks are applicable only when Omni-Path Express Switches or Chassis are being used.
- Tasks that are generally applicable to all environments are not marked.

Cornelis Omni-Path Express Fabric Design Generator for Cornelis Omni-Path Express Fabric

The Fabric Design Generator generates sample cluster configurations based on key cluster attributes, including a side-by-side comparison of up to four cluster configurations. The tool also generates parts lists and cluster diagrams.

To access the Fabric Design Generator for Omni-Path Express Fabric, go to Cornelis® Omni-Path® Fabric Design Generator.

License Agreements

Cornelis software and firmware are provided under one or more license agreements. 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 agreed 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 Cornelis products is available 24 hours a day, 365 days a year:

- Cornelis Networks Customer Support website
- Customer Support email address

1. Introduction

This document provides a high-level overview of the steps required to set up a Cornelis Omni-Path Express Fabric. Procedures and key reference documents, such as Cornelis Omni-Path Express user guides and installation guides are provided to clarify the process. Additional commands and BKMs are defined to facilitate the installation process and troubleshooting.

Cornelis recommends that you use the FastFabric Text-based User Interface (TUI) as the initial tool suite for installation, configuration, and validation of the fabric. This tool includes a set of automated features that are specifically used for standalone host, Ethernet, and Omni-Path Express Fabric connectivity validation.

This document includes recommendations for processes and procedures that complement the FastFabric tools to reduce the time required to install and configure the fabric.

You should check applicable release notes and technical advisories for key information that could influence installation steps outlined in this document.

This document assumes the following:

- Operating System (OS) Software is a release-supported OS. See the *Cornelis Omni-Path Express Fabric Software Release Notes* for the complete list of supported OSes.
- Single Management Node (with Fabric Manager running) configured with the Omni-Path Express Fabric Suite Software, also known as CornelisOPX-OPXS.
- Fabric Manager enabled on management nodes.
- Compute Nodes configured with the Omni-Path Express Fabric Software, also known as CornelisOPX-Basic.
- Password-less access enabled for all hosts and switches.

NOTE

Before you run top500 HPL (High Performance Linpack) runs or customer acceptance tests, Cornelis recommends that you follow all steps outlined in this guide.

1.1. Before You Begin

- Familiarize yourself with the available documentation by referring to Documentation Library in this document.
- Print checklists that contain an overview of the required steps for installing the Omni-Path Express Software are located in the *Cornelis Omni-Path Express Fabric Software Installation Guide*, “Software Installation Checklists” section.

- Go to the Cornelis Customer Center. Select the **Release Library** to download the software you need:
    - Fabric Host Software: Omni-Path Express Fabric Suite package for management nodes
    - Fabric Host Software: Basic package for compute nodes
    - Switch Firmware *.emfw: for externally-managed switches (optional)
    - Switch Firmware *.spkg: for managed switches
    - HFI Platform Firmware: for UEFI, TMM, and Firmware Tools (optional)
    - Fabric Manager GUI software
- Make sure you have access to OS packages and some extended packages that are prerequisites for installing the Omni-Path Express Fabric Suite software.
- Cornelis recommends that you develop a cable map topology file to validate the physical installation of the cables and ensure no accidental hot-spots are created through incorrect cabling. It is also a beneficial extension to any datacenter and rack layout diagrams developed for an installation. Refer to the instructions described in Section 5.1 "Defining Type in the Topology Spreadsheet" in this document.

2. Complete Installation Prerequisites

The recommended fabric installation prerequisites are defined in the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "Pre-Installation Requirements" section.

The RPMs required for the operating system you are using are defined on the website for the particular operating system.

Complete the following steps before starting software installation:

1. Install HFIs in servers, install externally-managed and managed switches in racks, and connect cabling.

   Details: *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* and *Cornelis Omni-Path Express Host Fabric Interface Installation Guide*.

   - All standard EMI protection should be adhered to whenever working with computer equipment. The same is true for switch and HFI installation.
   - HFIs should be installed in PCIe x16 slots of the server. The same PCIe slot should be used on all similar servers if possible to ensure homogeneity in the cluster.
   - Switches should be installed based on switch guidelines and pre-designed rack placement.
   - Care must be taken when connecting the cables. Adhere to bend radius requirements and use cable trays for management. Never use zip-ties to cinch the cables together, use soft fabric Velcro ties instead.

2. Power up all fabric hardware and verify LEDs operate as expected.

   Details: *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* and *Cornelis Omni-Path Express Host Fabric Interface Installation Guide*.

2.1. Configuring BIOS Settings

Cornelis recommends that you use UEFI BIOS. For optimal performance, refer to a recommended BIOS configuration in the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, "BIOS and Platform Settings" sections for the processors supported in the release you are installing. Refer to the "Supported Hardware" section of the Release Notes for supported hardware.

2.2. Configuring OS Settings

Before you install the Omni-Path Express Host Software, perform the following tasks:

- Confirm Operating System (OS) versions match the versions listed in the *Cornelis Omni-Path Express Fabric Software Release Notes*.
- Configure OS settings for optimal performance as described in the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*.

3. Resolve TCP/IP Host Names

For details on resolving TCP/IP Host Names, see the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "Fabric Setup Prerequisites" section.

Create a `/etc/hosts` file before starting the Omni-Path Express Host Software installation to simplify the process. In a typical installation, the server and switch names follow a local convention to indicate the physical location or purpose of the node.

- If using `/etc/hosts`, update the `/etc/hosts` file on the management node (the node with Omni-Path Express Fabric Suite installed) and copy it to all hosts.
- If using DNS, all Management Network and IPoIB hostnames must be added to DNS `/etc/resolve.conf` and configured on the management node.
- The `/etc/hosts` file should contain:
  - Local host, for subsequent single host verification using FastFabric TUI
  - Ethernet and IPoIB addresses and names for all hosts
  - Ethernet addresses and names of switches
  - Ethernet addresses of IPMI or remote management modules
  - Ethernet addresses of power domain

  An example of these recommendations follows:

```
# /etc/hosts example
# localhost (required)
127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain
# Ethernet Addresses of hosts
10.128.196.14    node1
10.128.196.15    node2
10.128.196.16    node3
#IPoIB Address of hosts should be outside Ethernet network
10.128.200.14    node1-opa
10.128.200.15    node2-opa
10.128.200.16    node3-opa
#RMM IP Addresses
10.127.240.121    node1-rmm
10.127.240.122    node2-rmm
# Chassis IP Address
10.128.198.250    opaedge1
10.128.198.249    opaedge2
# OPA director switch IP Address
10.128.198.251 opadirector1
10.128.198.252 opadirector2
```

Other files that may need adjustment according to specific site requirements include:

- `/etc/hostname`
- `/etc/resolv.conf`
- `/etc/network`
- `/etc/network-scripts/ifcfg-*` (for example, `ifcfg-enp5s0f0`)

4. Install Omni-Path Express Fabric Software

You should configure at least one node to run the Management Software including Fabric Manager (FM). This node is used to configure and validate all of the other hosts, switches, and chassis fabric devices. You must install the Omni-Path Express Fabric Suite software on this node.

The following document and sections describe the installation procedures:

- *Cornelis Omni-Path Express Fabric Software Installation Guide*
  - "Download the Omni-Path Express Fabric Software"
  - "Unpack the Tar File"
  - "Install the Omni-Path Express Fabric Software"

4.1. Installing Drivers

Driver installation is a two-step process. First, install the Omni-Path Express Fabric Suite package on your management servers, then use FastFabric to install the drivers on the rest of the systems in the cluster in parallel.

NOTE

If you use an existing cluster manager, installing the drivers to the image being installed on the compute nodes instead of using FastFabric is another option for quick driver installation.

4.1.1. Procedure

The following steps provide a summary for installation:

1. At the command prompt, change the directory to:

   `/CornelisOPX-OPXS.DISTRO.VERSION`.

2. Use the `./INSTALL` command to install the Omni-Path Express Fabric Suite software package on the management node(s) usually designated to run Subnet Manager (SM) and FastFabric Tools (including MPI applications).

NOTE

Cornelis recommends that you use the CLI commands for installation, configuration, and validation of the fabric. See "Install Using CLI Commands" in the *Cornelis Omni-Path Express Fabric Software Installation Guide*.

Examples include: `./INSTALL -n` or `./INSTALL -a -G`

Alternatively, FastFabric Text-based User Interface (TUI) as the tool suite can be used. See "Install Using the TUI Menus" in the *Cornelis Omni-Path Express Fabric Software Installation Guide*.

Install using the Omni-Path Express Fabric Suite-included Linux OS repository. See "Install Using Linux Distribution Software Packages Provided by Cornelis" in the *Cornelis Omni-Path Express Fabric Software Installation Guide*.

3. (optional) Upgrade the HFI UEFI firmware and install the Firmware Tools. Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "Install and Upgrade Standalone Firmware" section.
4. Cornelis recommends that you enable servers with IPMI interfaces to support ACPI or equivalent remote power management and reset control via an Ethernet network.
5. Apply Technical Advisories as needed.
6. (optional) Install NVIDIA software.

   For installation details, see the *Cornelis Omni-Path Express Fabric Software Installation Guide*.

NOTE

GPUDirect RDMA is an NVIDIA technology that allows third-party network adapters to directly read and write to CUDA host and device memory to enhance performance for latency and bandwidth.

For usage details, see the *Cornelis Omni-Path Express Fabric Host Software User Guide* and the *Cornelis Performance Scaled Messaging 2 (PSM2) Programmer's Guide*.

4.2. Verifying HFI Speed and Bus Width Using lspci

After the Omni-Path Express Fabric Suite installation, verify the Omni-Path Express HFI is configured and visible to the host OS as Gen3 x16 slot speed (values are in **bold** text):

```
lspci -d 8086:24f0 -vv |grep Width
LnkCap: Port #0, Speed 8GT/s, Width x16, ASPM L1, Exit Latency L0s
<4us, L1 <64us
LnkSta: Speed 8GT/s, Width x16, TrErr- Train- SlotClk+ DLActive- BWMgmt- ABWMgmt-
```

4.3. Performing Initial Fabric Verification

4.3.1. Procedure

Perform the following steps:

1. Run `opainfo` to verify the port state of host is Active.

   NOTE

   If the command fails or returns a port state other than Active, verify that the SM is running using `systemctl status opafm`.

2. Use `opacmdall` or `pdsh` to run `opainfo` on all nodes in the fabric.
3. Run `opaconfig -V` to verify the software version is the same on all nodes.
4. Run `opafabricinfo` as shown in the following example to verify all nodes, switches, SM, and ISLs are up.

```
opafabricinfo
Fabric 0:0 Information:
SM: node1 hfi1_0 Guid: 0x001175010165b116 State: Master
Number of HFIs: 126
Number of Switches: 9
Number of Links: 252
Number of HFI Links: 126            (Internal: 0   External: 126)
Number of ISLs: 126                 (Internal: 0   External: 126)
Number of Degraded Links: 0         (HFI Links: 0   ISLs: 0)
Number of Omitted Links: 0          (HFI Links: 0   ISLs: 0)
```

5. Review the number of HFIs, number of switches, and external ISLs, and confirm that they match the fabric design.

   NOTE

   The number of HFIs and external ISLs provide a fabric-blocking factor. If there are any degraded links, further troubleshooting is required.

4.4. Editing Hosts and Allhosts Files

Edit the following files, which are used by the `opafastfabric.conf` file.

- Edit `/etc/opa/hosts`

  This file contains all hosts except the management node running Omni-Path Express Fabric Suite.

- Edit `/etc/opa/allhosts`

This file contains the statement `include /etc/opa/hosts`. Edit the file to add the node(s) running Omni-Path Express Fabric Suite.

5. Generate Cable Map Topology Files

Two sample topology files, `detailed_topology.xlsx` and `minimal_topology.xlsx`, are provided in the release.

For complete details on the sample files, see the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*.

- Both sample topology files, `detailed_topology.xlsx` and `minimal_topology.xlsx`, are spreadsheets with 3 tabs:
  - Tab 1 Fabric is for you to customize with EXTERNAL links.
  - Tab 2 swd06 contains the internal links for an Omni-Path Express Edge Switch.
  - Tab 3 swd24 contains the internal links for an Omni-Path Express Director Class Switch.

  NOTE

  You should **not** modify tab 2 and 3.

- `README.topology` and `README.xlat_topology` describe best practices for editing the sample topology files.

For descriptions of other sample files provided in the package, see the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*.

5.1. Defining Type in the Topology Spreadsheet

All host nodes should be defined as Type = FI in column F of the spreadsheet. All Edge switches should be defined as Type = SW in column L (destination from host to Edge) and column F (source for Edge to core that is also Edge switch). The following example shows links between host and Edge switch.

```
R19 opahost1  1 FI  R19 opaedge1  13 SW opahost1_opae1p13 1m Cable CU
```

All links between Edge switch to core that is also an Edge switch should be defined as Type = SW, as shown in the following example:

```
row1 rack01 opaedge1  1 SW row1 rack04 opaedgecore1 2 SW opae1p1_opac1p2 5M Cable Fiber
```

All Director switches should be defined as Type = CL in column L (destination from Edge switch to Director switch). Column J (Name-2) should have the destination leaf and column K should have the port number on that leaf. The following example shows a link between an Edge switch to core that is a Director switch.

```
R19 opaedge1  5 SW  R72 opadirector1 01 L105B 11 CL opae1p5opad1L105Bp11 30m Fiber
```

All 24-leaf chassis Director switches should be defined as shown in the following example:

```
Core Name:opadirector1 Core Group:row1 Core Rack:rack72 Core Size:1152 Core Full:0
```

Set Core Full to 0 if the Director switch is not fully populated with all the leafs and spines. If it is fully populated, set Core Full to 1.

For complete details, see the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*.

5.2. Creating the Topology File

To create a topology file, perform the following steps:

1. Copy and save either `detailed_topology.xlsx` or `minimal_topology.xlsx` located in `/usr/share/opa/samples` from the Fabric Manager node to your local PC for editing.
2. Edit tab 1 in the spreadsheet to reflect your specific installation details. Save tab 1 as `<`*`topologyfile`*`>.csv` and copy this CSV file back to the Fabric Manager node.

   NOTE

   The cable label field can be up to 57 characters.

3. Generate the topology file in XML format using the following command and your customized CSV file as the source:

   ```
   # opaxlattopology <topologyfile>.csv <topologyfile>.xml
   ```

If there are Director switches defined in the CSV file, then `opaxlattopology` includes all the ISL (internal chassis links between leafs and spines) in the XML file.

For more details, see the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*, `opaxlattopology` section.

6. Configure FastFabric

The list of configuration files that are used by FastFabric are contained in the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*, Configuration Files for FastFabric section.

The `opafastfabric.conf` file provides default settings for most of the FastFabric command line options.

6.1. Formatting for IPoIB Host Names

By default, FastFabric uses the suffix `opa` for the IPoIB host name. You can change this to a prefix and you can also change from `opa` to another convention such as `ib`, as the customer requires in `/etc/opa/opafastfabric.conf`.

The following examples show how to change `opa` to `ib` as a prefix or suffix.

For suffix:

```
export FF_IPOIB_SUFFIX=${FF_IPOIB_SUFFIX:--opa to export FF_IPOIB_SUFFIX=$
{FF_IPOIB_SUFFIX:--ib
```

For prefix:

```
export FF_IPOIB_PREFIX=${FF_IPOIB_PREFIX:-opa- to export FF_IPOIB_PREFIX=$
{FF_IPOIB_PREFIX:-ib-
```

6.2. Specifying Test Areas for opaallanalysis

By default, `opaallanalysis` includes the fabric and chassis. These can be modified to include host SM, embedded SM, and externally-managed switches in `/etc/opa/opafastfabric.conf` as follows:

```
# pick appropriate type of SM to analyze
#export FF_ALL_ANALYSIS=${FF_ALL_ANALYSIS:-fabric chassis hostsm esm}
export FF_ALL_ANALYSIS=${FF_ALL_ANALYSIS:-fabric chassis hostsm}
```

6.3. Modifying the Location of mpi_apps Directory

By default, `opafastfabric` uses `mpi_apps` located in `/usr/src/opa/mpi_apps`. If a different path is set up for `mpi_apps`, then modify the following in `/etc/opa/opafastfabric.conf`:

```
export FF_MPI_APPS_DIR=${FF_MPI_APPS_DIR:-/usr/src/opa/mpi_apps}
```

NOTE

The default source location for `mpi_apps` is `/usr/src/opa/mpi_apps`, but the default compilation directory for `mpi_apps` is `$HOME/mpi_apps`.

7. Configure Managed Omni-Path Express Edge Switches

For a complete description of the configuration process, refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "Configure the Chassis" section.

For more information about the managed switch, refer to the *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* and *Cornelis Omni-Path Express Fabric Software Release Notes*.

7.1. Procedure

The following steps provide a summary for configuring the Omni-Path Express Edge Switch:

1. Install the driver file, `CDM v2.12.00 WHQL Certified.exe`, which can be downloaded from: http://www.ftdichip.com/Drivers/VCP.htm.
2. Set up the USB serial port terminal emulator using the following serial options:
   - Speed: **115200**
   - Data Bits: **8**
   - Stop Bits: **1**
   - Parity: **None**
   - Flow Control: **None**
3. Set up the switch TCP/IP address, gateway, netmask, and other options using a terminal emulator.

   NOTE

   The changes are effective immediately.

   For details, refer to the *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide*.

   a. Use the command `setChassisIpAddr -h` *`ipaddress`* `-m` *`netMask`*, where *`ipaddress`* is the new IP address in dotted decimal format (xxx.xxx.xxx.xxx), and *`netMask`* is the new subnet mask in dotted decimal format, to set the chassis IP address.

   b. Use the command `setDefaultRoute -h` *`ipaddress`*, where *`ipaddress`* is the new default gateway IP address in dotted decimal format, to change the chassis default gateway IP address.
4. Use the command `opagenchassis >> /etc/opa/chassis` to edit the chassis file.

> **NOTE**
>
> The chassis file contains the node name of managed switches corresponding to TCP/IP addresses as defined in the `/etc/hosts` file.

5. Type `opafastfabric` and press **Enter** to run the `opafastfabric` TUI.
6. Press **1** to access the **FastFabric OPA Chassis Setup/Admin menu**.
7. Select menu items `0-6` and press **P** to perform the selected operations.
8. (menu item 0) **Edit the Configuration and Select/Edit Chassis File**
   - Skip `opafastfabric.conf`. No changes needed.
   - Skip `ports`. No changes needed.
   - Open the `chassis` file in the editor.

   > **NOTE**
   >
   > Review the list of chassis selected. The setup of this file should have occurred above when setting up the Management Node by editing `/etc/opa/chassis` with the name corresponding to the Ethernet IP address of the chassis.

9. (menu item 1) **Verify Chassis via Ethernet Ping**. The test should pass without error.
10. (menu item 2) **Update Chassis Firmware**. Be sure to specify the location for the firmware file to use.
11. (menu item 3) **Set Up Chassis Basic Configuration**. When prompted for answers, provide the following:
    a. Password: Press **Enter** (no password).
    b. Syslog: y
       i. Syslog server: n
       ii. TCP/UDP port number: n - use default
       iii. Syslog facility: n - use default
    c. NTP: n - customer to assign.
    d. Timezone and DST: y - use local timezone of server.
    e. Do you wish to configure OPA Node Desc to match Ethernet chassis name? y - Enter y.
    f. Do you wish to configure the Link CRC Mode? n
12. (menu item 4) **Set Up Password-less ssh/scp**. Set up as required or press **Enter** to accept the default Chassis password.

13. (menu item 5) **Reboot the Chassis**. Reboot should pass without error
14. (menu item 6) **Get Basic Chassis Configuration**. Expected Summary output at end is shown below. Note that count should match the number of Edge switches.

```
Edgeswitch1:
Firmware Active          : 10.x.x.x.x
Firmware Primary         : 10.x.x.x.x
Syslog Configuration     : Syslog host set to: 0.0.0.0 port 514 facility 22
NTP                      : Configured to use the local clock
Time Zone                : Current time zone offset is: -5
LinkWidth Support        : 4X
Node Description         : switch1
Link CRC Mode            : 48b_or_14b_or_16b
```

15. Use an editor to review the following files:
    - `/root/test.res`
    - `/root/test.log`

8. Configure Cornelis Omni-Path Express Director Class Switch

Most Cornelis Omni-Path Express Director Class Switches (DCSs) are supplied with two Management Modules (MMs) for redundancy. In addition, DCSs have the following additional features:

- The switch has two Ethernet ports (one for each MM) and requires two Ethernet cables.
- The switch requires three IP addresses: one for each MM and one for the chassis, which is bound to the MM that is currently Master.
- It is useful to understand all reboot modes: `reboot all|-s|-m [slot #]` and how that causes failover.
- Default IP addresses of the Management Modules are:

  Chassis IP address: `192.168.100.9`

  Management Module M201: `192.168.100.10`

  Management Module M202: `192.168.100.11`

The chassis file, located in `/etc/opa/chassis`, contains the node names of the DCSs corresponding to TCP/IP addresses as defined in the `/etc/hosts` file. The chassis IP address is configured using the procedure for configuring managed switches, as described in Section 7 “Configure Managed Omni-Path Express Edge Switches”.

Configure the MM IP addresses using a serial connection as described in the following procedure:

1. Ensure that the module is connected to a COM port on a serial terminal device through the USB port.
2. Get to a `[boot]:` prompt by following either step a or b:
   a. If the management module is running and displays -> prompt, type the following command at the console: `reboot now` and press **ENTER**.
   b. If the management module is not running, power on the switch.
3. When the system displays `image1`, press the spacebar to interrupt the autoload sequence before the counter expires (within 5 seconds).
4. At the prompt, enter the command: `moduleip <`*`ip_address`*`>`

   The module reboots itself within five seconds and comes back with the new IP assigned to it. This module becomes the standby and the other MM becomes the master.

Repeat these steps for the second management module.

For more information, refer to the *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* and the *Cornelis Omni-Path Express Fabric Switches Release Notes*.

9. Configure Externally-Managed Omni-Path Express Edge Switches

For a complete description of the install process, refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "Configure the Externally-Managed Switches" section.

For more information about the managed switch, refer to the *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* and *Cornelis Omni-Path Express Fabric Switches Release Notes*.

The 100SWE48QF Edge switches do not have an Ethernet interface. Setup of these switches is performed using FastFabric via in-band commands.

9.1. Procedure

The following steps describe the preferred approach for configuring the externally-managed Edge Switch:

1. Use the command `opagenswitches >> /etc/opa/switches` to edit the switches file to replace the default switch name with the actual name that corresponds to the GUID for each switch for externally-managed switches. For example:

   Default: `0x00117501026a5683:0:0,`**`OmniPth00117501ff6a5602`**`,2`

   Edited: `0x00117501026a5683:0:0,`**`opaextmanagededge1`**`,2`

2. Type `opafastfabric` and press **Enter** to run the `opafastfabric` TUI.
3. Press **2** to access the **Externally Managed Switch Setup/Admin menu**.
4. Select menu items `0–9` and press **P** to perform the operations.
5. (menu item 0) **Edit the Configuration and Select/Edit Switch File**.
   a. Skip `opafastfabric.conf`. No changes needed.
   b. Skip `ports`. No changes needed.
   c. Open the `switches` file and review the list of switches selected. The `switches` file specifies:
      - switches by node GUID
      - (optional) hfi:port
      - (optional) Node Description (nodename) to be assigned to the switch
      - (optional) distance value indicating the relative distance from the FastFabric node for each switch

      The following snippet shows the switches file format and an example:

```
nodeguid:hfi:port,nodename,distance
0x00117501026a5683:0:0,opaextmanagededge1,2
```

6. (menu item 1) **Generate or Update Switch File**.
   a. To generate the switch file, type n if the switch file was generated in step 1. Type y if this is the first time, or additional externally-managed switches have been added or replaced.
   b. Type y to update switch names. Note that this step may take a few minutes.
7. (menu item 2) **Test for Switch Presence**. Test should pass without error.
8. (menu item 3) **Verify Switch Firmware**. Test should pass without error.
9. (menu item 4) **Update Switch Firmware**. Be sure to specify the location for the FW file (`.emfw`) to use.
10. (menu item 5) **Set Up Switch Basic Configuration**. Set up switch basic configuration and set the node description as shown below:

```
Performing Switch Admin: Setup Switch basic configuration
Executing: /usr/sbin/opaswitchadmin -L
/etc/opa/switches configure
Do you wish to configure the switch Link Width Options? [n]:
Do you wish to configure the switch Node Description as it is set in the switches
file? [n]: y
Do you wish to configure the switch FM Enabled option? [n]: Do you wish to configure
the switch Link CRC Mode? [n]: Executing configure Test Suite (configure) Fri Jan 15
11:11:12 EST 2016 ...
Executing TEST SUITE configure CASE
(configure.0x00117501026a5683:0:0,OmniPth00117501ff6a5602.i2c
.extmgd.switchconfigure) configure switch
0x00117501026a5683:0:0,OmniPth00117501ff6a5602 ...
TEST SUITE configure CASE
(configure.0x00117501026a5683:0:0,OmniPth00117501ff6a5602.i2c
.extmgd.switchconfigure) configure switch
0x00117501026a5683:0:0,OmniPth00117501ff6a5602 PASSED
TEST SUITE configure: 1 Cases; 1 PASSED
```

11. (menu item 6) **Reboot Switch**. Reboot should occur without error.
12. (menu item 7) **Report Switch Firmware & Hardware Info**. Review results for redundant power and FAN status. Expected summary output at end should be similar to the following (count should match number of externally-managed Edge switches):

```
0x00117501026a5683:0:0,opaextmanagededge1:
F/W ver:10.x.x.x.x    H/W ver:003-01    H/W pt num:H89344-003-
01    Fan status:Normal/Normal/Normal/Normal/Normal/Normal    PS1
Status:ONLINE    PS2 Status: ONLINE    Temperature
status:LTC2974:33C/MAX_QSFP:40C/PRR_ASIC:40C
```

    Any non-redundant or failed fans or power supplies found during this step are also reported in `/root/punchlist.csv`.

13. (menu item 8) **Get Basic Switch configuration**. Expected summary output at end should be similar to the following (count should match number of externally-managed Edge switches):

```
Link Width                     :    1,2,3,4
Link Speed                     :    25Gb
FM Enabled                     :    No
Link CRC Mode                  :    None
vCU                            :    0
External Loopback Allowed      :    Yes
Node Description               :    Edgeswitch1
```

14. (menu item 9) **Report Switch VPD Information**. Save the `test.res` output for future reference.
15. Use an editor to view the results in the following files:
    - `/root/test.res`
    - `/root/test.log`

10. Configure Host Setup

For more information about the installation process, refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide* and *Cornelis Omni-Path Express Fabric Software Release Notes*.

10.1. Procedure

The following steps provide a summary for configuring the Hosts:

1. Identify switch names. Make sure all hosts are booted. This is required to identify switch names. If hosts are not available, you can perform all configuration steps except setting the switch names.
2. Type `opafastfabric` and press **Enter** to run the `opafastfabric` TUI.
3. Press **3** to access the **FastFabric OPA Host Setup menu**.
4. Select items `0-8` and press **P** to perform the operations.

   NOTE

   Perform the following operations only as needed:

   a. (menu item 0) **Edit Configuration and Select/Edit Host File**
   b. (menu item 1) **Verify Hosts Pingable**
   c. (menu item 2) **Set Up Password-Less SSH/SCP**
   d. (menu item 3) **Copy /etc/hosts to All Hosts**
   e. (menu item 4) **Show uname -a for All Hosts**

5. (menu item 5) **Install/Upgrade OPA Software** will install CornelisOPX-Basic software package on all compute nodes defined in `/etc/opa/hosts`. Be sure to exclude management node(s) with Omni-Path Express Fabric Suite installed and the node where you are running `opafastfabric`.
   a. Provide the path to `CornelisOPX-Basic.DISTRO.VERSION.tgz` when prompted.
   b. Enter directory to get `CornelisOPX-Basic.DISTRO.VERSION.tgz` from (or none) `:/root`.
6. (menu item 6) **Configure IPoIB IP Address** to perform the IPoIB ping test.
7. (menu item 7) **Build Test Apps and Copy to Hosts**.
   a. Choose an MPI when prompted with the following: `Please Select MPI Directory`
   b. Select an MPI with -hfi extension, so it will build with PSM2 or the OPX provider. For example: `/usr/mpi/gcc/openmpi-x.x.x-hfi`.

c. When prompted to build base sample applications, select yes.

8. (menu item 8) **Reboot Hosts**.

11. Verify Cable Map Topology

This section describes how to use the fabric `topology.xml` file created in Section 5.1 "Defining Type in the Topology Spreadsheet" to verify that fabric topology (cabling) is consistent with the cable map.

The command `opareport -o verify* -T <topologyfilename>.xml` compares the live fabric interconnect against the topology file created based on the cable map. These commands test links, switches, and SM topology. If successful, the output reports a total of 0 Incorrect Links found, 0 Missing, 0 Unexpected, 0 Misconnected, 0 Duplicate, and 0 Different.

```
# opareport -o verifyfis -T <topologyfilename>.xml
# opareport -o verifyextlinks -T <topologyfilename>.xml
# opareport -o verifyall -T <topologyfilename>.xml
```

In most cases, links reported with errors are either due to incorrect cabling to the wrong port or the `topology.csv` file has incorrect source and port destinations.

Verify the physical interconnect against the cable map using `opaextractsellinks` as shown in the following examples:

- List all the links in the fabric: `opaextractsellinks`
- List all the links to a switch named `OmniPth00117501ffffffff`: `opaextractsellinks -F "node:OmniPth00117501ffffffff"`
- List all the connections to end-nodes: `opaextractsellinks -F "nodetype:FI"`
- List all the links on the second HFI's fabric of a multi-plane fabric: `opaextractsellinks -h 2`

After all topology issues have been resolved, copy the `topologyfile.xml` from the local working directory to `cat /etc/opa/topology.0\:0.xml`.

Refer to the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide* for more information about using `opareport` in general, and using `opareport` for Advanced Topology Verification.

12. Verify Server and Fabric

Validation of servers and the fabric is initiated from the Management Node using the FastFabric TUI using `opafastfabric` on hosts defined in `/etc/opa/allhosts`.

12.1. Procedure

Perform the following steps to verify servers and fabric:

1. Type `opafastfabric` and press **Enter** to run the `opafastfabric` TUI.
2. Press **4** to access the **Host Verification/Admin** menu.
3. Select items `3`, `4`, `6`, and `8` and press **P** to perform the operations.
4. (menu item 3) **Perform Single Host Verification**.
    a. Enter **y** when prompted with `Would you like to specify tests to run? [n]:`
    b. Enter **y** when prompted with `View Load on hosts prior to verification? [y]:`

       This option checks CPU load by running `/usr/sbin/opacheckload -f /etc/opa/allhosts`.
5. (menu item 4) **Verify OPA Fabric Status and Topology**. Accept the default for all prompts.

    > **NOTE**
    >
    > This option goes through a fabric error and topology verification. Edit `/root/linkanalysis.res` to view results.

6. (menu item 6) **Verify Hosts Ping via IPoIB**. This option pings all IPoIB interfaces.
7. (menu item 8) **Check MPI Performance**. Accept the default for all prompts.

    > **NOTE**
    >
    > This option tests Latency and Bandwidth deviation between all hosts. Edit `/root/test.log` to view results.

For more information, refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*.

NOTE

A punchlist file is generated during execution of the FastFabric TUI and CLI commands, which can be used to track issues identified by the Omni-Path Express tools. The punch list file is located in `$FF_RESULT_DIR /punchlist.csv`, typically `/root/punchlist.csv`.

Two additional files, `/root/test.res` and `/root/test.log`, are created during Omni-Path Express test commands and are useful for tracking test failures and issues.

13. Apply Best Known Methods for Site Installation

This section contains commands useful for configuring and debugging issues during fabric installation.

13.1. Enabling the Fabric Manager GUI

By default, the Fabric Manager GUI is disabled after installation of the Omni-Path Express Fabric Suite software. To quickly enable for early debug, use the following steps. For complete details, refer to the *Cornelis Omni-Path Express Fabric Suite Fabric Manager GUI User Guide*.

> **NOTE**
>
> This method bypasses the SSH key authorization and is not intended for end-customer installs.

13.1.1. Procedure

Perform the following steps to enable the Fabric Manager GUI:

1. Open the `/etc/opa-fm/opafm.xml` file on the Management Node to make the two changes shown in **bold** for SslSecurityEnabled and default FE startup:

```
<SslSecurityEnabled>0</SslSecurityEnabled>

<!-- Common FE (Fabric Executive) attributes -->
<Fe>
<!-- The FE is required by the Omni-Path FM GUI. -->
<!-- To enable the FE, configure the SslSecurity parameters in this file -->
<!-- as desired. -->
<!-- For Host FM then set Start to 1. -->
<!-- For Embedded FM the Start parameter in this file is not used;    -->
<!-- enable the FE via the smConfig and smPmStart chassis CLI commands. -->
<Start>1</Start> <!-- default FE startup for all instances -->
<!-- Overrides of the Common.Shared parameters if desired -->
<!-- <SyslogFacility>Local6</SyslogFacility> -->
```

2. Type: `systemctl restart opafm` to restart the Fabric Manager to enable the changes and start the FE process required by the Fabric Manager GUI.
3. Download and install the Fabric Manager GUI application to a Windows PC or Linux system. Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, Install Cornelis Omni-Path Express Fabric Suite Fabric Manager GUI section.
4. Start the Fabric Manager GUI application.
5. Open the **Configuration** tab and enter the hostname or IP address of the Management Node running the Fabric Manager in your system into the FE Connection.

6. Uncheck the Secure tab.
7. Click **Apply** to run the connection test.
8. Click **Run** to start the Fabric Manager GUI application.

> **NOTE**
>
> The Fabric Manager GUI does not operate through network proxies. Network firewall access may also need to be disabled. For a quick go/no-go verification, complete the connection test in the configuration tab as previously described.

13.2. Review Server and Fabric Verification Test Results

During fabric validation, unexpected loads on Host CPUs may result in inconsistent performance results. As a debug step, isolate the issue using the following:

Use the Omni-Path Express tool to verify CPU host load. By default, it captures the top ten most heavily loaded hosts.

```
# /usr/sbin/opacheckload -f /etc/opa/allhosts
```

After the high load hosts have been identified, the next step is to root cause the issues.

Procedure

Perform the following steps to verify servers and fabric:

1. Use `lspci` (Section 4.2 “Verifying HFI Speed and Bus Width Using lspci”) or `opahfirev` (Section 13.6 “Decoding the Physical Configuration of an HFI”) to verify the HFI PCIe cards' operating speed and bus width. >Possible sources for narrow PCIe width are:
   - Be aware that Omni-Path Express does support different width PCIe cards, including dual HFI cards using two x8 slices of a x16 physical connector. `opahfirev` is very useful for detecting this configuration.
   - HFI Card partial insertion into x16 slots. Initially this appears to be a narrow width issue but re-inserting the card often resolves the issue. This may occur after a server is shipped. This step has resolved most width issues.
   - Server physical configuration: Many servers support different PCIe logical widths based on riser card configuration. The slot may be physically x16 but internally limited to x8. Check other servers of the same configuration in the fabric. Check the server configuration. *This is also a common issue*.
   - Swap the HFI to another server to determine if the problem follows the card or the server.
2. Use the Linux `top` command to identify the key CPU load processes.

> **NOTE**
>
> `opatop` may be useful for checking for loads that vary over time. Use the `r` (rev), `f` (forward), and `L` (live) options to look through PM snapshots of system activity. This is also helpful for monitoring application startup versus run time loads. The PM captures high resolution statistics, with very low system overhead, over periods up to two days. The tools that harvest the PM stats are `opatop` and the FM GUI.

3. Check for high CPU percent processes. For example:
   - **Screen savers** - when a Linux GUI is enabled on hosts, the screen that runs when the user interface is idle may have a high CPU load.
   - **Test applications** - look for MPI jobs or similar applications running in the background. This is a common issue particularly in a shared fabric bring-up environment. Use `kill -p` *`process`* to stop orphan applications or reboot the server to debug the issue.
4. Review BIOS settings to isolate nodes with different or incorrect settings. Refer to Section 2.1 “Configuring BIOS Settings”.

13.3. Debugging Omni-Path Express Physical Link Issues

After you have run the FastFabric tool suite and identified issues with links, it is useful to start root-causing the issues. This section focuses on the Omni-Path Express Fabric physical links and not PCIe bus link issues.

The Omni-Path Express reporting tools are robust, but it can be confusing for new users to understand the difference between error counters and actual failures.

From an installation perspective, it is important to watch for physical issues with cabling, both copper and optical. In general, bend radius, cable insertion issues, and physical compression or damage to cables can result in transmission issues. The software recovers from many issues transparently.

The following information can help you root-cause solid failures as well as marginal links. Most often the issue is resolved simply by re-installing a cable and verifying that it clicks into the connector socket on the HFI or switch.

- View the QSFP/cable details of a specific switch port using the command:

```
opasmaquery -o cableinfo -d 10 -l <lid> -m <switch portnumber>
```

- To debug a particular switch, a useful technique is to get a snapshot of it using the command:

```
opareport -o snapshot -F portguid:value
```

13.3.1. Omni-Path Express Link Transition Flow

To debug link issues, it is helpful to understand the four key link states, starting from Offline and running properly in the final Active state.

NOTE

The Fabric Manager, `opafm`, must be running to transition physical links from the Init state to the Active state. If you subsequently stop the Fabric Manager when a link is in the Active state, the link remains active. You can safely make changes to the `opafm.xml` file for the Fabric Manager and restart the service without dropping active links. As of the 10.0.0.696 software release, by default, the `opafm` service is not configured for autostart after Omni-Path Express Fabric Suite installation.

PortState:

- Offline: link down. QSFP not present or not visible to the HFI driver.
- Polling: physical link training in progress. At this point you do not know whether the other end of the QSFP is connected to a working Omni-Path Express device.
- Init: Link training has completed, both sides are present. Typically waiting for the Fabric Manager to enable the link.
- Active: Normal operating state of a fully functional link.

13.3.2. Verify the Fabric Manager is Running

From the Management Node, run the following command to report all HFIs and Switches.

```
# opafabricinfo
```

If it fails, try the following steps:

- Check the status of the Fabric Manager process using the command:

  ```
  # systemctl status opafm
  ```

- Restart the Fabric Manager using the command:

  ```
  # systemctl start opafm
  ```

13.3.3. Check the State of All Links in the System

The `opaextractsellinks` command generates a CSV output representing the entire link state of the fabric.

- To generate the CSV output, run the following:

```
# opaextractsellinks > link_status.csv
```

- For links with errors, run the `opaextracterror` command:

```
# opaextracterror > link_status.csv
```

13.3.4. Check the State of HFI Links from a Server

If you are debugging server link issues, the `opainfo` command may be useful for a single server view.

`opainfo` captures a variety of data useful for debugging server-related link issues. Multiple Omni-Path Express commands can be used to extract individual data elements, however, this command is unique in the combination of data it provides.

- PortState: see Section 13.3.1 “Omni-Path Express Link Transition Flow”.
- LinkWidth: a fully functional link should indicate Act:4 and En:4.
- QSFP: Physical cable information for the QSFP, in this case a 5M Optical (AOC) Finisar cable.
- Link Quality: Range = 0 - 5 where 5 is Excellent.

```
# opainfo
hfi1_0:1                 PortGID:0xfe80000000000000:001175010165b19c
   PortState:     Active
   LinkSpeed      Act: 25Gb         En: 25Gb
   LinkWidth      Act: 4            En: 4
   LinkWidthDnGrd ActTx: 4  Rx: 4   En: 3,4
   LCRC           Act: 14-bit       En: 14-bit,16-bit,48-bit Mgmt: True
   LID: 0x00000001-0x00000001       SM LID: 0x00000002 SL: 0
   QSFP: PassiveCu, 1m   FCI Electronics   P/N 10131941-2010LF  Rev 5
   Xmit Data:           22581581 MB Pkts:           5100825193
   Recv Data:           18725619 MB Pkts:           4024569756
   Link Quality: 5 (Excellent)
```

13.3.5. Link Width, Downgrades, and opafm.xml

By default, Omni-Path Express links run in x4 link width mode. Omni-Path Express has a highly robust link mechanism, as compared to InfiniBand, and it allows links to run in reduced widths with no data loss.

Three things to know:

1. By default, the `opafm.xml` configuration file requires links to start up in x4 link width mode. This is configurable separately for HFI and ISL links using the **WidthPolicy** parameter.
2. Link downgrade ranges are also configurable in the `opafm.xml` file, using the **MaxDroppedLanes** parameter.

3. Default configuration example - A link that successfully starts up in x4 width and subsequently downgrades to x3 width continues to operate. If the link is restarted, by a server reboot, for example, and attempts to run by less than x4 width, then the link is disabled by the Fabric Manager and does not enter the Active state.

The `opainfo` command for HFIs is useful for checking the link width and link downgrade configuration on servers.

For a system view of all links that are running in less than x4 width mode, use the command:

```
# opareport -o errors -o slowlinks
```

13.3.6. How to Check Fabric Connectivity

For large fabrics, follow the flow described in Section 5.1 “Defining Type in the Topology Spreadsheet”.

13.3.7. Physical Links Stability Test Using opacabletest

Omni-Path Express uses a quality metric for reporting status (`opainfo`). The quality metric ranges from 5 (excellent) to 1 (poor). For a more quantitative metric, use `cabletest` to generate traffic from on the HFI and ISL links, and `opaextractperf` and `opaextracterrors` to harvest the data.

13.3.7.1. Procedure

Perform the following steps:

1. Prior to running the cable tests, perform the following:

   > **NOTE**
   >
   > Use `opatop` to monitor fabric and link utilization.

   a. Use `/usr/sbin/opacabletest -A -n 3 -f '/etc/opa/allhosts' stop_fi stop_isl # opareport -o none –clearall` to clear error counters. Check the error counters after the test.

   b. Use `opareport -o errors` to ensure there are no errors in the fabric.

2. Start and stop cable test on the management node either from the `opafastfabric` TUI or using CLI commands.

   - Using `opafastfabric` TUI

     a. Type `# opafastfabric` at the prompt to start the FastFabric TUI.

     b. >Select menu item **4** for Host Verification/Admin.

     c. Select menu item **a** for Start or Stop Bit Error Rate Cable Test.

d. Press **P** to perform the operations.

- Using CLI commands. Manually run each test for a reasonable time, typically 5 - 15 minutes.

  a. Run the cable test for hosts with `/usr/sbin/opacabletest -A -n 3 -f '/etc/opa/allhosts' start_fi`

  b. Run the cable test for ISLs with `/usr/sbin/opacabletest -A -n 3 -f '/etc/opa/allhosts' stop_fi start_isl`

  c. Use the following commands to stop the cable test and collect performance counters and statistics in csv files for viewing:

```
# /usr/sbin/opacabletest  -A -n 3 -f '/etc/opa/allhosts'  stop_isl stop_fi

# opaextractperf > link_stability_perf.csv

# opaextracterrors > link_stability_counters.csv
```

3. For large fabrics, check stability using a long run (typically 4 - 8 hours) of `opacabletest`.

> **NOTE**
>
> Short runs of 10-15 minutes are fine for initial validation.

13.3.7.2. How to Interpret the Results

The `opaextracterrors` command is a misnomer. It captures interesting statistics for evaluating links, but most of the content is not indicative of failures. The Omni-Path Express Fabric has robust end-to-end recovery mechanisms that handle issues.

Suggest looking specifically at the following columns:

- LinkWidthDnGradeTxActive - expect to see x4 Width
- LinkWidthDnGradeRxActive - expect to see x4 Width
- LinkQualityIndicator - 5 is excellent, 4 is acceptable, 3 is marginal and clearly an issue.
- LinkDowned - when an HFI is reset, the link down count increases, so rebooting a server results in small increments. If you see a link with significantly higher counts than its reboot expectations, then take a look at the server `/var/log/messages` file to determine whether the server is rebooting or the link is re-initializing.

For the other error counters, run a column sort and look for high error counts (greater than 100x) versus other links and take a look at the link types. Optical links have higher retry rates. This is not typically an issue unless they far exceed their peers.

The output is useful for verifying that every link is being tested. Unusual fabric `opaextractperf` topologies may result in non-optimum cabletest results. One workaround is to separately run isl and fi (HFI) link tests, then look at the total error results.

13.3.8. How to Debug and Fix Physical Link Issues

Check the topology before and after each of the debug steps using:

```
# opareport -o verifyall -T test_topology.xml
```

If the original issue was marginal operation rather than a hard failure, then re-run `cabletest` and analyze the `opaextracterrors` results to verify whether the issues were resolved.

At this point, you have a list of links with issues. Cornelis recommends the following approach for physical link resolution:

1. Unplug and re-insert each end of a physical cable. Check that the cable actually clicks into place. It may be useful to do this step separately for each end of the cable. Re-run `opacabletest` and verify whether the issue has been resolved or not.

   NOTE

   This step has resolved more link issues in fabric installs than all others.

2. Swap the questionable cable with a known good cable to isolate whether it is an HFI/Switch issue or cable issue.
3. If step 2 worked, then install the questionable cable into another location and verify whether it works.
4. If the issue is corrected, then the issue may be a mechanical latching issue on the HFI/Switch connector.
5. If the original issue was marginal operation rather than a hard failure, then re-run `opacabletest` and analyze the `opaextracterrors` results to verify whether the issues were resolved.
6. Re-run the physical links stability test using `opacabletest`.

13.3.9. Link Debug CLI Commands

Task	CLI Command
Identify fabric errors.	`opareport -o errors`
Identify slow links (< x4 width).	`opareport -o slowlinks`
Obtain LID of the switch.	Use `opaextractlids` for the links.
Obtain the port and links.	Use `opaextractsellinks` for the links.
If a link is not coming up as Active, first bounce the link, then check the link state.	`opaportconfig -l <lid> -m <port> bounce`
Get detailed link info for all nodes connected to an Edge switch or leaf and their neighbor.	`opareport -M -m -A -s –o comps -d 20 -F lid:<lid of edge switch>:node`

Task	CLI Command
Find links that are not plugged in or not seen by the interface. Find all links stuck in the Offline state.	`opareport -A -m -F portphysstate:offline -o comps -d 5`
Find all links stuck in the Polling state. **NOTE** A link stuck in Polling may indicate that the other end of the cable is not inserted correctly. In this case, typically, one end is Polling and the other end is Offline.	`opareport -A -m -F portphysstate:polling -o comps -d 5`
Identify bad links.	`opaextractbadlinks`
Disable all bad links and append `/etc/opa/disabled.0:0.csv` with a list of all bad links disabled. **NOTE** As a debug step, disabling bad links should be temporary.	`opaextractbadlinks | opadisableports`
Enable links previously disabled.	`cat /etc/opa/disabled.0:0.csv | opaenableports`
To bounce a link, simulate a cable pull and re-insert on a server. **NOTE** It may take up to 60 seconds for the port to re-enter the active state.	`opaportconfig bounce`
Check status of local HFI ports.	`opainfo`
`opaportconfig` and `opaportinfo` are key commands for port debugging.	Run the commands with the `-help` option to see available parameters.

Task	CLI Command
Disable a set of links by extracting them to a csv file using opaextractsellinks.	In the following example, links are extracted to linkstodisable.csv. 1. To disable a set of links, run: opadisableports < linkstodisable.csv By default, all disabled links are appended to the file /etc/opa/disabled\:1\:1.csv. 2. To enable the disabled ports, run: opaenableports < /etc/opa/disabled\:1\:1.csv After enabling the ports, the file /etc/opa/disabled\:1\:1.csv purges the links that are enabled. **NOTE** For each listed link, the switch port closer to this node is disabled. 3. Run opaportinfo -l *<lid of switch>* -m *<port number>*. 4. Check the port state by running: opaportinfo -l 3 -m 0x10

NOTE

Be sure to exclude the SM node on the Edge switch you are on and run `disableports` from the `linkstodisable` file to prevent cutting off this node from the fabric.

13.4. Using opatop to View Bandwidth and Error Summary

Use the the Fabric Performance Monitor TUI, `opatop`, to look at the bandwidth and error summary of HFIs and switches. This section provides a high-level overview of `opatop`.

1. To start up the Fabric Performance Monitor TUI, type `opatop` at the prompt.
2. Select `1` for HFIs or `2` for SW.

   Cornelis recommends that you select `2` to view SWs bandwidth and error summary.

   NOTE

   In this display, HFIs show as `Send/Rcv` and ISLs show as `Int`.

3. On the Group Information screen:
   - Select `P` for Group Performance/Bandwidth Utilization.

   - Select S for Group Statistics/Error Category Stats.

4. Use u to move to an upper level.

For additional details, see the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*.

13.5. Using the Beacon LED to Identify HFI and Switch Ports

The LED beaconing flash pattern can be turned ON/OFF with the `opaportconfig` command. This can be used to identify the HFI and switches/ports installed in racks that need attention.

- For HFI:

```
opaportconfig -l 0x001 ledoff
Disabling Led at LID 0x00000001 Port 0 via local port 1 (0x0011750101671ed9)

opaportconfig -l 0x001 ledon
Enabling LED at LID 0x00000001 Port 0 via local port 1 (0x0011750101671ed9)
```

- For Switch port:

```
opaportconfig -l 0x002 -m 40 ledon (where –m 40 is port number)
Enabling LED at LID 0x00000002 Port 40 via local port 1 (0x0011750101671ed9)

opaportconfig -l 0x002 -m 40 ledoff
Disabling Led at LID 0x00000002 Port 40 via local port 1 (0x0011750101671ed9)
```

13.6. Decoding the Physical Configuration of an HFI

The `opahfirev` command provides a quick snapshot of an Omni-Path Express HFI, providing both PCIe status and physical configuration state, complementary to the `opainfo` command.

```
# opahfirev
######################
phsmpriv07.ph.intel.com  - HFI 0000:81:00.0
HFI:   hfi1_0
Board: ChipABI 3.0, ChipRev 7.17, SW Compat 3
SN:    0x0063be82
Location:Discrete  Socket:1 PCISlot:00 NUMANode:1  HFI0
Bus:   Speed 8GT/s, Width x16
GUID:  0011:7501:0163:be82
SiRev: B1 (11)
TMM:   10.4.0.0.146
######################
```

Note the field for Thermal Monitoring Module (TMM) firmware version is an optional micro-controller for thermal monitoring on vendor-specific HFI adapters using the SMBus. Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "*Install and*

Upgrade Standalone Firmware" for details on installing TMM. For more information on the `opatmmtool`, see the *Cornelis Omni-Path Express Fabric Host Software User Guide*.

- Check the current TMM firmware version using `opatmmtool -fwversion`.

  Note that the `opatmmtool` is installed under `/usr/sbin`.
- Check the TMM firmware version in the `hfi1_smbus.fw` file using:

  Note that the `hfi1_smbus.fw` is installed under `/lib/firmware/updates/`.

```
opatmmtool -f /lib/firmware/updates/hfi1_smbus.fw fileversion
```

- If the `fwversion` is less than `fileversion`, then update the TMM firmware version using:

```
opatmmtool -f /lib/firmware/updates/hfi1_smbus.fw update
```

- After the TMM is updated, restart the TMM using:

```
opatmmtool reboot
```

NOTE

Data traffic is not interrupted during a reboot of the TMM.

13.7. Programming and Verifying Option ROM EEPROM Device

This section describes how to program and verify the Option ROM EEPROM device on an Omni-Path Express HFI.

13.7.1. Before you Begin

- If you have an Omni-Path Express HFI, continue with this process.
- If you have an HFI from another manufacturer, contact the manufacturer's support team for the supported file versions and other specific instructions.

13.7.2. Overview

The Option ROM contains three files that are pre-installed on the HFI at the time of manufacture.

- HFI1 UEFI Option ROM: `HfiPcieGen3_x.x.x.x.x.efi`

  Installed under `/usr/share/opa/bios_images/`.
- UEFI UNDI Loader: `HfiPcieGen3Loader_x.x.x.x.x.rom`

Installed under `/usr/share/opa/bios_images/`.

- HFI1 platform file: `hfi1_platform.dat`

  Installed under `/lib/firmware/updates`.

> **NOTE**
>
> The `hfi1_platform.dat` file is unique to the HFI hardware and should not need to be updated. If you do need to update this file, contact Cornelis Networks Customer Support.

Go to the Cornelis Customer Center. Select the **Release Library** to locate and download the following files and tools that are required to program the Options ROM EEPROM:

- Omni-Path Express Fabric Firmware Tools
- Omni-Path Express HFI UEFI Firmware

Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "*Install and Upgrade Standalone Firmware*" for details.

Note that the `hfi1_platform.dat` file is unique to the HFI hardware and is not provided on the Cornelis Customer Portal. For HFI cards, this file may be obtained by contacting Customer Support email address. For non-Cornelis HFIs, contact your hardware supplier. If needed and `hfi1.platform.dat` is provided, put it in the `/lib/firmware/updates` folder.

13.7.3. Single Rail (one HFI) Example

The following example shows how to program a single HFI with three partitions.

1. Enter the following commands, replacing `x.x.x.x.x` with the versions provided in the release:

   In normal mode user can only update HfiPcieGen3Loader.x.x.rom and HfiPcieGen3_x.x.x.efi using only one command.

   ```
   # hfi1_eprom -u /usr/share/opa/bios_images/*
   ```

   Individual files can be only updated in service mode.

   ```
   # hfi1_eprom -S -w -c /lib/firmware/updates/hfi1_platform.dat
   # hfi1_eprom -S -w -o /usr/share/opa/bios_images/HfiPcieGen3Loader_x.x.x.x.x.rom
   # hfi1_eprom -S -w -b /usr/share/opa/bios_images/HfiPcieGen3_x.x.x.x.x.efi
   ```

   > **NOTE**
   >
   > Ensure you select the correct partitions for your files.

2. Verify programmed versions using the following command:

```
# hfi1_eprom -V
```

3. Reboot the server for the firmware updates to take effect.

13.7.4. Dual Rail (two HFIs) Example

The following example shows how to program two HFIs with three partitions for each HFI.

1. Obtain the device assignment using the following Linux command:

```
# lspci | grep HFI
05:00.0 Fabric controller: Cornelis Omni-Path HFI Silicon 100 Series [discrete] (rev
11)
81:00.0 Fabric controller: Cornelis Omni-Path HFI Silicon 100 Series [discrete] (rev
11)
```

2. Program the devices using the following commands, replacing `x.x.x.x.x` with the versions provided in the release:

NOTE

Ensure you use the correct device assignment.

In normal mode user can only update HfiPcieGen3Loader.x.x.rom and HfiPcieGen3_x.x.x.efi using only one command.

```
# hfi1_eprom -d /sys/bus/pci/devices/0000:05:00.0/resource0 -u /usr/share/opa/
bios_images/*

hfi1_eprom -d /sys/bus/pci/devices/0000:81:00.0/resource0  -u /usr/share/opa/
bios_images/*
```

Individual files can be only updated in service mode.

```
# hfi1_eprom -d /sys/bus/pci/devices/0000:05:00.0/resource0 -S -w -c /lib/firmware/
updates/hfi1_platform.dat
# hfi1_eprom -d /sys/bus/pci/devices/0000:05:00.0/resource0 -S -w -o /usr/share/opa/
bios_images/HfiPcieGen3Loader_x.x.x.x.x.rom
# hfi1_eprom -d /sys/bus/pci/devices/0000:05:00.0/resource0 -S -w -b /usr/share/opa/
bios_images/HfiPcieGen3_x.x.x.x.x.efi

# hfi1_eprom -d /sys/bus/pci/devices/0000:81:00.0/resource0 -S -w -c /lib/firmware/
updates/hfi1_platform.dat
# hfi1_eprom -d /sys/bus/pci/devices/0000:81:00.0/resource0 -S -w -o /usr/share/opa/
bios_images/HfiPcieGen3Loader_x.x.x.x.x.rom
# hfi1_eprom -d /sys/bus/pci/devices/0000:81:00.0/resource0 -S -w -b /usr/share/opa/
bios_images/HfiPcieGen3_x.x.x.x.x.efi
```

3. Verify programmed versions using the following commands:

```
# hfi1_eprom -d /sys/bus/pci/devices/0000:05:00.0/resource0 -V

# hfi1_eprom -d /sys/bus/pci/devices/0000:81:00.0/resource0 -V
```

4. Reboot the server for the firmware updates to take effect.

13.8. Verifying Fabric Manager Sweep

By default, Fabric Manager sweeps every five minutes as defined in the `/etc/opa-fm/opafm.xml` file. Sweeps are triggered sooner if there are fabric changes such as hosts, switches, or links going up or down. Edit `/var/log/messages` and search for `CYCLE START`. Each cycle start has a complementary cycle end. Any links with errors are noted during this sweep cycle.

An example of a clean SM sweep follows:

```
Feb 16 16:12:08 hds1fnb8261 fm0_sm[3946]: PROGR[topology]: SM: topology_main: TT:
DISCOVERY CYCLE START - REASON: Scheduled sweep interval
Feb 16 16:12:08 hds1fnb8261 fm0_sm[3946]: PROGR[topology]: SM: topology_main: DISCOVERY
CYCLE END. 9 SWs, 131 HFIs, 131 end ports, 523 total ports, 1 SM(s), 1902 packets, 0
retries, 0.350 sec sweep
```

Compare the sweep results with `opafabricinfo` and the fabric topology.

13.9. Verifying PM Sweep Duration

To show the PM sweep duration, perform the following steps:

1. Open `opatop`, then select `i`.

```
opatop: Img:Tue Feb 16 01:54:43 2016, Hist  Now:Tue Feb 16 09:53:26 2016
Image Info:
 Sweep Start: Tue Feb 16 01:54:43 2016
 Sweep Duration: 0.001 Seconds

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

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

2. Select `r` to traverse the previous sweep duration time from history files. By default, PM sweeps every ten seconds. The latest ten image files (100 sec) are stored in RAM and up to 24 hours of history is stored in `/var/usr/lib/opa-fm`.

13.10. Checking Credit Loop Operation

For details on credit loops, see the *Cornelis Omni-Path Express Fabric Suite Fabric Manager User Guide*.

To verify that a fabric does not have a credit loop issue, use:

```
# opareport -o validatecreditloops
```

The output should report similar to the following where no credit loops are detected:

```
Fabric summary: 135 devices, 126 HFIs, 9 switches,
504 connections, 16880 routing decisions,
15750 analyzed routes, 0 incomplete routes
Done Building Graphical Layout of All Routes
Routes are deadlock free (No credit loops detected)
```

13.11. Modifying the Fabric Manager Routing Algorithm

If long Fabric Manager (FM) sweep times are observed or FM sweeps do not finish when a large number of nodes are bounced, consider changing the FM routing algorithm to `fattree` from the default `shortestpath`. You can do this by updating the `/etc/opa-fm/opafm.xml` file as shown in the following example:

```
<!-- **************** Fabric Routing **************************** -->
<!-- The following Routing Algorithms are supported -->
<!-- shortestpath - pick shortest path and balance lids on ISLs -->
<!-- dgshortestpath - A variation of shortestpath that uses the        -->
<!--            RoutingOrder parameter to control the order in which   -->
<!--            switch egress ports are assigned to LIDs being routed  -->
<!--            through the fabric. This can provide a better balance  -->
<!--            of traffic through fabrics with multiple types of end  -->
<!--            nodes.                                                 -->
<!--            See the <DGShortestPathTopology> section, below, for   -->
<!--            more information.                                      -->
<!-- fattree -  A variation of shortestpath with better balancing      -->
<!--            and improved SM performance on fat tree-like fabrics.  -->
<RoutingAlgorithm>fattree</RoutingAlgorithm>
```

14. Run Benchmark and Stress Tests

For details on the tests provided with the Omni-Path Express Fabric Software, refer to the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*.

For optimal performance when running benchmark tests, configure systems according to the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*.

14.1. Running Bandwidth Test

From `$FF_MPI_APPS_DIR` run:

```
# ./run_bw3
```

This test uses hosts defined in the `mpi_hosts` file, however, only the first two hosts in the file are used.

14.2. Running Latency Test

From `$FF_MPI_APPS_DIR` run:

```
# ./run_lat3
```

This test uses hosts defined in the `mpi_hosts` file, however, only the first two hosts in the file are used.

14.3. Running MPI Deviation Test

From `$FF_MPI_APPS_DIR` run:

```
# ./run_deviation 20 20 50
```

This test uses hosts defined in the `mpi_hosts` file.

14.4. Running run_mpi_stress

The default traffic pattern is "all-to-all" for this test.

Refer to *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide* for detailed information.

The test is located in `$FF_MPI_APPS_DIR`.

NOTE

These steps assume that you have defined hosts in the `/usr/src/opa/mpi_apps/mpi_hosts` file.

1. Clear error counters using

   ```
   # opareport -o none --clearall
   ```

2. Confirm no errors exist using

   ```
   # opareport -o error
   ```

3. Run a 60-minute stress test using

   ```
   # ./run_mpi_stress all -t 60
   ```

4. Run `opatop` to monitor the link utilization during the test.
5. Check error counts after the test using

   ```
   # opareport -o errors
   ```

6. View the log file that is available for analysis in `$FF_MPI_APPS_DIR/logs`. The log filename format is `mpi_stress.`*`date_time`*`.`
7. Extract the log file in CSV format for errors and performance using

   ```
   # opaextracterror
   # opaextractperf
   ```

14.4.1. Example

The following example demonstrates this test being run on four nodes.

```
# ./run_mpi_stress all -t 60
Running MPI tests with 4 processes
 logfile /usr/src/opa/mpi_apps/logs/mpi_stress.12Apr17150628
OpenMPI Detected, running with mpirun.
 Running Mpi Stress ... Running Mpi Stress ...
Using hosts list: /usr/src/opa/mpi_apps/mpi_hosts
Hosts in run:
node1
node2
node3
node4

+ /usr/mpi/gcc/openmpi-1.10.4-hfi/bin/mpirun -np 4 -map-by node
--allow-run-as-root -machinefile /usr/src/opa/mpi_apps/mpi_hosts
-mca plm_rsh_no_tree_spawn 1
/usr/mpi/gcc/openmpi-1.10.4-hfi/tests/intel/mpi_stress -t 60
```

15. Take State Dump of a Switch

> **NOTE**
>
> Taking a state dump is a disruptive process and requires reboot of the switch after the state dump is taken. A state dump should only be taken if required to debug an issue.

You can take a state dump of any switch in the fabric, using its LID.

15.1. Prerequisites

- Find the LID of the switch whose state you want to dump by running the `opaextractlids|grep` *`switch name`* command.
- Contact Cornelis Customer Support to get the correct username and password for the `supportLogin` command.

15.2. Procedure

The following example describes how to take a state dump of a switch.

1. Use the default username (`admin`) and password (`adminpass`) to log in to a managed switch.
2. Using the support username and password obtained in **Prerequisites**, run the `supportLogin` command as shown.

   ```
   -> supportLogin
   username: support
   password:
   ```

3. Depending on your type of switch, run `ismTakeStateDump` using one of the two methods below.
   - For a locally managed switch, run the `ismTakeStateDump` command.
   - For a remote-managed or externally-managed switch, run the command below, where <*`lid`*> identifies the desired switch.

     ```
     -> ismTakeStateDump -lid <lid>
     Dumping state of the switch at lid 4 to /firmware/prr-LID0004.gz
     ```

4. From the Management Node, SFTP to the managed switch used for running the state dump command to retrieve the log.

   ```
   sftp admin@<managed switch> with password adminpass.
   admin@10.228.222.20's password:
   ```

```
Connected to 10.228.222.20.
sftp> dir
admin       operator      prr-LID0004.gz   prr-LID0005.gz   prr-LID0015.gz
get prr-LID0004.gz
```

5. Reboot the switch on which the state dump was taken to clear the state dump. For externally-managed switches, use FastFabric to reboot the switch.

16. Perform Final Fabric Checks

After addressing all issues, perform final fabric checks as described in Section 12 “Verify Server and Fabric”.

17. Customizing MPI

This section provides information for customizing and rebuilding MPI Libraries and MPI Applications, as needed.

17.1. Building MPI Libraries

Cornelis Omni-Path Express Fabric Suite comes complete with prebuilt versions of OpenMPI enabled to use the OPX provider and PSM2, and MVAPICH2 that is enabled to use PSM2. The source code for the provided MPI versions is included with the Omni-Path Express Fabric Suite and can be used with the FastFabric TUI to rebuild the MPI library along with the selection of a few options, such as compiler (refer to the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*, Managing the Host Configuration). The src rpms for the MPI libraries are installed under `/usr/src/opa/MPI/` along with a `do_build` script that FastFabric uses to prompt the user to rebuild the MPI library.

If you want more advanced customization of the MPI options, refer to the `do_openmpi_build` and `do_mvapich2_build` scripts. Each of these scripts provide build options to the respective source rpms allowing you to rebuild the MPI library along with some standard test programs and benchmarks. Also, you can use these scripts as samples of the options needed to rebuild OpenMPI and MVAPICH2. For further details, the OpenMPI and MVAPICH2 libraries are documented in https://www.open-mpi.org/ and http://mvapich.cse.ohio-state.edu/, respectively.

Intel MPI is shipped separately from Omni-Path Express Fabric Suite as part of Intel Parallel Studio. For information on Intel MPI, refer to https://software.intel.com/en-us/parallel-studio-xe. For more information on using Intel MPI with Omni-Path Express, see the *Cornelis Omni-Path Express Fabric Host Software User Guide*, Using MPI.

17.2. Building and Running MPI Applications

The Omni-Path Express Fabric Suite comes with the source code for a number of MPI benchmarks. You can use the FastFabric TUI to rebuild the MPI benchmarks and distribute them to the compute nodes in the cluster (refer to the *Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*, Managing the Host Configuration and MPI Sample Applications). The source code for the benchmarks are provided in `/usr/src/opa/mpi_apps` together with a `makefile` for building them as well as some sample `run_*` scripts. You may use these benchmarks and scripts to verify basic fabric and cluster performance during deployment. As such, they directly invoke the appropriate `mpirun` command for the selected MPI library using a local file listing the hostname for each rank in the job.

Typical production clusters use a job scheduler to select nodes for a given job, launch the MPI application, and monitor for its completion. Documentation on job schedulers is beyond the scope of this document. Consult with your system integrator for further details on using the job scheduler installed on the cluster.

Appendix A. Glossary of Acronyms

The following acronyms are used throughout the Omni-Path Express Documentation Set.

ACPI	Advanced Configuration and Power Interface - an industry specification that allows the BIOS, OS, and peripherals to communicate with each other about power usage
AOC	Active Optic Cable - an optic cable used to connect modules in theOmni-Path Express Fabric
BIOS	Basic Input Output System - firmware used to initialize and test system hardware components, and load a boot loader from a mass storage device which then initializes a kernel
BKM	Best Known Methods
CLI	Command Line Interface - text-based user interface used to run programs, manage files, and interact with a computer
CPU	Central Processing Unit - primary chip in a computer used to execute the instructions comprising a computer program
CRC	Cycle Redundancy Check - an error-detecting code used to detect accidental changes to digital data
CSV	Comma Separated Values - a file format commonly associated with spreadsheets
CUDA	Compute Unified Device Architecture - A platform and API created by Nvidia that allows software to use certain types of GPUs for general-purpose processing
DCS	Director Class Switch
DNS	Domain Name System - a database that translates human readable website names into their numeric IP addresses
DST	Delivered to System Test
EEPROM	Electrically Erasable Programmable Read-only Memory - user-modifiable read-only memory
EMI	Electromagnetic Interference - an external disturbance that may degrade the performance of a circuit or even stop it from functioning
FE	Fabric Executive - an Omni-Path Express component that provides out-of-band access to the Fabric Manager

FM	Fabric Manager - an Omni-Path Express Fabric Suite component responsible for managing the fabric using management packets over a dedicated virtual lane (VL15)
GUI	Graphical User Interface - an interface that allows users to interact with computers through graphical icons
GUID	Globally Unique Identifier - a 128-bit text string generated when a unique reference number is needed to identify a component (e.g., hardware, software, account, node, etc.) on a computer or network
HFI	Host Fabric Interface - the hardware and software that allows one module to talk to another in an Omni-Path Express Fabric.
IP	Internet Protocol - a set of rules for communicating over the internet
IPMI	Intelligent Platform Management Interface - a set of interfaces used by system administrators for out-of-band management of computer systems and monitoring of their operation
IPoIB	Internet Protocol over InfiniBand
ISL	Inter-Switch Link - a Cisco Systems proprietary protocol that maintains VLAN information in Ethernet frames as traffic flows between switches and routers, or switches and switches
LED	Light Emitting Diode - a semiconductor that emits light when current flows through it
LID	Local Identifier
MM	Management Module
MPI	Message Passing Interface - a standardized means of exchanging messages between multiple computers running a parallel program across distributed memory
NTP	Network Time Protocol - a networking protocol for clock synchronization between computer systems over packet-switched, variable-latency data networks
Omni-Path Express HFI	Omni-Path Express Host Fabric Interface
OPX Provider or OPX Libfrabic Provider	OPX provider is a new enhanced libfabric provider that takes full advantage of the libfabric acceleration features while running over existing and future Omni-Path Express hardware.

OPXS	Omni-Path Express Fabric Suite
OS	Operating System - software that manages computer hardware, software resources, and provides common services for computer programs
PCIe	Peripheral Component Interconnect Express - a high-speed serial computer expansion bus standard for connecting a computer to one or more peripheral devices
PM	Performance Manager, Performance Management, Program Manager, Program Management
PSM	Performance Scaled Messaging - a low-level user-level communications interface
QSFP	Quad Small Form-factor Pluggable - a compact, hot-pluggable transceiver used for data communications applications
RDMA	Remote Direct Memory Access - technology that enables two networked computers to exchange data in main memory without CPU intervention
RPM	Red Hat Package Manager
SCP	System Control Processor
SFTP	Secure File Transfer Protocol - a network protocol for securely accessing, transferring, and managing large files and sensitive data
SM	Subnet Manager or Subnet Management
SSH	Secure Shell - a cryptographic network protocol for operating network services securely over an unsecured network
TCP	Transmission Control Protocol - a standard that defines how to establish and maintain a network conversation by which applications can exchange data
TCP/IP	Transmission Control Protocol/Internet Protocol - a framework for organizing the set of communication protocols used in the Internet and similar computer networks according to functional criteria
TMM	Thermal Management Module - an electronically controlled thermostat
TUI	Text (or Text-based) User Interface

UDP	User Datagram Protocol - a communication protocol used across the Internet for especially time-sensitive transmissions
UEFI	Unified Extensible Firmware Interface - a publicly available specification that defines a software interface between an operating system and platform firmware
USB	Universal Serial Bus - a plug and play interface that allows a computer to communicate with peripheral devices
VPD	Vital Product Data - information that uniquely records each hardware and licensed-internal-code element in the nodes
XML	Extended Markup Language - a markup language (similar to html) and file format for storing, transmitting, and reconstructing arbitrary data; it defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

Appendix B. Older Revisions

Date	Rev	Description
Sep 2020	14.0	Updated “Configure Omni-Path Director Class Switch” to remove contentious terms from text.
Aug 2020	13.0	Updated “Best Practices” to include pointer to “Updating the Certificate” in the *Cornelis Omni-Path Switches GUI User Guide*.
Jan 2020	12.0	Updated Download Software links in “Before You Begin” and “Programming and Verifying Option ROM EEPROM Device”.
Oct 2019	11.0	Added new section “Customizing MPI” including “Building MPI Libraries” and “Building and Running MPI Applications”.
Dec 2018	10.0	Updated “Installing Drivers” to include installing using the Omni-Path Express Fabric Software-included Linux OS repository.
Sep 2018	9.0	Globally, performed minor edits to text, format, and structure for consistency and usability.
Apr 2018	8.0	Made several minor updates throughout.
Oct 2017	7.0	The *Intel Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide* has been merged into the *Intel Omni-Path FastFabric User Guide*. See the “Intel Omni-Path Documentation Library” for details.
Aug 2017	6.0	Added NVIDIA bullet in the “Installing Drivers” section of the “Install Intel Omni-Path Software”.
Apr 2017	5.0	Added “Programming and Verifying Option ROM EEPROM Device” section.
Dec 2016	4.0	Added “Cluster Configurator for Intel Omni-Path” to “Preface”.
Oct 2016	3.0	Made several minor changes throughout.
Aug 2016	2.0	Updated “Generate Cable Map Topology Files” to include Intel Omni-Path Director Class Switch information.
May 2016	1.0	Initial release.