Cornelis™ Omni-Path Express™ Fabric Software

Release Notes for V10.12.0

Table of Contents

1. Overview of the Release

This document provides a brief overview of the changes introduced into the Cornelis Omni-Path Express Fabric Software (OPX Fabric Software) in 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

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.

1.3. If You Need Help

Technical support for OPX products is available 24 hours a day, 365 days a year.

For additional details:

- Visit the Cornelis Support page at http://www.cornelisnetworks.com/support
- Contact Cornelis Networks Customer Support at support@cornelisnetworks.com

1.4. Product Improvements

The following improvements are included in this release:

- Support for Red Hat Enterprise Linux (RHEL) 7.9, 8.5, and 8.6 (See Table 1 "Supported Operating Systems")
- Support for Rocky Linux 8.5, and 8.6 (See Table 1 "Supported Operating Systems")
- Support for OPX Provider (See Section 1.13 "OPX Provider")
- Support for libfabric OFI version 1.16.1
- Support for Intel OneAPI 2022, Update 2
- Support for Open MPI version 4.1.4
- Support for CUDA Toolkit 11.7

- Support for MVAPICH2-2.3.7 (PSM2 only)

> **NOTE**
>
> This release includes support for a limited set of supported distros as listed above. A follow-on release will include updates for additional distros, including SLES. See Table 1 for the latest release that supports a particular distro.

1.5. Release Packages

There are two OPX Fabric Software packages:

- **BASIC** for compute, service, and login nodes

  Includes:

  – Software that installs the following packages:

    > **NOTE**
    >
    > These packages will replace or update the packages included in the OS distribution.

    - hfi1-firmware, libpsm2 (for RHEL) and libpsm2-2 (for SLES), hfi1-diagtools-sw
    - Open MPI and MVAPICH2. See Section 1.10 "MPI Libraries" for details.
    - mpitests
    - mpi-selector
    - Open Fabrics Interface (OFI) libfabric

- **OPXS** for management nodes

  Includes the BASIC package plus:

  – Fabric Manager, which allows comprehensive control of administrative functions using a mature Subnet Manager. Fabric Manager simplifies subnet, fabric, and individual component management, easing the deployment and optimization of large fabrics.

  – FastFabric Toolset, which enables rapid, error-free installation and configuration of OPX Host Software and management software tools, as well as simplified installation, configuration, validation, and optimization of HPC fabrics.

1.6. Release Compatibility

This release is backward compatible with the most recent minor release version. For example, Release 10.N is backward compatible with Release 10.N-1 and Release 10.N-1.x.

1.7. Operating Systems

This release of the OPX Fabric Software supports the operating systems listed in the following table.

> **NOTE**
>
> The 10.12.0 release contains release packages for RHEL 7.9, 8.5, and 8.6, and Rocky Linux 8.5, and 8.6 only. All other OS release packages are still available with the 10.11.1 release.

> **NOTE**
>
> RHEL 7.9 is supported through RHEL 7.8 Release 10.12.0 OPX packages.

Table 1. Supported Operating Systems

Operating System	Update	Base Kernel Version	Release
Red Hat Enterprise Linux (RHEL) 7.7 X86_64	7.7	3.10.0-1062.el7.x86_64	10.11.1
Red Hat Enterprise Linux (RHEL) 7.8 X86_64	7.8	3.10.0-1127.el7.x86_64	10.11.1
Red Hat Enterprise Linux (RHEL) 7.9 X86_64	7.9	3.10.0-1160.el7.x86_64	10.12.0
Red Hat Enterprise Linux (RHEL) 8.3 X86_64	8.3	4.18.0-240.el8.x86_64	10.11.1
Red Hat Enterprise Linux (RHEL) 8.4 X86_64	8.4	4.18.0-305.el8.x86_64	10.11.1
Red Hat Enterprise Linux (RHEL) 8.5 X86_64	8.5	4.18.0-348.el8.x86_64	10.12.0
Red Hat Enterprise Linux (RHEL) 8.6 X86_64	8.6	4.18.0-372.9.1.el8.x86_64	10.12.0
CentOS-7 (1908) X86_64 (corresponds to RHEL 7.7)	1908	3.10.0-1062.el7.x86_64	10.11.1
CentOS-7 (2003) X86_64 (corresponds to RHEL 7.8)	2003	3.10.0-1127.el7.x86_64	10.11.1
CentOS-7 (2009) X86_64 (corresponds to RHEL 7.9)	2009	3.10.0-1160.el7.x86_64	10.11.1
CentOS-8 (2011) X86_64 (corresponds to RHEL 8.3)	2011	4.18.0-240.el8.x86_64	10.11.1
CentOS-8 (2105) X86_64 (corresponds to RHEL 8.4)	2105	4.18.0-305.el8.x86_64	10.11.1
CentOS-8 (2111) X86_64 (corresponds to RHEL 8.5)	2111	4.18.0-348.el8.x86_64	10.11.1
Rocky Linux 8.4 (corresponds to RHEL 8.4)	8.4	4.18.0-305.el8.x86_64	10.11.1
Rocky Linux 8.5 (corresponds to RHEL 8.5)	8.5	4.18.0-348.el8.x86_64	10.12.0
Rocky Linux 8.6 (corresponds to RHEL 8.6)	8.6	4.18.0-372.9.1.el8.x86_64	10.12.0
Scientific Linux 7.7 X86_64	7.7	3.10.0-1062.el7.x86_64	10.11.1
Scientific Linux 7.8 X86_64	7.8	3.10.0-1127.el7.x86_64	10.11.1
Scientific Linux 7.9 X86_64	7.9	3.10.0-1160.el7.x86_64	10.11.1

Operating System	Update	Base Kernel Version	Release
SUSE Linux Enterprise Server (SLES) 12.4 X86_64	SP 4	4.12.14-94.41_default	10.11.1
SUSE Linux Enterprise Server (SLES) 12.5 X86_64	SP 5	4.12.14-120_default	10.11.1
SUSE Linux Enterprise Server (SLES) 15.2 X86_64	SP 2	5.3.18-22_default	10.11.1
SUSE Linux Enterprise Server (SLES) 15.3 X86_64	SP 3	5.3.18-57-default	10.11.1

1.8. CUDA Support

In general, PSM2 GPUDirect RDMA with CUDA is supported as shown below:

- CUDA Toolkit 11.7 is supported on RHEL 7.9, 8.5, and 8.6
- CUDA Toolkit 11.5 is supported on RHEL 7.7, 7.8, 8.3, 8.4, and SLES 15 SP2, 15 SP3
- CUDA Toolkit 10.2 is supported on SLES 12 SP4, 12 SP5

For information on compatible driver versions, refer to NVIDIA's CUDA Compatibility, "CUDA Toolkit and Compatible Driver Versions" table.

1.9. Kernel Modules for hfi1 Driver

This release of the OPX Fabric Software contains the hfi1 driver kernel modules listed in the following table.

> **NOTE**
>
> The version number is shown in bold.

Table 2. Kernel Modules for hfi1 Driver

OS	Non-CUDA/CUDA	RPM Version
RHEL 7.7	Non-CUDA	kmod-ifs-kernel-updates-3.10.0_1062.el7.x86_64-**2351**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-3.10.0_1062.el7.x86_64-**2351cuda**.x86_64.rpm
RHEL 7.8	Non-CUDA	kmod-ifs-kernel-updates-3.10.0_1127.el7.x86_64-**2351**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-3.10.0_1127.el7.x86_64-**2351cuda**.x86_64.rpm
RHEL 7.9 (built during installation)	Non-CUDA	kmod-ifs-kernel-updates-3.10.0_1160.el7.x86_64-**2351**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-3.10.0_1160.el7.x86_64-**2351cuda**.x86_64.rpm
RHEL 8.3	Non-CUDA	kmod-ifs-kernel-updates-4.18.0_240.el8.x86_64-**2351**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-4.18.0_240.el8.x86_64-**2351cuda**.x86_64.rpm

OS	Non-CUDA/ CUDA	RPM Version
RHEL 8.4	Non-CUDA	kmod-ifs-kernel-updates-4.18.0_305.el8.x86_64-**2351**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-4.18.0_305.el8.x86_64-**2351cuda**.x86_64.rpm
RHEL 8.5	Non-CUDA	kmod-ifs-kernel-updates-4.18.0-348.el8.x86_64-**2369**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-4.18.0-348.el8.x86_64-**2369cuda**.x86_64.rpm
RHEL 8.6	Non-CUDA	kmod-ifs-kernel-updates-4.18.0_372.9.1.el8.x86_64-**2369**.x86_64.rpm
	CUDA	kmod-ifs-kernel-updates-4.18.0_372.9.1.el8.x86_64-**2369cuda**.x86_64.rpm
SLES 12 SP4	Non-CUDA	ifs-kernel-updates-kmp-default-4.12.14_94.41_default_k4.12.14_94.41-**2351**.x86_64.rpm
	CUDA	ifs-kernel-updates-kmp-default-4.12.14_94.41_default_k4.12.14_94.41-**2351cuda**.x86_64.rpm
SLES 12 SP5	Non-CUDA	ifs-kernel-updates-kmp-default-4.12.14_120_default_k4.12.14_120-**2351**.x86_64.rpm
	CUDA	ifs-kernel-updates-kmp-default-4.12.14_120_default_k4.12.14_120-**2351cuda**.x86_64.rpm
SLES 15 SP2	Non-CUDA	ifs-kernel-updates-kmp-default-5.3.18-57-default_k5.3.18_22-**2351**.x86_64.rpm
	CUDA	ifs-kernel-updates-kmp-default-5.3.18-57-default_k5.3.18_22-**2351cuda**.x86_64.rpm
SLES 15 SP3	Non-CUDA	ifs-kernel-updates-kmp-default-5.3.18_22.default_k5.3.18_22-**2351**.x86_64.rpm
	CUDA	ifs-kernel-updates-kmp-default-5.3.18_22.default_k5.3.18_22-**2351cuda**.x86_64.rpm

1.10. MPI Libraries

NOTE

Cornelis recommends that you source the Intel MPI mpivars.sh or setvars.sh script when using Intel MPI with OPX and OFI. Refer to the Intel MPI Library documentation (https://software.intel.com/content/www/us/en/develop/tools/oneapi/components/mpi-library.html)

If you are not using Intel MPI, do not run this script.

1.10.1. Supported MPI Libraries

The table below lists the different MPI libraries supported by OPX Fabric Software with the corresponding version, fabric support, and compiler used. Note that the second column indicates if the MPI library is included in the OPX Fabric Software package.

NOTE

When using newer releases of Open MPI with PSM2 MTL or OFI MTL, there are new performance limitations to be aware of. The 5.0.x builds of Open MPI have removed support for sending "one sided communications" via the Open MPI MTL (Message Transport Layer) and replaced it with using the OFI BTL (Basic Transport Layer) instead.

While this yields some performance improvements in Open MPI it also consumes twice the number of hardware resources on Omni-Path (and other RDMA) hardware. This can have the effect of reducing the total number of ranks-per-node that Open MPI can support on an Omni-Path fabric.

For this reason, Cornelis recommends using the 4.0.6, 4.0.7, 4.1.1, 4.1.2, or 4.1.4 releases of Open MPI on Omni-Path fabrics. When using Open MPI 4.1.x or later, if Open MPI initializes both the MTL and BTL with both using PSM2, each rank will open two PSM2 endpoints and the maximum number of ranks per node will be restricted to 1/2 the number of user contexts on the node. The number of user contexts is typically equal to the number of physical CPU cores on the node. For example, if a node has 44 physical cores, the number of user contexts will typically be 44. If an Open MPI job initializes the PSM2 or OFI MTL and the OFI BTL using the PSM2-OFI provider underneath, the job will now be restricted to 22 PPN.

Alternately, the system administrator can update the HFI1 module settings to increase the number of contexts reserved for users by increasing the `num_user_contexts` parameter. This can be found in the `/etc/modprobe.d/hfi1.conf`, for example, but may have other performance penalties. See the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide* for more information.

- Do not specify `-mca btl` or `--mca pml ob1` if not needed.

- If the job does need BTL but does not need internode BTL, specify `--mca btl self,vader`.

- If the job needs BTL but not MTL, do not specify `--mca mtl ...` or specify `--mca mtl self`. If OFI BTL is selected with PSM2-OFI provider underneath, specify `--mca mtl self` to restrict MTL to intrarank communication only.

- If the job needs BTL and higher PPN than `num_phys_cores`, reload hfi1 with higher `num_user_contexts`, up to `num_user_contexts=160`. The hfi1 reserves some contexts for driver purposes so actual number of free user contexts will be less than 160, typically greater than 150.

- If pt2pt OSC is still present in Open MPI 4.1.x; it can run with `--mca osc pt2pt` for applications that need one-sided operations. N/A for Open MPI 5.0.x and later.

For 4.1.x versions of Open MPI, the ability to use the pt2pt for one-sided communications is still present but may no longer be selected by default. For this reason, when running with Open MPI versions 4.1.x, if pt2pt OSC is desired, we recommend using the settings:

```
--mca mtl ofi --mca osc pt2pt --mca btl self,vader
```

NOTE

As of 10.10.0, the Open MPI build in OPXS, no longer includes native verbs support (openib BTL).

Table 3. Supported MPI Libraries

MPI Implementation	Included in Basic Package?	Runs Over	Compiled With
Open MPI 4.0.6, 4.0.7, 4.1.1, 4.1.2[1], 4.1.4	Yes	PSM2, OFI, OPX provider	GCC
Open MPI 4.0.6-cuda[1]	Yes	PSM2	GCC
MVAPICH2-2.3.7	Yes	PSM2	GCC
Intel OneAPI 2021 Update 3 or 4 [2]	No	OFI	N/A
Intel OneAPI 2022 Update 2	No	OFI	N/A

NOTE:
  1. Open MPI 4.1.0 is not supported at this time.
  2. PSM2 is compiled using the Intel Compiler (ICC).

1.10.2. Compiler Versions and Distributions

The MPI libraries listed in the preceding section that are included in the release and built with PSM2 support were built using the distro default compilers.

NOTE

Refer to the *Cornelis Omni-Path Express Fabric Host Software User Guide* for setup information when using Open MPI with the SLURM PMI launcher and PSM2.

1.11. Supported Hardware

The following table lists the hardware supported in this release. The table does not include OEM-specific hardware, such as custom adapters and switches.

NOTE

The PSM2 implementation has a limit of four (4) HFIs per server.

Table 4. Supported Hardware

Hardware	Description
Cornelis Omni-Path Express Host Fabric Interface Adapter 100HFA016 (x16)	Single Port PCIe Host Fabric Interface (HFI)
Cornelis Omni-Path Express Host Fabric Interface Adapter 100HFA018 (x8)	Single Port PCIe Host Fabric Interface (HFI)
Cornelis Omni-Path Express Accelerated Open Compute Project (OCP) 3.0 Host Fabric Interface Adapter 100HFP016KS (x16)	Single Port OCP Host Fabric Interface Adapter (OCP HFA)
2nd Gen AMD EPYC Processor	Rome Zen 2 microarchitecture
3rd Gen AMD EPYC Processor	Milan Zen 3 microarchitecture
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 Processor	Skylake CPU-based servers
2nd Generation Intel Xeon Scalable Processor	Cascade Lake CPU-based servers
3rd Generation Intel Xeon Scalable Processor (i3, i5, i7)	Ice Lake CPU-based servers

1.12. Switch Firmware

The following firmware is supported for OPX Switches:

- OPX Switch Firmware 10.8.x revision (managed and externally-managed switches)
- OPX Switch Firmware 10.7.x revision (managed and externally-managed switches)

Cornelis recommends using switch firmware 10.8.4 or greater.

Refer to the *Cornelis Omni-Path Express Fabric Switches Release Notes* for more information.

1.13. OPX Provider

Cornelis is introducing a new enhanced libfabric provider, called the OPX provider.

The OPX provider is written to take full advantage of the libfabric acceleration features while running over existing and future OPX hardware.

The OPX provider started as a fork of the libfabric Blue Gene Q (BGQ) provider, having all of the hardware-specific parts re-written for OPX. Therefore, the OPX provider inherits several desirable characteristics of embedded systems design constraints. Instruction counts and cache line footprint of most HPC operations show the OPX provider being lighter weight than PSM2 to the host software stack, which provides better overall performance.

The OPX provider provides excellent latency characteristics and message rates at smaller message sizes (under 16K message length). Bulk transfer via SDMA is present in the OPX provider giving some added performance for large messages. However, RDMA support is not

implemented in the OPX provider at this time, so performance on large messages is still better with the PSM2 provider.

For further information on the OPX provider, refer to the *Cornelis Omni-Path Express OPX_Provider Installation and Setup Application Note*.

1.14. Installation Requirements

This section provides installation requirements for this release.

1.14.1. Best Practices

Note the following Cornelis recommendations:

- Update to the latest versions of OPX firmware and software to obtain the most recent functional and security updates.

- To improve security:

  – administrators should log out users and disable multi-user logins prior to performing provisioning and similar tasks.

  – update the default HTTPS certificate (refer to the *Cornelis Omni-Path Express Fabric Switches GUI User Guide*, "Updating the Certificate" for details).

- To improve security, Cornelis recommends configuring the `MgmtAllowed` setting and consider limiting access to port configuration changes by limiting access to Userspace Management Datagrams (UMADs). Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide*, "About User Queries Settings" for more information.

1.14.2. Software and Firmware Requirements

Section 1.7 "Operating Systems" lists the operating systems supported by this release. For the required packages, refer to Section 1.14.3 "OS RPMs Installation Prerequisites".

1.14.3. OS RPMs Installation Prerequisites

Ensure that the following requirements are met before installing the software.

- Along with normal OS installation options, additional OS RPMs must be installed before you can install the OPX Fabric Software.

- Refer to the applicable section below to verify that all required RPMs are installed for the specific version of your OS distribution.

- Depending on the packages you choose, there may be additional prerequisites. For additional information, refer to the Release Notes for your specific release and installation type.

1.14.3.1. RHEL OS RPMs

The tables below list the RPMs for each supported RHEL release.

> **NOTE**
> - Some RHEL RPMs are available in a Server-Optional repository. Please install the RHEL Server-Optional version from Red Hat, which contains additional, required development packages.
> - rdma-ndd is part of rdma-core.

Table 5. RHEL 7.7 , 7.8, and 7.9 Distribution RPMs

IB/OPA Centric	System Centric			Other/Generally Installed	Build Requirements
ibacm	atlas	libstdc++-devel		bash	bison
infinipath-psm	bc	ncurses-libs		irqbalance	expat-devel
libibumad	coreutils	numactl-libs		kernel	flex
libibverbs	createrepo	openssl		kmod	libnl3-devel
libnl3	expat	openssl-devel		libgcc	libpfm
librdmacm	expect	openssl-libs		perl	libuuid-devel
opensm-libs	gcc-gfortran	pciutils		perl-Getopt-Long	ncurses-devel
perftest	glibc	redhat-rpm-config		perl-PathTools	numactl-devel
qperf	kernel-devel	rpm-build		perl-Socket	opensm-libs
rdma-core	libatomic	sysfsutils		pkgconfig	openssl-devel (1.0.1 or higher)
rdma-core-devel	libgfortran	tcl		python	tcl-devel
	libgomp	tcsh		systemd	valgrind-devel
	libquadmath	zlib		systemd-libs	
	libstdc++				

Table 6. RHEL 8.3, 8.4, 8.5 and 8.6 Distribution RPMs

IB/OPA Centric	System Centric			Other/Generally Installed	Build Requirements
ibacm	atlas	libstdc++-devel		bash	autoconf
libibumad	bc	ncurses-compat-libs		irqbalance	automake
libibverbs	coreutils	ncurses-libs		kernel	bison
librdmacm	createrepo	numactl-libs		kernel-modules-extra	elfutils-libelf-devel

IB/OPA Centric	System Centric		Other/Generally Installed	Build Requirements
opensm-libs	expat	openssl	kmod	expat-devel
perftest	expect	openssl-devel	libgcc	flex
qperf	gcc-gfortran	openssl-libs	perl	gcc-c++.x86_64
rdma-core	glibc	pciutils	perl-Getopt-Long	kernel-abi-whitelists
rdma-core-devel	kernel-devel	redhat-rpm-config	perl-Socket	kernel-rpm-macros
	libatomic	rpm-build	pkgconf	libnl3-devel
	libgfortran	sysfsutils	python2	libpfm
	libgomp	tcl	systemd	libtool
	libquadmath	tcsh		libuuid-devel
	libstdc++	zlib		ncurses-devel
				numactl-devel
				opensm-libs
				openssl-devel (1.1.1 or higher)
				tcl-devel

1.14.3.2. SLES OS RPMs

The tables below list the RPMs for each supported SLES release.

NOTE

Some SLES RPMs are available in SLES Software Development Kit (SDK). Please install the SLESSLES SDK iso from SUSE, which contains additional, required development packages.

Table 7. SLES 12.4 Distribution RPMs

IB/OPA Centric	System Centric		Other/Generally Installed	Build Requirements
ibacm	bc	libnuma1	bash	bison
libibcm1	createrepo	libopenssl1_0_0	glibc	flex
libibmad5	expect	libopenssl-devel	grep	libexpat-devel
libibumad3	gcc-fortran	libquadmath0	irqbalance	libnuma-devel
libibverbs1	kernel-devel	libudev-devel	kmod	libopenssl-devel (1.0.1 or higher)
libpsm_infinipath1	kernel-syms	libz1	libedit0	libuuid-devel

IB/OPA Centric	System Centric		Other/Generally Installed	Build Requirements
librdmacm1	libatomic1	openssl	libgcc_s1	ncurses-devel
mpi-selector	libexpat1	rpm-build	libstdc++6	opensm-libs3
opensm-devel	libgfortran3	tcl	perl	tcl-devel
opensm-libs3	libgomp1	tcsh	perl-base	valgrind-devel
perftest	libncurses5		pkg-config	
qperf			python-base	
rdma-core			systemd	
rdma-core-devel			udev	
rdma-ndd				

Table 8. SLES 12.5 Distribution RPMs

IB/OPA Centric	System Centric		Other/Generally Installed	Build Requirements
ibacm	bc	libnuma1	bash	bison
libibmad5	createrepo	libopenssl1_0_0	glibc	flex
libibumad3	expect	libopenssl-devel	grep	libexpat-devel
libibverbs1	gcc-fortran	libquadmath0	irqbalance	libnuma-devel
libpsm_infinipath1	kernel-devel	libudev-devel	kmod	libopenssl-devel (1.0.1 or higher)
librdmacm1	kernel-syms	libz1	libedit0	libuuid-devel
mpi-selector	libatomic1	openssl	libgcc_s1	ncurses-devel
opensm-devel	libexpat1	rpm-build	libstdc++6	opensm-libs3
opensm-libs3	libgfortran3	tcl	perl	tcl-devel
perftest	libgomp1	tcsh	perl-base	valgrind-devel
qperf	libncurses5		pkg-config	
rdma-core			python-base	
rdma-core-devel			systemd	
rdma-ndd			udev	

Table 9. SLES 15.2, and 15.3 Distribution RPMs

IB/OPX Centric	System Centric		Other/Generally Installed	Build Requirements
ibacm	bc	libnuma1	bash	bison
libibmad5	coreutils	libopenssl1_1	glibc	flex
libibumad3	createrepo_c	libopenssl-devel	grep	kernel-devel

IB/OPX Centric	System Centric		Other/Generally Installed	Build Requirements
libibverbs1	expect	libosmcomp3	irqbalance	libexpat-devel
libnl3-200	gcc-fortran	libquadmath0	kmod	libnuma-devel
libpsm_infinipath1	kernel-devel	libudev-devel	libedit0	libopenssl-devel (1.0.1 or higher)
libquadmath0	kernel-syms	libz1	libgcc_s1	libuuid-devel
librdmacm1	libatomic1	openssl	libncurses5	ncurses-devel
libverbs1	libexpat1	rpm-build	libstdc++6	opensm-libs3
mpi-selector	libgfortran4	tcl	perl	tcl-devel
opensm-devel	libgomp1	tcsh	perl-base	valgrind-devel
opensm-libs3	libncurses6		pkg-config	
perftest			python-base	
qperf			systemd	
rdma-core			udev	
rdma-core-devel				
rdma-ndd				

1.14.3.3. Example

The example below shows the installation of a group of OS RPMs on a SLES OS server. Not all required OS RPMs are included in this example and some OS RPMs in this example might not be needed.

```
# yum install libibmad libibverbs librdmacm qperf perftest
rdma infinipath-psm expat libstdc++-devel gcc-gfortran atlas tcl expect
tcsh sysfsutils pciutils bc libibumad libibumad-devel libibumad
libibumad-devel libibverbs-devel libibmaddevel librdmacm-devel ibacm-devel
openssl-devel libuuid-devel expatdevel infinipath-psm-devel valgrind-devel
libgnome libibverbs opensm-libs ncurses-devel hwloc hwloc-gui
```

Cornelis recommends that you build your own list of OS RPMs for installation.

1.14.4. Installation Instructions

There are two OPX Fabric Software packages:

- `CornelisOPX-OPXS.<distro>-x86_64.<version>.tgz` for the management node.
- `CornelisOPX-Basic.<distro>-x86_64.<version>.tgz` for compute, service, and login nodes.

The packages in the tgz file are RPMs. Installing individual RPMs is not supported in this release.

> **(!) IMPORTANT**
>
> If you want to install the CUDA versions of the OPXS software using the INSTALL command, you must use the `-G` option.

Refer to the *Cornelis Omni-Path Express Fabric Software Installation Guide* for related software requirements and complete installation procedures. Refer to the *Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide* for related firmware requirements.

1.15. Product Constraints

- Power class 2 AOCs are supported. You must use 10.5 (or newer) host software and 1.5 (or newer) UEFI for proper operation. Integrated HFI (-F) requires a specific BIOS level to support power class 2 AOC. Contact your BIOS vendor for more information.

- Some newer distros include an updated version of IPoIB which has an incompatibility with Fabric Managers prior to this release. This could result in the inability to establish IPoIB connections. Ensure that OS versions are consistent across the fabric.

1.16. Product Limitations

This release has the following product limitations:

- The embedded version of the Fabric Manager supports a maximum of 100 HFI ports involving less than 20 Switch ASICs. Calculate the number of Switch ASICs in your fabric as follows:

  – One ASIC per Cornelis Omni-Path Express Edge Switch 100 Series

  – Two ASICs per OPX Chassis Leaf module

  – Two ASICs per OPX Chassis Spine module

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

  To disable PA failover, edit the `/etc/opa-fm/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.

1.17. Accelerated RDMA Information

Accelerated RDMA is a Verbs protocol extension to improve the performance of RDMA write and RDMA read operations on OPX hardware.

This extension improves the efficiency of large message transfers to provide performance benefits for storage protocols and other Verbs-based protocols. The benefits include

increased achievable bandwidth with reduced CPU utilization. The Accelerated (or Token ID (TID)) RDMA protocol accelerates the OpenFabrics Alliance (OFA) Verbs API with no changes required to API consumers. The acceleration technique is performed by the host driver and the application running over the OFA Verbs API does not need to make any code changes.

Accelerated RDMA is off by default.

To enable it, add `cap_mask=0x4c09a09cbba` to the `/etc/modprobe.d/hfi1.conf` file. Instructions on how to do this are in the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, "Setting HFI1 Driver Parameters" section.

1.18. hfi1_IRQ_Balancing

`irqbalance` daemon and udev can be used to ensure optimal HFI IRQ => CPU core mapping. Optimal HFI IRQ => CPU core mapping may yield better performance. There are two parts needed for hfi1 IRQ balancing to work:

- `irqbalance` policy script that is called for each IRQ belonging to an HFI. Tells irqbalance which CPU core(s) a given HFI IRQ should be processed on.

- An `udev.rules` file that tells `irqbalance` to rescan IRQs when an HFI appears or disappears from the system bus.

OPXS provides the `irqbalance` script but not the `udev.rules` file.

Here is an example `udev.rules` file to trigger irqbalance when an HFI appears or disappears:

```
&amp;lt;preformatted text&amp;gt;
ACTION=="add|change|remove",
SUBSYSTEM=="hfi1_user",
RUN+="/usr/bin/killall -SIGHUP
irqbalance"
&amp;lt;/preformatted text&amp;gt;
```

Important notes about the above rule:

- Keys off of `hfi1_user` virtual device instead of 'hfi1' module load. This rule will be triggered for each HFI added rather than just the initial hfi1 module load/unload.

- Necessary to key off of `hfi1_user` since this is created at about the same time the IRQ affinity hints needed by our policy script are created. Keying off of 'hfi1' module load runs the risk of a race between when the affinity hints are populated and when irqbalance rescan is triggered.

- Sends `SIGHUP` to the irqbalance daemon rather than doing `systemctl restart irqbalance`. `SIGHUP` was the recommended method to rescan from our distro's documentation. `systemctl restart` did not work correctly in our testing.

See the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide* for information about `irqbalance` and custom HFI IRQ affinity.

2. Issues

This section lists the resolved and open issues in the OPX Fabric Software.

2.1. Resolved Issues

2.1.1. Issues Resolved in this Release

The following table lists issues that are resolved in this release.

Table 10. Issues Resolved in this Release

ID	Description	Resolved in Release
STL-63687	PSM2 prior to version 11.2.230 has an issue where silent data errors will occur when all four of the following conditions are met: • OPA Congestion Control Architecture (CCA) is enabled in the FM configuration. • The application receives into an unaligned receive buffer, a buffer whose start address is not a 4B multiple. • The unaligned receive is for a large receive, one where PSM2 will use the TID receive mechanism. • One or more packets for the unaligned TID receive encounters congestion and is marked with the forward explicit congestion notification (FECN) bit. When these conditions are met, PSM2 is responsible for copying the payload to the destination memory rather than the TID hardware. But the PSM2 code did not take into account the unaligned buffer start. As a result, the data was copied to the wrong location in the receive buffer. This PSM2 issue has been resolved starting with PSM2 11.2.230 [this release] to correctly account for the unaligned start.	10.12.0
STL-63901	When the Fabric Manager SM has discovered and is configuring a switch, the SM must initialize the SCVL_nt Table of each neighboring port. This is done via a MAD request to the neighboring port. Currently, when the SM encounters a problem with that MAD request, the SM quarantines the switch port-0 (instead of quarantining the neighboring port that failed the request). This eventually causes the SM to report the disappearance of all end-nodes connected to the quarantined switch.	10.12.0

ID	Description	Resolved in Release

STL-63732	When IPoIB multicast (MC) join, requests were issued to an existing IPoIB MC group, via host port(s) configured to be limited-member(s) of a Virtual Fabric (VF), the FM/SA would not permit VF limited-members to join the MC group and would report an error message. This action taken by the FM/SA was incorrect. The IPoIB Architecture spec (RFC4392) states, that both MC full-members and MC limited-members may join an 'existing' MC group. The following is an example of the error message reported by the FM/SA:
	
	`sa_McMemberRecord_Set: Failing group join request for PortGUID`
	`0x001175010170c3fc`
	`from opahsx74 hfi1_0 Port 1, PortGUID 0x001175010170c3fc, LID 0x00000001`
	`because`
	`it is not a full member of virtual fabric with MGid`
	`0xff12401b81010000:00000000ffffffff`
		10.12.0

2.1.2. Issues Resolved in Prior Releases

The following table lists issues that were resolved in prior releases.

Table 11. Issues Resolved in Prior Releases

ID	Description	Resolved in Release
STL-56557	The primary IPoIB network device associated with any RDMA device may fail to join certain multicast groups. This can prevent IPv6 neighbor discovery and possibly other network ULPs from working correctly. Note that the IPv4 broadcast group is not affected as the IPoIB network device handles joining that multicast group directly. This problem does not affect IPoIB child network devices.	10.11.1
STL-58332	When upgrading from RHEL 7.5 to a newer version, if a value is set for `ifs_sel_mode` in `hfi1.conf`, the HFI driver will not load.	10.11.1

STL-60129	When attempting to launch large scale jobs, you may see messages such as:
	
	`Received eager message(s) ptype=0x1 opcode=0xcc from`
	`an unknown process (err=49)`
	
	PSM2 connections are timing out and OFI is not honoring the error returned by PSM2 and later attempting to send messages over the failed connection.
	
	OFI overrides the PSM2_CONNECT_TIMEOUT variable to 5 seconds (`.conn_timeout = 5` in `psmx2_init.c`), whereas the PSM2 code specifies 0 (no timeout).
		10.11.1

STL-61241 STL-61299	On some platforms, the hfi1 device may not show up in BIOS/UEFI boot menus and may not be available as a PXE boot device. This is caused by the platform not loading the UEFI driver for the hfi1 adapter.	10.11.1
STL-61605	OMPI 4.1.1 fails to build with PSM2 support when PSM2 is installed in a non-default location.	10.11.1

ID	Description	Resolved in Release
STL-62722	When freeing dma memory associated with eager buffers, the test for "is allocated" incorrectly checks the dma address. The dma allocation is then leaked.	10.11.1
STL-62926	The OPXS installation fails on newer distro kernels due to the prerequisite package kernel-abi-whitelists package being replaced by kernel-abi-stablelists.	10.11.1
STL-62933	A kernel panic occurred when the AIP send queue size was increased from its default value.	10.11.1
STL-63090	Under rare circumstances when a job is killed while a rank is being brought up a panic could occur.	10.11.1
STL-61143	Certain BIOS versions or settings may cause the BIOS to report a numa_node for hfi1 as `NO_NUMA_NODE`, resulting in a probe time panic.	10.11.0.1
STL-61504 STL-61602	The addition of an upstream patch (633d61021298 "RDMA/ipoib: Remove racy Subnet Manager sendonly join checks") to 3.10.0-1160.25.1 caused issues with both the INSTALL RPM build and with the FM handling of a new join type.	10.11.0.1
STL-61604	Certain RedHat Z-stream kernel releases fail to build due to newly incorporated changes. RedHat backported a patch that the OPXS driver needs to handle.	10.11.0.1
STL-61689	A build issue occurs with SLES security kernel update and OPXS driver due to redefined components in the common_compat.h file.	10.11.0.1
STL-60901	Child process can access the parent memory management through the hfi dev file handle.	10.11.0
STL-61152	atomic_fetch_add_unless() was added to the Kernel API in 4.18.0-193.28.1. OPX OPXS driver cannot be compiled and installed on RHEL 8.2 and its equivalent CentOS-8 (2004).	10.11.0
STL-59934	Customers experience errors similar to `hfi1_0: Send Context 8(151) Error: WriteOverflow` in dmesg or console logs.	10.11.0

2.2. Open Issues

The following table lists the open issues for this release.

Table 12. Open Issues

ID	Description	Workaround
134819 (STL-49733)	In KNL-F EFI shell, the command `ifconfig –l` does not correctly display the IP address after being assigned via DHCP.	Launch a newer version of the EFI shell from the embedded shell.
136822 (STL-49728)	The UEFI driver contained in the server BIOS must be executed for proper support of Active Optical Cables (AOC) in an integrated HFI environment. Some 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 the use of Legacy BIOS boot mode if your platform does not execute the HFI driver in that mode. Avoid BIOS settings or other configuration settings that do not execute the HFI driver during boot.

ID	Description	Workaround
139613 (STL-49732)	The Subsystem Vendor and Subsystem Device ID in the PCI configuration space of OPX 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.
139995 (STL-49724)	When installing OPX Fabric Software on a clean system (with no prior versions installed), the following error message may be displayed: `cat: //etc/opa/version_delta: No such file or directory`	This message can be safely ignored. The installer is looking for an OPXS version file before it has been created on the system. The installation is not impacted.
143174 (STL-47003)	Due to a SLES 15 kernel setting, hfidiags cannot work while the HFI driver is loaded. The tool and driver are mutually exclusive.	Boot the kernel with `iomem=relaxed`. This turns off the resource exclusivity check.
STL-48921	On some Distributions, if you attempt to unload or restart the hfi1 driver while the ib_ipoib driver is loaded, it may fail with a message similar to this: `modprobe: FATAL: Module hfi1 is in use`	To avoid this issue, unload the ib_ipoib driver before unloading or restarting the hfi1 driver.
STL-49313	On OS versions (RHEL 7.7 and newer and SLES 15, SLES 12.4 and newer) where it is available, configuring IPoIB module parameter `ipoib_enhanced=0` disables AIP on same node.	To run with AIP, ensure enhanced IPoIB is enabled (`ipoib_enhanced=1`).
STL-57127	Restarting the irqbalance service while the hfi1 driver is loaded may cause improper or inconsistent distribution of interrupts, which may result in low network transfer performance.	Irqbalance must be stopped before starting the hfi1 module. Perform the following command sequence: ```systemctl stop irqbalance modprobe -r ib_ipoib modprobe -r hfi1 modprobe hfi1 modprobe ib_ipoib systemctl start irqbalance``` **NOTE:** Additional steps (like stopping Lustre) may be required on more complicated fabrics.
STL-59413	Using PSM2 with CUDA enabled on top of a non-CUDA enabled hfi1 driver may result in invalid behavior.	Ensure the PSM2 CUDA library is used only in conjunction with the hfi1 driver with GPUDirect/CUDA support. **NOTE:** If the gpu-direct enabled driver is present, the string `gpu-direct` will be returned in the Driver version output of `hfi1_control -i`.

ID	Description	Workaround
STL-59663	Certain conditions result in hfi1 driver interrupts not being distributed correctly, even with irqbalance running. This may result in a large run to run performance variation or low overall performance.	Refer to the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, section for help with determining which cores are running the hfi1 interrupts. If interrupts are arbitrarily distributed or significantly reusing select cpu cores, refer to "Manually Changing IRQ Affinity" section. Contact Cornelis Networks Customer Support if you need further guidance.
STL-60932	On 3rd Generation Intel Xeon Scalable Processors, IPoFabric small message latency may measure higher than expected depending on the tunings being used.	Cornelis recommends that you use the Tuned latency-performance profile to significantly reduce IPoFabric small message latency. Tuned is a utility (for example, tuned-2.10.0-15.el8.noarch) that allows for dynamic and adaptive tuning of CPU behavior. Note that using Tuned latency-performance profile may have a negative performance impact to other bandwidth/message rate measurements.
| STL-61176 | Running perftest, the following text may be displayed:<br><br>```
Couldn't post send: qp 0 scnt=189
Failed to complete run_iter_bw function
successfully
```<br><br>**NOTE:** This failure is only seen on perftest and does not seem to impact any other upper layer protocols (ULPs). However, it is worse on RHEL 8.3. | Disable TID RDMA. |
STL-61215	In rare instances, the host may boot and the HFI device is not visible on the PCI bus, and is therefore not usable.	Reboot/power cycle the host.
STL-61244	Servers with 2nd Gen AMD EPYC Processor may experience an unexpected server reboot when running under load with a high number of processes per node.	Contact Cornelis Networks Customer Support for more information.
STL-61256	On 2nd Gen AMD EPYC Processors, AIP interrupts do not map optimally in all situations, especially when certain BIOS settings (such as NPS) are modified. This can impact IPoFabric performance.	Manually remap the AIP interrupts to give the best performance. Refer to the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, "Driver IRQ Affinity Assignments" section for more details. Or, contact Cornelis Networks Customer Support if you need further guidance."

ID	Description	Workaround
STL-61329	On RHEL 7.x servers with two HFIs installed (dual-rail), unbinding both HFIs from the hfi1 driver may result in a kernel warning trace.	None.
STL-61331	When installing on SLES 12.x from a yum/zypper repository, the installation may fail due to missing dependencies.	Do not use `-y` when performing the yum/zypper installation and select the option to ignore the dependencies.
STL-61527	Bidirectional host to host workloads using a GPU OPXS installation do not achieve maximum performance.	Use a non-GPU OPXS install for host-host workloads, or set `PSM2_CUDA=0`, `PSM2_GPUDIRECT=0` in the job environment for host-only workloads.
STL-61590	Open MPI (OMPI) has changed the support for one-sided operations since OMPI 4.0.5. Because of this change, OMPI is now launching both PSM2 MTL and OFI Byte Transfer Layer (BTL), which consumes more contexts than it actually needs. This can result in runtime warnings/errors on some high PPN jobs.	Run -mca btl self,vader. This string explicitly tells OMPI to use -mca btl self,vader as the BTL instead of OFI and prevents OMPI from consuming more contexts than it actually needs. **Note:** Applications that require BTL to communicate may be adversely affected by this change. These are limited to a subset of MPI applications. Intel MPI benchmarks (IMB) will not be impacted.
STL-62119	Issues running the allreduce IMB benchmark using older versions of Intel MPI (ie: 2020.4) on AMD Milan systems have been observed.	Upgrade to use a more recent version of IMPI and IMB if this issue is seen.
STL-62604 STL-62605	Running certain IMB benchmarks using the the IMB source code packaged with oneAPI 2021.3 (or earlier) can result in benchmark crashes.	Upgrade to use a more recent version of IMB or upgrade to oneAPI 2021.4
STL-62613	The capmask for Accelerated IP (AIP) has changed starting in OPXS 10.11.1.0. This can cause AIP to be disabled if you are using the old cap mask referenced in the performance tuning guide, which is used to enable Accelerated RDMA.	The correct capmask is to use is now 0x4c09a09cbba.
STL-63231	For applications requiring tagged messaging where messages are matched out of order, OPX tag matching may underperform by an order of magnitude or more due to serialization. OPX searches tag matching lists serially in the order the receive was posted or an unexpected message was received, which can be very slow.	Use internal MPI tag matching where available. Some caveats apply. Contact Cornelis Networks Customer Support for details.
STL-63374	False credit loops may be reported by `opareport -o validatecreditloops` on a fabric configured in non-fat tree topology.	None, `opareport -o validatecreditloops` is only supported on fat tree topologies.

2.2.1. Third Party Open Issues

The following table lists the third party open issues for this release.

Table 13. Third Party Open Issues

ID	Description	Workaround

129563 (STL-47095)	Memory allocation errors with MVAPICH2-2.1/Verbs.	**NOTE:** To avoid this issue, use MPIs over PSM2.
		
		If you are using MPIs over verbs, the following workaround is required:
		
		• When running MVAPICH2 jobs with a large number of ranks (for example, > 36 ranks but ≤ 72 ranks), you must set the following parameters in `/etc/security/limits.conf`:
		
		  – hard memlock unlimited
		
		  – soft memlock unlimited
		
		• Increase the lkey_table_size:LKEY table size in bits (2^n, where $1 \le n \le 23$) from its default of 16 to 17. For instructions on setting module parameters, refer to the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, "HFI1 Driver Module Parameters" chapter.

141273 (STL-46935)	The in-distro version of perftests has bugs.	Use the upstream version of perftest from https://github.com/linux-rdma/perftest.
STL-47571	When trying to run an MPI/PSM job with more MPI ranks than CPU cores (oversubscribing), the job may fail with the following error message: `hfi_userinit: assign_context command failed: Device or resource busy PSM2 can't open hfi unit: -1 (err=23)`	Set `PSM2_MULTI_EP=0` (user environment variable) before or during job launch. For details, see the *Cornelis Performance Scaled Messaging 2 (PSM2) Programmer's Guide*.
STL-57040	If Intel MPI 2019, Update 2 and OpenMPI are installed on the same cluster, OpenMPI jobs could link to the Intel MPI internal libfabric library after running Intel MPI mpivars.sh.	Before running an OpenMPI job, perform the following to reset the libfabric environment variables that had been previously set by Intel MPI. `source <installdir>/ intel64/bin/mpivars.sh -ofi_internal=0`

ID	Description	Workaround
STL-59449	Occasionally, an unload of the hfi1 driver will produce a message similar to the following: `WARNING: CPU: 15 PID: 119002 at ../kernel/workqueue.c:4091 destroy_workqueue+0x61/0x240()` This message will be followed by a call trace showing the hfi1 unload. This issue is related to memory reclaim activity just before the unload.	None. **NOTE:** This is a bug in the Linux kernel. Cornelis Networks is working with the Linux community to resolve this issue.
STL-59919	Creating a ram disk using `dmsetup` on RHEL 7.7 will cause a kernel panic. The creation sequence is: 1. `modprobe brd rd_nr=1 rd_size=33554432 max_part=1` 2. `size=$(blockdev --getsize /dev/ram0)` 3. **`dmsetup`** `create mdevice0 --table "0 ${size} linear /dev/ram0 0"`	None.
STL-60633	When you install `gpfs.callhome-ecc-client-5.0.5-0` under SLES using zypper, you may receive an error code if there are no existing credentials.	Ignore the error code. **NOTE:** To avoid this error, you can touch an empty file `/usr/lpp/mmfs/bin/ecc/ECCBase/com.ibm.ws.webservices.thinclient_7.0.0.jar` to force zypper to exit with a success code. For example, run the following command before you install `gpfs.callhome-ecc-client-5.0.5-0`: `touch /usr/lpp/mmfs/bin/ecc/ECCBase/com.ibm.ws.webservices.thinclient_7.0.0.jar`
STL-61762	osu_mbw_mr can report bandwidth greater than line-rate in OSU micro benchmarks version 4.1 and higher.	None.

ID	Description	Workaround
LU-14733 (STL-61911)	The lustre Lnet layer can emit the following message: `dk.opal63.llnl.gov.7:00000001:00020000:43.0:1622598261.714620:0:129525:0:(brw_test.c:415:brw_bulk_ready()) BRW bulk READ failed for RPC from 12345-192.168.128.126@o2ib18: -103` The issue happens when an FRMR invaldation operation is failed in driver. RHEL 8.4 is exposed to the issue due to the release removing the older FMR kernel registration scheme. Lustre is sending an old previously used invalidate.	Use Lustre client release 2.12.8 or Build Lustre from source since the patches are in the Lustre source
STL-63861	Large size IPv6 pings (greater than 15,975) may experience packet loss on SLES 15.4 and RHEL 9.0.	Contact Cornelis Networks Customer Support or Linux distro provider for the latest information on the issue.

3. Related Information

3.1. Omni-Path Express Rebranding

The rebranding changes included in this release update the names of the installation packages for the management and compute nodes.

The naming convention of "Intel Fabric Suite (IFS)" package, installed for fabric management/head nodes is repnow Cornelis Omni-Path Express Fabric Suite (OPXS) to align with our rebranding efforts as Cornelis Networks.

The Basic package naming convention is retained for non-management nodes/compute nodes.

The package names are changed as follows:

Table 14. Package Name Rebranding

Old Package Name	New Package Name
IntelOPA-Basic.<distro>-x86_64.<version>.tgz	CornelisOPX-Basic.<distro>-x86_64.<version>.tgz
IntelOPA-IFS.<distro>-x86_64.<version>.tgz	CornelisOPX-OPXS.<distro>-x86_64.<version>.tgz

The subdirectory structure in the OPXS and Basic packages also replace references to "IntelOPA" with "CornelisOPX." The subdirectory names are changed as follows:

Table 15. Subdirectory Structure Rebranding

IntelOPA-OFA_Delta	CornelisOPX-OFA_DELTA
IntelOPA-Tools	CornelisOPX-Tools[1]
IntelOPA-Tools-FF	CornelisOPX-Tools-FF[2]
IntelOPA-FM	CornelisOPX-FM[2]

Note:
1. Basic only
2. OPXS only

3.2. Documentation Library

Go to the Cornelis Customer Center to download the publications from the Release Library.

Use the tasks listed in this table to find the corresponding document.

Task	Document Title	Description
Using the OPX documentation set	*Cornelis Omni-Path Express Fabric Quick Start Guide*	A roadmap to Cornelis' comprehensive library of publications describing all aspects of the product family. This document outlines the basic steps for installing your OPX cluster and ensuring it is operational.

Task	Document Title	Description
Setting up an OPX cluster	*Cornelis Omni-Path Express Fabric Setup Guide*	Provides a high-level overview of the steps required to stage a customer-based installation of the OPX Fabric. Procedures and key reference documents, such as OPX user and installation guides, are provided to clarify the process. Additional commands and best known methods are defined to facilitate the installation process and troubleshooting.
Installing hardware	*Cornelis Omni-Path Express Fabric Switches Hardware Installation Guide*	Describes the hardware installation and initial configuration tasks for OPX Director Class Switches and OPX Edge Switches.
	Cornelis Omni-Path Express Host Fabric Interface Installation Guide	Contains instructions for installing the OPX HFI in an OPX cluster.
	Cornelis Omni-Path Express Gateway Installation and Setup Guide	Describes the hardware installation and set up tasks for the OPX Gateways.
Installing Host Software Installing HFI firmware Installing switch firmware (externally-managed switches)	*Cornelis Omni-Path Express Fabric Software Installation Guide*	Describes using a Text-based User Interface (TUI) to guide you through the installation process. You have the option of using command line interface (CLI) commands to perform the installation or install using the Linux distribution software.
Managing a switch using Chassis Viewer GUI Installing switch firmware (managed switches)	*Cornelis Omni-Path Express Fabric Switches GUI User Guide*	Describes the graphical user interface (GUI) of the OPX Chassis Viewer GUI. This document provides task-oriented procedures for configuring and managing the OPX Switch family. Help: GUI embedded help files
Managing a switch using the CLI Installing switch firmware (managed switches)	*Cornelis Omni-Path Express Fabric Switches Command Line Interface Reference Guide*	Describes the CLI task information for the OPX Switch family. Help: -help for each CLI
Managing a fabric using FastFabric	*Cornelis Omni-Path Express Fabric Suite FastFabric User Guide*	Provides instructions for using the set of fabric management tools designed to simplify and optimize common fabric management tasks. The management tools consist of TUI menus and CLI commands. Help: -help and man pages for each CLI. Also, all host CLI commands can be accessed as console help in the Fabric Manager GUI.

Task	Document Title	Description
Managing a fabric using Fabric Manager	*Cornelis Omni-Path Express Fabric Suite Fabric Manager User Guide*	The Fabric Manager uses a well-defined management protocol to communicate with management agents in every OPX HFI and switch. Through these interfaces the Fabric Manager can discover, configure, and monitor the fabric.
	Cornelis Omni-Path Express Fabric Suite Fabric Manager GUI User Guide	Provides an intuitive, scalable dashboard and set of analysis tools for graphically monitoring fabric status and configuration. This document is a user-friendly alternative to traditional command-line tools for day-to-day monitoring of fabric health. Help: Fabric Manager GUI embedded help files
Configuring and administering HFI and IPoIB driver Running MPI applications on OPX	*Cornelis Omni-Path Express Fabric Host Software User Guide*	Describes how to set up and administer the OPX HFI after the software has been installed. This document is for cluster administrators and Message-Passing Interface (MPI) application programmers.
Writing and running middleware that uses OPX software	*Cornelis Performance Scaled Messaging 2 (PSM2) Programmer's Guide*	Provides a reference for programmers working with the PSM2 Application Programming Interface (API). The Performance Scaled Messaging 2 API (PSM2 API) is a low-level user-level communications interface.
	Cornelis Omni-Path Express OPX_Provider Installation and Setup Application Note	Provides a reference for programmers working with the OPX provider.
Optimizing system performance	*Cornelis Omni-Path Express Fabric Performance Tuning User Guide*	Describes BIOS settings and parameters that have been shown to ensure best performance, or make performance more consistent, on the OPX Architecture. If you are interested in benchmarking the performance of your system, these tips may help you obtain better performance.
Designing an IP or LNet router on OPX	*Cornelis Omni-Path Express IP and LNet Router Design Guide*	Describes how to install, configure, and administer an IPoIB router solution (Linux IP or LNet) for inter-operating between an OPX Fabric and a legacy InfiniBand fabric.
Building Containers for OPX Fabrics	*Building Containers for Cornelis Omni-Path Express Fabrics using Docker and Singularity Application Note*	Provides basic information for building and running Docker and Singularity containers on Linux-based computer platforms that incorporate OPX networking technology.

Task	Document Title	Description
Writing management applications that interface with OPX	*Cornelis Omni-Path Express Management API Programmer's Guide*	Contains a reference for programmers working with the OPX Architecture Management (OPAMGT) Application Programming Interface (API). The OPAMGT API is a C-API permitting in-band and out-of-band queries of the FM's Subnet Administrator and Performance Administrator.
Using NVM over Fabrics on OPX	*Configuring Non-Volatile Memory Express (NVMe) over Fabrics on Cornelis Omni-Path Express Application Note*	Describes how to implement a simple OPX Architecture-based point-to-point configuration with one target and one host server.
Learning about new release features, open issues, and resolved issues for a particular release	*Cornelis Omni-Path Express Fabric Software Release Notes*	
	Cornelis Omni-Path Express Fabric Manager GUI Software Release Notes	
	Cornelis Omni-Path Express Fabric Switches Release Notes (includes managed and externally-managed switches)	
	Cornelis Omni-Path Express Fabric Unified Extensible Firmware Interface (UEFI) Release Notes	
	Cornelis Omni-Path Express Fabric Thermal Management Microchip (TMM) Release Notes	
	Cornelis Omni-Path Express Fabric Firmware Tools Release Notes	

3.2.1. How to Search the Cornelis Omni-Path Express Documentation Set

Many PDF readers, such as Adobe Reader and Foxit Reader, allow you to search across multiple PDFs in a folder.

Follow these steps:

1. Download and unzip all the publications into a single folder.
2. Open Acrobat Reader and use **CTRL-SHIFT-F** to open the Advanced Search window.
3. Select **All PDF documents in...**
4. Select **Browse for Location** in the dropdown menu and navigate to the folder containing the PDFs.
5. Enter the string you are looking for and click **Search**.

Use advanced features to further refine your search criteria. Refer to your PDF reader Help for details.