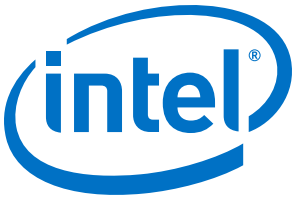


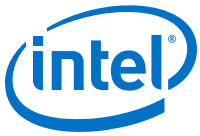
Intel® Omni-Path Fabric Software in SUSE* Linux* Enterprise Server 15

Release Notes

Rev. 3.0

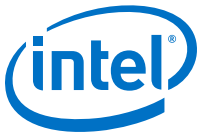
April 2020

Doc. No.: J98644, Rev.: 3.0



Contents

Tables



1.0 Overview of the Release

These Release Notes are intended for Intel® Omni-Path IFS software provided inbox with the OS release. This document provides a brief overview of the changes introduced into the Intel® Omni-Path Software by this release. References to more detailed information are provided where necessary. The information contained in this document is intended as supplemental information only; it should be used in conjunction with the documentation provided for each component.

These Release Notes list the features supported in this software release, open issues, and issues that were resolved during release development.

1.1 Audience

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

1.2 Software License Agreement

This software is provided 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.

1.3 If You Need Help

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

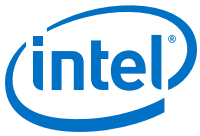
1.4 Packages in This Release

Intel® Omni-Path Software Packages
Packages created by Intel
opa-address-resolution-10.6.0-3.13.x86_64
opa-basic-tools-10.6.0-3.13.x86_64
opa-fastfabric-10.6.0-3.13.x86_64
opa-fm-10.6.0-2.44.x86_64
libfabric1-1.5.3-2.7.x86_64
libopamgt0-10.6.0-3.13.x86_64
libpsm2-2-10.3.37-1.15.x86_64
Firmware binaries delivered by Intel
8051 firmware version 1.24.0
continued...

Intel® Omni-Path Software Packages
SBus Master firmware version 0x10130001
PCIe SerDes firmware version 0x4755
Fabric SerDes firmware version 0x1055
Packages used by Intel
rdma-core-16.2-1.3.x86_64 (libhfi1)
openmpi2-2.1.3-1.14.x86_64
mvapich2-2.2-7.17.x86_64
mpitests-openmpi2-3.2-1.17.x86_64
mvapich2-psm2-2.2-7.13.x86_64
mpitests-mvapich2-psm2-3.2-1.9.x86_64

1.5 Supported Features

- The list of supported hardware is in Table 2 on page 7.
- Topology-aware job scheduling, which is enabled by the opamgt library and allows developers to write code that interfaces to the SA/PA.New features in this release include: Switch Cost Record query and Asynchronous Trap subscriptions. See the *Intel® Omni-Path Management API Programmer's Guide* for details.
- Active Optical Cables. For details, see the Cable Matrix at: http://www.intel.com/content/www/us/en/high-performance-computing-fabrics/omni-path-cables.html
  - Support for active optical cables (AOC) on server platforms using integrated HFI for OPA (commonly known as "-F").
  - Support for Power Class 2 active optical cables (AOC). See Product Constraints on page 10 for more information.
- Legacy BIOS Boot Mode Enhancements to support boot over fabric, custom board descriptions, and pre-boot platform configuration data for AOC support.
- Multi-endpoint functionality. See the *Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide* for details.
- Support for OpenFabrics Interfaces (OFI), a framework that includes libraries (including libfabric) and applications used to export fabric communication services to applications.
- Support for NVMe over Fabric Protocol
- Support for IBM* Platform MPI and IBM* Spectrum MPI. See Supported MPI Libraries on page 7 for details.
- Virtual Fabric creation has been enhanced to better support advanced topologies, including the ability to place multicast traffic on a separate SL from unicast traffic. For details, see the *Intel® Omni-Path Fabric Suite Fabric Manager User Guide*, section 2.
- Support for the Enhanced Hypercube Routing Engine is outside the scope of Intel® OPA support. However, Intel partners may offer such support as part of their solutions. In addition there is an open source community who may be able to answer specific questions and provide guidance with respect to the Enhanced Hypercube Routing Engine.



- Additional Product Constraints described in Product Constraints on page 10.

1.6 Supported MPI Libraries

The table below lists the different MPI libraries supported by Intel® Omni-Path Fabric Software. Note that the second column indicates whether the MPI library is included in the distribution or not.

Table 1. Supported MPI Libraries

MPI Implementation	Included in Distribution?	Runs Over
Open MPI 1.10.7	Yes	PSM2
MVAPICH2-2.2	Yes	PSM2

1.7 Intel Hardware

The following table lists the Intel hardware supported in this release.

NOTE

The Intel® PSM2 implementation has a limit of four (4) HFIs.

Table 2. Supported Hardware

Hardware	Description
Intel® Xeon® Processor E5-2600 v3 product family	Haswell CPU-based servers
Intel® Xeon® Processor E5-2600 v4 product family	Broadwell CPU-based servers
Intel® Xeon® Scalable Processors	Skylake CPU-based servers
Intel® Xeon Phi™ x200 Product Family	Knights Landing CPU-based servers
Intel® Xeon Phi™ 72x5 Processor Family	Knights Mill CPU-based servers
Intel® Omni-Path Host Fabric Interface 100HFA016 (x16)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Host Fabric Interface 100HFA018 (x8)	Single Port Host Fabric Interface (HFI)
Intel® Omni-Path Switch 100SWE48Q	Managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48U	Externally-managed 48-port Edge Switch
Intel® Omni-Path Switch 100SWE48UFH	Externally-managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE48QFH	Managed 48-port Edge Switch, hot-swap power and fans
Intel® Omni-Path Switch 100SWE24Q	Managed 24-port Edge Switch
Intel® Omni-Path Switch 100SWE24U	Externally-managed 24-port Edge Switch
Intel® Omni-Path Director Class Switch 100SWD24	Director Class Switch 100 Series, up to 768 ports
Intel® Omni-Path Director Class Switch 100SWD06	Director Class Switch 100 Series, up to 192 ports

1.8 Intel® OPA Compatibility Matrix

The following component versions are compatible with Intel® Omni-Path software in SLES* 15.

Table 3. Intel® OPA Compatibility Matrix

UEFI	TMM	Managed Switch	Externally-Managed Switch	FM GUI
1.7.2.0.0	10.7.0.0.3	10.7.0.0.146	10.7.0.0.144	10.7.0.0.145
1.6.0.0.0	10.4.0.0.146	10.6.1.0.3	10.6.1.0.1	10.6.0.0.136
1.5.2.0.0	10.4.0.0.146	10.5.1.0.2	10.5.0.0.143	10.5.0.0.141

1.9 Installation Requirements

This section provides instructions and information on installing the software.

1.9.1 Best Practices

- Intel recommends that users update to the latest versions of Intel® Omni-Path firmware and software to obtain the most recent functional and security updates.
- To improve security, the administrator should log out users and disable multi-user logins prior to performing provisioning and similar tasks.

1.9.2 Installation Instructions

Perform the steps in this section to install the default Intel® Omni-Path Software configuration.

Assumptions

- You are logged in as root or with root privileges.
- You have a list of IPv4 addresses and netmasks for each IPoIB interface you are going to set up.

References

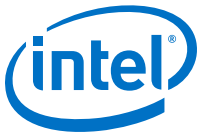
- Refer to the *Intel® Omni-Path Fabric Software Installation Guide* for related software requirements and next steps.
- Refer to the *Intel® Omni-Path Fabric Switches Hardware Installation Guide* for related firmware requirements.

Procedures

Perform the following steps to install the default Intel® OP Software configuration using SLES* OS:

Step	Task/Prompt	Action
Install OPA-Basic Software		
1.	At the command prompt, enter the installation command for `opa-basic-tools`.	Type **`zypper install -y opa-basic-tools`** and press **Enter**.

continued...



Step	Task/Prompt	Action
2.	At the command prompt, reboot the server.	Type **reboot** and press **Enter**.

3.	Check your link using `opainfo`.	Type **opainfo** and press **Enter**.
		Example output:
		`hfi1_0:1 PortGID: 0xfe80000000000000:001175010163f931`
		`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: 0x00000010-0x00000010 SM LID: 0x0000000c SL: 0`
		`QSFP: AOC , 5m FINISAR CORP P/N FCBN425QB1C05 Rev A`
		`Xmit Data: 0 MB Pkts: 251`
		`Recv Data: 0 MB Pkts: 251`
		`Link Quality: 5 (Excellent)`

4.	Install the `rdma-core` rpm.	Type **zypper install -y rdma-core** and press **Enter**.
5.	On all compute nodes: install the PSM2 library.	Type **zypper install -y libpsm2-2** and press **Enter**.
Install Intel® Omni-Path Fabric Suite Components on the Management Node		
6.	Install FastFabric.	Type **zypper install -y opa-fastfabric** and press **Enter**.
8.	Install Fabric Manager.	Type **zypper install -y opa-fm** and press **Enter**.
9.	Start the Fabric Manager.	Type **systemctl start opafm** and press **Enter**.
Set up IPoIB IPV4 Configuration		

10.	Manually edit or create the `ifcfg-ibX` file.	*Note:* Use the OS distribution-supplied instructions for setting up network interfaces.
		Type **cat /etc/network/ifcfg-ib0** and press **Enter**.
		Example output:
		`BOOTPROTO=static`
		`IPADDR=192.168.0.1`
		`BROADCAST=192.168.0.255`
		`NETWORK=192.168.0.0`
		`NETMASK=255.255.255.0`
		`STARTMODE=auto`
		`IPOIB_MODE='connected'`
		`MTU=65520`

11.	Bring up the ib0 interface.	Type **ifup ib0** and press **Enter**.

12.	Perform a test ping.	Type **ping <remote IPoIB address>** and press **Enter**.
		For example:
		`ping 10.228.200.161`
		`PING 10.228.200.161 (10.228.200.161) 56(84) bytes of data.`
		`64 bytes from 10.228.200.161: icmp_seq=1 ttl=64 time=0.863 ms`

continued...

Step	Task/Prompt	Action
(Optional) Install the Fabric Manager GUI		
13.	On one node in the fabric: install the Fabric Manager GUI.	*Note:* Intel recommends not to install the Fabric Manager GUI on the Management Node where the Fabric Manager is being used. Type **`zypper install -y opa-fmgui`** and press **Enter**.
End Task		

1.10 Product Constraints

- Power class 2 AOC are supported. Specifically, 1.5 level UEFI or higher are required for proper operation. Integrated HFI (-F) requires a specific BIOS level to support power class 2 AOC; contact your BIOS vendor for more information.
- The minimum firmware version for Intel® Omni-Path Host Fabric Interface Silicon 100 Series Switch ASIC is 10.2.

1.11 Product Limitations

This release has the following product limitations:

- Performance Administration (PA) Failover should not be enabled with FMs running on differing software versions.

  To disable PA failover, edit the `/etc/sysconfig/opafm.xml` file and in the `<Pm>` section, change `<ImageUpdateInterval>` to `0`.
- Enabling UEFI Optimized Boot on some platforms can prevent the HFI UEFI driver from loading during boot. To prevent this, do not enable UEFI Optimized Boot.

1.12 Document Versions

The following table lists the end user document versions supported by this release.

Table 4. Supported Document Versions

Title	Doc. Number	Revision
Key: Shading indicates the URL to use for accessing the particular document.		
• Intel® Omni-Path Switches Installation, User, and Reference Guides: http://www.intel.com/omnipath/SwitchPublications		
• Intel® Omni-Path Software Installation, User, and Reference Guides (includes HFI documents): http://www.intel.com/omnipath/FabricSoftwarePublications		
• Drivers and Software (including Release Notes): http://www.intel.com/omnipath/Downloads		
Intel® Omni-Path Fabric Quick Start Guide	J57479	3.0
Intel® Omni-Path Fabric Setup Guide (Old title: *Intel® Omni-Path Fabric Staging Guide*)	J27600	7.0
Intel® Omni-Path Fabric Switches Hardware Installation Guide	H76456	6.0
Intel® Omni-Path Host Fabric Interface Installation Guide	H76466	5.0

continued...

Title	Doc. Number	Revision
Intel® Omni-Path Fabric Software Installation Guide	H76467	8.0
Intel® Omni-Path Fabric Switches GUI User Guide	H76457	8.0
Intel® Omni-Path Fabric Switches Command Line Interface Reference Guide	H76458	8.0
Intel® Omni-Path Fabric Suite FastFabric User Guide (Merged with: *Intel® Omni-Path Fabric Suite FastFabric Command Line Interface Reference Guide*)	H76469	8.0
Intel® Omni-Path Fabric Suite Fabric Manager User Guide	H76468	8.0
Intel® Omni-Path Fabric Suite Fabric Manager GUI User Guide	H76471	8.0
Intel® Omni-Path Fabric Host Software User Guide	H76470	8.0
Intel® Performance Scaled Messaging 2 (PSM2) Programmer's Guide	H76473	8.0
Intel® Omni-Path Fabric Performance Tuning User Guide	H93143	10.0
Intel® Omni-Path IP and LNet Router Design Guide	H99668	5.0
Building Lustre Servers with Intel® Omni-Path Architecture Application Note*	J10040	1.0
Building Containers for Intel® Omni-Path Fabrics using Docker and Singularity* Application Note*	J57474	4.0
Intel® Omni-Path Management API Programmer's Guide	J68876	2.0
Intel® Omni-Path Fabric Software Release Notes	J82662	1.0
Intel® Omni-Path Fabric Manager GUI Release Notes	J82663	1.0
Intel® Omni-Path Fabric Switches Release Notes (includes managed and externally- managed switches)	J82661	1.0

2.0 Issues

This section lists the open issues in the Intel® Omni-Path Software.

2.1 Open Issues

The following table lists the open issues for this release.

Table 5. Open Issues

ID	Description	Workaround
129563	Memory allocation errors with MVAPICH2-2.1/Verbs.	When running MVAPICH2 jobs with a large number of ranks (for example, between 36 and 72 ranks), you must set the following parameters in `/etc/security/limits.conf`: • hard memlock unlimited • soft memlock unlimited Also, you must increase the lkey_table_size:LKEY table size in bits (2^n, 1 <= n <= 23) from its default of 16 to 17. For instructions on setting module parameters, refer to Appendix A in the *Intel® Omni-Path Fabric Performance Tuning User Guide* .
134494	Open MPI uses srand() family functions at MPI_Init() time. Therefore, if the user sets srand() before calling MPI_Init(), the values will be altered.	1. Fixed in Open MPI 2.0.1. 2. Call srand() functions family after calling MPI_Init().

135390	Very old HFI adapters may be programmed with an obsolete version of the AOC platform configuration file. In these cases, errors such as the following may be observed:
	`[ 26.903186] hfi1 0000:d5:00.0: hfi1_0: parse_platform_config:Bad config file`
	`[ 26.903186] hfi1 0000:d5:00.0: hfi1_0: parse_platform_config:File claims to be larger than read size`
	`[ 27.351555] hfi1 0000:d5:00.0: hfi1_0: tune_serdes: Unknown port type`
		Update the platform configuration file on the HFI to the current version. For details, see the *Intel® Omni-Path Fabric Software Installation Guide*, section B.1.

135975	After performing an OPA software configuration update, switches will show the new settings when queried by `opaswitchadmin` tools, however, individual ports will continue to operate using the previous settings, including LinkWidth enable.	Perform one of the following: • Reboot the switch to apply the new settings. • Bounce the affected port(s).
136728	If hundreds of links are bouncing while the FM is sweeping, the FM sweep time may be significantly extended. This can result in unexpected delays in FM responsiveness to fabric changes or host reboots. (The issue is that active links bounce between the time FM discovers one side of the link versus the other side of the link.)	Ensure the following value is present in the `<Sm>` section of the `opafm.xml` file: `<CumulativeTimeoutLimit>300</ CumulativeTime outLimit>`
136822	The Intel UEFI driver contained in the server BIOS must be executed for proper support of Active Optical Cables (AOC) in an integrated HFI environment. Some	Avoid the use of Legacy BIOS boot mode if your platform does not execute the HFI driver in that mode.

continued...

ID	Description	Workaround
	BIOS do not execute the UEFI in Legacy BIOS Boot mode, and there are BIOS configuration settings that may prevent the UEFI from executing in any mode.	Avoid BIOS settings or other configuration settings that do not execute the HFI driver during boot.
137951	In the HFI BIOS screen for Advanced NIC Configuration, a warning message about incorrect custom P_Key value is not completely displayed.	The valid range for custom P_Key value is 0x8001 to 0xFFFE.
139550	Infrequently, an AOC may exhibit an unexpectedly high local link integrity error rate after the link comes up, relative to the error rate on previous link up occasions. This can be determined by observing a link quality of <5. These links may eventually experience a link width downgrade.	Bounce the link.
139613	The Subsystem Vendor and Subsystem Device ID in the PCI configuration space of Intel® Omni-Path discrete HFI cards may not indicate the correct OEM vendor and device. As a result, the `lspci` command may show incorrect Subsystem Vendor and Device ID information. This issue affects Intel server boards for Intel® Xeon® Processor v3 and v4 Product Family configured in Legacy OS boot mode.	Reconfigure the system from Legacy OS boot mode to UEFI boot mode.
140691	When running `opaswitchadmin` against multiple externally-managed switches simultaneously, it sends schedule requests in parallel to those hosts. It is possible that some hosts may intermittently fail due to timeouts at high levels of parallelism.	The number of hosts that are queried in parallel can be limited by setting the FF_MAX_PARALLEL value in `/etc/opa/opafastfabric.conf`. Results may vary by fabric. Intel recommends reducing this to 2.
140881	In rare cases when an LNI failure occurs, the link will not come up after manually disabling and re-enabling the link.	Reload the driver.
140911	The OFI verbs provider does not support FI_EP_RDM End Point type. This End Point type is needed for Open MPI OFI support. Therefore, Open MPI OFI support will not run over the verbs provider.	If needed to run Open MPI over verbs, use the native verbs support in the Open MPI build included in the distro.

142330	MPI applications that leverage the PSM2 library's access to the HFI ASICs Memory Mapped IO and that access the MMIO directly (not via PSM2) can potentially cause an "unsupported opcode" error which some servers handle as a critical error.	Disable upstream error reporting using the AER mask register.
		• For discrete HFI ASICs (e.g., CHF PCIe card), use
		`setpci -d 8086:24f0 ECAP_AER +8.l=00100000:00100000`
		• For integrated HFIs (e.g., KNL-F and SKX-F), use
		`setpci -d 8086:24f1 ECAP_AER +8.l=00100000:00100000`