Cornelis® Omni-Path Express® Fabric Software

Release Notes for V10.14.1

May 2024
Doc. No. A00375, Rev. 2.1

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 allCornelis 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 email: support@cornelisnetworks.com

1.4. Product Improvements and Changes

The following improvements are included in this release:

- Support for Red Hat Enterprise Linux (RHEL) 8.7, 8.8, 9.1, and 9.2.
- Support for Rocky Linux 8.7, 8.8, 9.1, and 9.2.
- Support for SLES 15.5.
- Support for Nvidia GPU on latest OS with PSM2.
- Deprecate hfidiags RPM.

1.5. Release Packages

There are two OPX Fabric Software packages:

Table 1. Release Packages

Package	Includes
BASIC • For compute, service, and login nodes	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) • Open MPI and MVAPICH2. See Section 1.9 "MPI Libraries" for details. • opa-basic-tools • mpitests • mpi-selector • Open Fabrics Interface (OFI) libfabric
OPXS • For management nodes	Includes the BASIC package plus: • Fabric Manager – 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 – Enables rapid, error-free installation and configuration of OPX Host Software and management software tools – 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.

Table 2. Supported Operating Systems for Release 10.14.1

Operating System	Update	Base Kernel Version
Red Hat Enterprise Linux (RHEL) 8.7 X86_64	8.7	4.18.0-425.3.1.el8.x86_64
Red Hat Enterprise Linux (RHEL) 8.8 X86_64	8.8	4.18.0-477.10.1.el8_8.x86_64
Red Hat Enterprise Linux (RHEL) 9.1 X86_64	9.1	5.14.0-162.6.1.el9_1.x86_64
Red Hat Enterprise Linux (RHEL) 9.2 X86_64	9.2	5.18.0-284.11.el9_2.x86_64
Rocky Linux 8.7 (corresponds to RHEL 8.7)	8.7	4.18.0-372.9.1.el8.x86_64
Rocky Linux 8.8 (corresponds to RHEL 8.8)	8.8	4.18.0-477.10.1.el8_8.x86_64
Rocky Linux 9.1 (corresponds to RHEL 9.1)	9.1	5.14.0-162.6.1.el9_1.x86_64
Rocky Linux 9.2 (corresponds to RHEL 9.2)	9.2	5.18.0-284.11.el9_2.86_64
SUSE Linux Enterprise Server (SLES) 15.5 X86_64	SP 5	5.14.21-150500.55.44.1.x86_64

1.8. CUDA Support

GPUDirect RDMA with CUDA is supported using PSM2.

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

1.9. MPI Libraries

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

NOTE

If you are not using Intel MPI, do not run these scripts. If you need to load the Intel compilers, use the compiler-specific vars.sh included in OneAPI.

1.9.1. Supported MPI Libraries

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

NOTE

Contact Cornelis Networks Customer Support at support@cornelisnetworks.com if problems are found using Open MPI releases newer than 4.1.4.

Table 3. MPI Implimentation

MPI Implementation	Supported versions	In Package Tar file	Runs over	Compiled with
Open MPI	4.1.X	No	ofi mtl (psm2/opx), psm2 mtl	GCC
	4.1.4	Yes		GCC
	4.1.4-cuda	Yes		GCC
MVAPICH	2.3.7	Yes	PSM2	GCC
Intel MPI Library	2021 Update 3 and above	No	OFI (OPX, PSM2)	N/A
Sandia OpenSHMEM	1.5.x	No	PSM2, OPX provider	N/A

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.

Refer to the table below for the list of Cornelis commit IDs.

Table 4. OFI Component Git Commit IDs

Component	Version	Branch	Commit ID
Kernel Driver	RHEL8.7	pub/rhel8.7	e3dc210
Kernel Driver	RHEL8.8	pub/rhel8.8	73f085d
Kernel Driver	RHEL9.1	pub/rhel9.1	ca4dc36
Kernel Driver	RHEL9.2	pub/rhel9.2	188507b
Kernel Driver	SLES15.5	pub/sles15.5	2f29dde
PSM2	12.0.1.0	master	ad5dd1b
OPX	Libfabric 1.20.1+	10.14.0-prerelease	dd7781e
OPENIB_INSTALL	10.14.1.0.6	opa-10_14	853f026
FABRIC_SIM	10.14.1.0.5	opa-10_14	3a61178
FM (VIEO_HOST)	10.14.1.0.5	opa-10_14	f405e33
FM (OPENIB_FF)	10.14.1.0.5	opa-10_14	9121215
OFED_MPIS	10.14.1.0.6	opa-10_14	c5af337
OFED DELTA	10.14.1.0.6	opa-10_14	6d9322a

1.10. 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 provider has been tested with up to eight (8) HFIs per server when running in non-multirail mode (PSM2_MULTIRAIL=0) and four (4) HFIs per server when running with PSM2_MULTIRAIL!=0.

Table 5. 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 CPU-based servers
3rd Gen AMD EPYC Processor	Milan Zen 3 CPU-based servers
4th Gen AMD EPYC Processor	Genoa Zen 4 CPU-based servers
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	Ice Lake CPU-based servers
4th Generation Intel Xeon Scalable Processor	Sapphire Rapids CPU-based servers
NVIDIA Graphics Processor	Tesla Series PCIe-based GPUs

1.11. 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. 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. RDMA support is implemented in the OPX provider via a non-default mechanism and can be enabled via the environment variable FI_OPX_EXPECTED_RECEIVE_ENABLE=1.

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

1.12. Installation Requirements

This section provides installation requirements for this release.

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.

1.12.1. 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. These are found in the Cornelis Customer Center.
- 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).
  - Configure 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.

1.12.2. Installation Instructions

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.

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

The `-G` option must be used with the INSTALL command to enable OPXS components with NVIDIA-CUDA support.

1.13. Product Constraints

Power class 2 AOCs are supported. This support requires a minimum Host software version of 10.5 and a minimum UEFI version of 1.5. Integrated HFI (-F) requires a specific BIOS level to support power class 2 AOC. Contact your BIOS vendor for more information.

1.14. Product Limitations

This release has the following product limitations:

- On systems where the total number of CPU cores exceeds the limit of the HFI Driver contexts (160 for Omni-Path HFI), Accelerated IPoFabric (AIP) does not enable correctly. For example, on dual-socket 96 core CPUs totaling 192 cores, a message similar to the following may be seen in dmesg: `hfi1_0: No receive contexts available for netdevs`. If AIP Is required for IPoFabric performance, the hfi1 driver must be configured to allow AIP to function properly by restricting the hfi1 module parameter `num_user_contexts` to a lower value such as 149 (this works with the default `hfi1 krcvqs` value of 2). See the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, "Setting HFI1 Module Parameters" for more detail. Contact Cornelis Networks Customer Support for more technical detail or guidance for your exact system.
- The OPX Libfabric provider has a limit on the number of processes (or MPI ranks) per Omni-Path host fabric interface (HFI) card due to direct binding of processes to hardware contexts on the HFI. This limit may be encountered when running with a single HFI and one MPI rank per core on some high core count, multi-socket servers. For more detailed information about this limit, please contact Cornelis Networks Customer Support.

  If this issue impacts your workload, please consider:

  - Using the PSM2 fabric provider to allow sharing contexts between ranks (which may cause resource contention under heavy load).
  - Executing your workload using fewer MPI ranks per server node, either using a reduced MPI rank count or in combination with an increased number of threads (such as pthreads or OpenMP threads) where applicable.
  - Deploying additional HFIs in the server to increase the total available hardware contexts.
- 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.

1.15. Accelerated RDMA

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

1.16. 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:

```
ACTION=="add|change|remove", SUBSYSTEM=="hfi1_user", RUN+="/usr/bin/killall -SIGHUP
irqbalance"
```

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. Issues Resolved in this Release

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

Table 6. Issues Resolved in this Release

ID	Description
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.
STL-61331	When installing on SLES 12.x from a yum/zypper repository, the installation may fail due to missing dependencies.
STL-64192	The OPXS repo based installation fails on RHEL 9 and SLES 15.4 due to missing dependencies for Python2 which has been deprecated in both OSes. OPXS packages a utility, hfidiags, which requires Python2. The utility is deprecated and will be removed in an upcoming OPXS release.
STL-64193	In SLES15 distros, libpsm_infinipath1 is designated to obsolete libpsm2-compat even though the rpms can coexist. For some use cases (such as rebuilding MPI from source), both rpms must be installed. When a user installs OPXS using zypper, the process fails when installing libpsm2-compat due to this conflict.
STL-65954	OPX does not currently support separating traffic to specified Virtual Lanes.
STL-67571	Open MPI collectives functions (e.g., MPI_Alltoall) may require the smcuda BTL to work on GPU buffers for intranode and intrarank communications. Not providing the smcuda BTL may result in a segfault.

2.2. Open Issues

The following table lists the open issues for this release.

Table 7. Open Issues

ID	Description	Workaround
139613 STL-49732	The Subsystem Vendor and Subsystem Device ID in the OPX PCI configuration space of 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.
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-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-64175	When Installing OPXS software, dracut warnings can be observed on the console. These warnings are not a result of OPXS software and do not impact functionality.	None. See SLES documentation for more information.

ID	Description	Workaround
STL-64620	If using the OPXS repo-based installation on SLES15.4 and the suggested zypper workarounds are followed to remove the conflicting `libpsm_infinipath1` package in favor of the `libpsm2-compat` package, a segfault may result when running libfabric due to a missing link to `libpsm_infinipath.so.1`.	The libpsm2-compat package provides `libpsm_infinipath.so.1` that is required by libfabric. To instruct libfabric on how to find the path to the library: • Add a config file for `psm2-compat` that points to the location of the `libpsm_infinipath.so.1` provided by `libpsm2-compat` `cat /etc/ld.so.conf.d/psm2compat.conf /usr/lib64/psm2-compat` • Load the new config file using ldconfig
STL-64974	Failures have been observed when running the IMB-EXT "Window" benchmark using the OPX Provider and ppn greater than (num_user_contexts/2).	To run the IMB-EXT benchmark with the OPX provider and higher PPN, reload the hfi1 module with an increased num_user_contexts value. For example, if the server has 40 physical cores and you want to run one IMB-EXT process per core, increase the number of hfi1 contexts to 80. See the *Cornelis Omni-Path Express Fabric Performance Tuning User Guide*, section "HFI1 Driver Module Parameters" for instructions. The benchmark can now be run with #contexts/2 = 40 PPN, one for each CPU core.
STL-65516 STL-67638	On AMD systems running RHEL 9.1, IMB benchmarks running over libfabric may fail with an "OFI get address vector map" error.	IOMMU should be disabled in the bios. See *Cornelis Omni-Path Express Fabric Performance Tuning User Guide* for recommended IOMMU settings for AMD systems.
STL-65757	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)In release 10.13, the oversubscription case may result in a segfault with the following errormessage: *** Process received signal *** Signal: Segmentation fault (11) Signal code: (-6)	Set PSM2_MULTI_EP=0 (user environment variable) before or during job launch.

ID	Description	Workaround

STL-65811	During installation, the following warning messages can be observed:
	`Not Found: ./CornelisOPX-Tools-FF.RHEL87-x86_64.10.13.0.0.6/RPMS/*/opa-address-resolution-devel user`
	`Not Found: ./CornelisOPX-Tools-FF.RHEL87-x86_64.10.13.0.0.6/RPMS/*/opa-snapconfig user`
		None. These warnings can be safely ignored and installation will proceed successfully without impact to functionality.

STL-66199	On RHEL 9.x, the `opaswitchadmin` command used to update unmanaged switch firmware fails with the following error:
	`Invalid MAD: Invalid Session ID`
	`Error: Failed to get ini descriptors - status 3`
	`Error: Failed to acquire node description - status -1`
	`Invalid MAD: Invalid Session ID`
	`Error: Failed to get ini descriptors - status 3`
		Contact Cornelis Customer Support

STL-67178	NVIDIA GPU support with PSM2 is unavailable when the NVIDIA driver is not loaded prior to loading the hfi1 driver.	Review boot configurations to ensure NVIDIA GPU driver is available prior to running OPXS INSTALL script or when loading the hfi1 driver.
STL-67691	On AMD CPUs with SMT enabled and IOMMU disabled, the hfi1 driver does not allocate one context per physical CPU core as intended. The PSM2 provider will run in context-sharing mode, but the OPX Provider will not run due to lack of context sharing support.	Disable SMT or run with the PSM2 provider in context sharing mode.
STL-68089	During installation of OPXS packaging, a message similar to the example below may be observed: `Not Found: ./CornelisOPX-OFA_DELTA.SLES155-x86_64.10.14.0.0.12/RPMS/SuSE-ES155/libfabric-psm user`	This message may be ignored.
STL-68168	When enabling FI_OPX_EXPECTED_RECEIVE_ENABLE=1, this feature may fail with the following warning: modinfo: command not found . On SLES systems modinfo is found in /sbin/ .	Include the path to modinfo in your mpirun command line. For Open MPI, use "`-x PATH=/sbin/:$PATH`". For Intel MPI, use "`-genv PATH=/sbin/:$PATH`".
STL-68208	When using Ipoib in connected mode with high traffic and bouncing links at the same time, errors may occur and, in rare circumstances, result in a kernel panic.	Use Ipoib in datagram mode, or use connected mode but do not bounce links unnecessarily.

ID	Description	Workaround
STL-69062	The PSM2 error "Unhandled error in TID Update: Bad address" may occur when running applications that use CUDA Unified Memory.	None. GPUDirect does not support CUDA Unified Memory. Contact your application vendor to inquire about not using CUDA managed memory for psm2 operations.
STL-69063	Repo-based install for OPXS 10.14.1.1.1 on SLES15 SP5 reports a file conflict for `/usr/lib64/libopasadb.so.1.0.0` between OPXS opa-address-resolution and in-distro libopasadb1_0_0.	For non-interactive repo-based install, zypper must include the --replacefiles option: 1. `zypper install --from CornelisOPX-OPXS -y --replacefiles opanode_compute` 2. After uninstalling OPXS software, restore the missing /usr/lib64/libopasadb.so.1.0.0 file by re-installing in-distro libopasadb package (--force is needed to force reinstallation of a package that's already installed): `zypper install --force -y libopasadb1_0_0`

2.2.1. Third Party Open Issues

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

Table 8. 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 $1 \leq n \leq 23$) 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.

ID	Description	Workaround
STL-61762	osu_mbw_mr can report bandwidth greater than line-rate in OSU micro benchmarks version 4.1 and higher.	None.
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-64238	When running the IMB-RMA tests that are a component of the Intel MPI Microbenchmarks, if a version of libfabric is used other than the one provided with Intel OneAPI, certain tests (Accumulate, Get_accumulate, Fetch_and_op, and Compare_and_swap) will hang if the environment variable `I_MPI_ROOT` is set. The Intel OneAPI `env vars` script, which may be sourced before using Intel OneAPI MPI, does set this environment variable.	To prevent possible hangs, Cornelis recommends making sure the `I_MPI_ROOT` environment variable is not set before using the OPX provider libfabric with an Intel OneAPI MPI distribution, especially if making use of MPI RMA functionality.
STL-66803	kdump does not work by default on RHEL 9.2	In some RHEL releases there is a known issue (Chapter 8. Known issues Red Hat Enterprise Linux 9 | Red Hat Customer Portal) assigning memory for crash dumps. This is indicated by `kdump: No memory reserved for crash kernel` in the system logs. The result is the system will be unable to collect crash dumps for analysis of kernel, panics. Customers can work around the issue with a command similar to `grubby --args crashkernel=512M --update-kernel ALL`. It is suggested to contact the OS vendor for the appropriate value to use.
STL-67113	Timeouts, hangs, and memory allocation issues can occur with MPI jobs if the `memlock` limit is not sufficient.	Set the following parameters in `/etc/security/limits.conf` • `- hard memlock unlimited` • `- soft memlock unlimited`

ID	Description	Workaround
STL-68521	Running `modprobe -r hfi1` results in error message, `modprobe: FATAL: Module nvidia is in use.`	This concerns NVIDIA's module unloading and is outside of Cornelis control. It does not prevent unloading/reloading hfi1. This message can be ignored.

3. Related Information

3.1. 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.
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

Task	Document Title	Description
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.
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 Performance Scaled Messaging 2 (PSM2) Application Programming Interface (API). The 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.

Task	Document Title	Description
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.
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	

3.1.1. 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.