Mellanox HPC-X™ Software Toolkit Release Notes

Rev 2.3

NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "AS-IS" WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCT(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANTTHAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Mellanox Technologies
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 9703400
Fax: (408) 9703403

Table of Contents

List Of Tables

Release Update History

Table 1 - Release Update History

Release	Date	Description
Rev 2.3	December 3, 2018	Initial version of this HPC-X version.

1 Overview

These are the release notes for the Mellanox HPC-X™ Rev 2.3. The Mellanox HPC-X™ Software Toolkit is a comprehensive software package that includes Open MPI, OpenSHMEM, PGAS, MXM, UCX, FCA tool suite for high performance computing environments. HPC-X provides enhancements to significantly increase the scalability and performance of message communications in the network. HPC-X™ enables you to rapidly deploy and deliver maximum application performance without the complexity and costs of licensed third-party tools and libraries.

1.1 HPC-X™ Requirements

The platform and requirements for HPC-X are detailed in the following table:

Platform	Drivers and HCAs
OFED / MLNX_OFED	• OFED 1.5.3 • MLNX_OFED 4.3-x.x.x and above
HCAs	• ConnectX®-5 / ConnectX®-5 Ex • ConnectX®-4 / ConnectX®-4 Lx • ConnectX®-3 / ConnectX®-3 Pro • Connect-IB®

1.2 HPC-X™ Content

The following communications libraries and acceleration packages are part of Mellanox HPC-X™ Rev 2.3 package:

Library/Acceleration Package	Version Number
Open MPI	4.0.x
Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)	1.7.2
HCOLL	4.2
UCX	1.5
Open SHMEM specification compliant	1.4[a]

a. Open SHMEM v1.4 compliance is at beta level.

1.3 Important Notes

When HPC-X is launched in an environment without resource manager (slurm, pbs, ...) installed, or from a compute node, it will use Open MPI default rsh/ssh based launcher which does not propagate the library path to the compute nodes.

In such case, pass the `LD_LIBRARY_PATH` variable as following:

```
% mpirun -x LD_LIBRARY_PATH -np 2 $HPCX_MPI_TESTS_DIR/examples/hello_c
```

2 Changes and New Features

HPC-X™ Rev 2.3 provides the following changes and new features:

Table 2 - Changes and New Features

Category	Description
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • Open MPI version 4.0.x • Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) version 1.7.2 • HCOLL version 4.2 • UCX version 1.5 • OpenSHMEM version 1.4
UCX	UCX is now compiled without JAVA bindings.
	Added support for running UCX over rdma-core, for DC transport and direct verbs.
	Emulation layer: Added the ability to run UCX over software emulation of remote memory access and atomic operations. This provides full support of SHMEM and MPI-RMA over shared memory, TCP, and older RDMA hardware, such as ConnectX-3 HCA.
HCOLL	HCOLL and Mellanox SHARP are now compiled with CUDA support.
	Added support for CUDA buffers over SRA allreduce algorithm.
MXM	Removed support for MXM library.
OpenMPI	Added the following configuration options to OMPI: • `--with-libevent=internal` • `--enable-mpi1-compatibility`
	Updated the configuration file platform/mellanox/optimized config in OMPI upstream by removing BTL OpenIB and UCT support and removing links to MXM/FCA usage.
	Removed PMI2 support.

3 Known Issues

The following is a list of general limitations and known issues of the various components of this HPC-X release.

Table 3 - Known Issues (Sheet 1 of 3)

Internal Ref.	Issue
1582208	**Description:** Sending data over multiple SHMEM contexts may lead to memory corruption or segmentation fault.
	Workaround: Add the command line "`-x UCX_SELF_DEVICES=`" when running `oshrun`, or set it before the run.
	Keywords: Open SHMEM, segmentation fault
	Discovered in Version: 2.3 (Open MPI v4.0.x, OpenSHMEM v1.4)
-	**Description:** OSC UCX module is not selected by default on ConnectX-4/ConnectX-5 HCAs.
	Workaround: Add the command line argument "`-mca osc ucx`" when running `mpirun.`
	Keywords: OSC UCX, one-sided, Open MPI
	Discovered in Version: 2.3 (Open MPI v4.0.x, OpenSHMEM v1.4)
-	**Description:** Zero-length OpenSHMEM collectives may fail due to incomplete implementation.
	Workaround: N/A
	Keywords: OpenSHMEM atomic, Open MPI
	Discovered in Version: 2.3 (Open MPI v4.0.x, OpenSHMEM v1.4)
-	**Description:** OpenSHMEM atomic operations AND/OR/XOR for datatypes int32/int64/uint32/uint64 are not implemented, which may cause build failures.
	Workaround: N/A
	Keywords: OpenSHMEM atomic, Open MPI
	Discovered in Version: 2.3 (Open MPI v4.0.x, OpenSHMEM v1.4)
2934	**Description:** OpenMPI and OpenSHMEM applications may hang with DC transport. (Github issue: https://github.com/openucx/ucx/issues/2934)
	Workaround: Use RC or UD instead of DC. To achieve that, make sure UCX pml (or UCX osc/smpl for one-sided applications) is used, and UCX_TLS variable is set. Example: `UCX_TLS=rc,self,sm`
	Keywords: UCX, Open MPI, DC
	Discovered in Version: 2.3 (Open MPI v4.0.x, OpenSHMEM v1.4)

Table 3 - Known Issues (Sheet 2 of 3)

Internal Ref.	Issue
1307243	**Description:** One-sided tests may fail with a segmentation fault.
	Workaround: In order to run one-sided tests, make sure to add `-mca osc ucx` to the command line.
	Keywords: OSC UCX, Open MPI, one-sided
	Discovered in Version: 2.1 (Open MPI 3.1.x)
-	**Description:** In ConnectX-4 and Connect-IB HCAs, when the DC transport is used on a large scale, “Retry exceeded” messages may be printed from UCX.
	Workaround: Configure SL2VL on your OpenSM in the fabric and make UCX use SL=1 when using the InfiniBand transports via '-x UCX_IB_SL=1'.
	Keywords: UCX, DC transport, ConnectX-4, Connect-IB
	Discovered in Version: 2.1 (UCX 1.3)
-	**Description:** When UCX requires more memory utilization than the memory space defined in /proc/sys/kernel/shmmni file, the following message is printed from UCX: “`... total number of segments in the system (%lu) would exceed the limit in /proc/sys/kernel/shmmni (=%lu)... please check shared memory limits by 'ipcs -l`”.
	Workaround: Follow the instructions in the error message above and increase the value of shared memory segments in /proc/sys/kernel/shmmni file.
	Keywords: UCX, memory
	Discovered in Version: 2.1 (UCX 1.3)
2226	**Description:** The following assertion may fail in certain cases: `Assertion `ep->rx.ooo_pkts.head_sn == neth->psn' failed` (Gihub issue: https://github.com/openucx/ucx/issues/2226)
	Workaround: Set the DC transport using the `UCX_TLS` parameter.
	Keywords: UCX, assertion
	Discovered in Version: 2.1 (UCX 1.3)
-	**Description:** Mellanox SHARP library is not available in HPC-X for the Community OFED and Inbox OFED.
	Workaround: N/A
	Keywords: Mellanox SHARP library
	Discovered in Version: 2.0
1162	**Description:** UCX currently does not support canceling send requests. (Github issue: https://github.com/openucx/ucx/issues/1162)
	Workaround: N/A
	Keywords: UCX
	Discovered in Version: 2.0

Table 3 - Known Issues (Sheet 3 of 3)

Internal Ref.	Issue
-	**Description:** UCX job hangs with SocketDirect/MultiHost/SR-IOV.
	Workaround: Set `UCX_IB_ADDR_TYPE=ib_global`
	Keywords: UCX
-	**Description:** As UCX embedded in the HPC-X is compiled with AVX support, UCX cannot be run on hosts without AVX support. In case the AVX is not available, recompile the UCX that is available in the HPC-X with the option: `--with-avx=no`
	Workaround: Recompile UCX with AVX disabled: `$ ./utils/hpcx_rebuild.sh --rebuild-ucx --ucx-extra-config "--with-avx=no"`
	Keywords: UCX

4 Bug Fixes History

Table 4 lists the bugs fixed in this release.

Table 4 - Bug Fixes History (Sheet 1 of 5)

Internal Ref.	Issue
-	**Description:** Fixed the issue where using UCX on ARM hosts may result in hangs due to a known issue in Open MPI when running on ARM.
	Keywords: UCX
	Discovered in Version: 1.3 (Open MPI 1.8.2)
	Fixed in Version: 2.3 (Open MPI 4.0.x)
-	**Description:** MCA options `rmaps_dist_device` and `rmaps_base_mapping_policy` are now functional.
	Keywords: Process binding policy, NUMA/HCA locality
	Discovered in Version: 2.0 (Open MPI 3.0.0)
	Fixed in Version: 2.3 (Open MPI 4.0.x)
2111	**Description:** Fixed the issue of when UCX was used in the multi-threaded mode, it might have taken the `osu_latency_mt` test a long time to be completed. (Github issue: https://github.com/openucx/ucx/issues/2111)
	Keywords: UCX, multi-threaded
	Discovered in Version: 2.1 (UCX 1.3)
	Fixed in Version: 2.3 (UCX 1.5)
2267	**Description:** Fixed the issue where the following error message might have appeared when running at the scale of 256 ranks with the RC transport, when UD is used for wireup only: "`Fatal: send completion with error: Endpoint timeout`". (Github issue: https://github.com/openucx/ucx/issues/2267)
	Keywords: UCX
	Discovered in Version: 2.1 (UCX 1.3)
	Fixed in Version: 2.3 (UCX 1.5)

Table 4 - Bug Fixes History (Sheet 2 of 5)

Internal Ref.	Issue
2702	**Description:** Fixed the issue of when using the Hardware Tag Matching feature, the following error messages may have been printed: • `"rcache.c:481 UCX WARN failed to register region 0xdec25a0 [0x2b7139ae0020..0x2b7139ae2020]: Input/output error"` • `"ucp_mm.c:105 UCX ERROR failed to register address 0x2b7139ae0020 length 8192 on md[1]=ib/mlx5_0: Input/output error"` • `"ucp_request.c:259 UCX ERROR failed to register user buffer datatype 0x20 address 0x2b7139ae0020 len 8192: Input/output error"` (Github issue: https://github.com/openucx/ucx/issues/2702)
	Keywords: Hardware Tag Matching
	Discovered in Version: 2.2 (UCX 1.4)
	Fixed in Version: 2.3 (UCX 1.5)
2454	**Description:** Fixed the issue where some one-sided benchmarks may have hung when using "osc ucx". For example: osu-micro-benchmarks-5.3.2/osu_get_acc_latency (Latency Test for accumulate with Active/Passive Synchronization). (Github issue: https://github.com/openucx/ucx/issues/2454)
	Keywords: UCX, one_sided
	Discovered in Version: 2.2 (UCX 1.4)
	Fixed in Version: 2.3 (UCX 1.5)
2670	**Description:** Fixed the issue of when enabling the Hardware Tag Matching feature on a large scale, the following error message may have been printed due to the increased threshold for BCOPY messages: `"mpool.c:177 UCX ERROR Failed to allocate memory pool chunk: Out of memory."` (Github issue: https://github.com/openucx/ucx/issues/2670)
	Keywords: Hardware Tag Matching
	Discovered in Version: 2.2 (UCX 1.4)
	Fixed in Version: 2.3 (UCX 1.5)
1295679	**Description:** Fixed the issue where OpenSHMEM group cache had a default limit of 100 entries, which might have resulted in OpenSHMEM application exiting with the following message: "`group cache overflow on rank xxx: cache_size = 100`".
	Keywords: OpenSHMEM, Open MPI
	Discovered in Version: 2.1 (Open MPI 3.1.x)
	Fixed in Version: 2.2 (Open MPI 3.1.x)

Table 4 - Bug Fixes History (Sheet 3 of 5)

Internal Ref.	Issue
-	**Description:** Fixed the issue where UCX did not work out-of-the-box with CUDA support.
	Keywords: UCX, CUDA
	Discovered in Version: 2.2 (UCX 1.4)
	Fixed in Version: 2.1 (UCX 1.3)
1926	**Description:** Fixed the issue of when using multiple transports, invalid data was sent out-of-sync with Hardware Tag Matching traffic. (Github issue: https://github.com/openucx/ucx/issues/1926)
	Keywords: Hardware Tag Matching
	Discovered in Version: 2.1 (UCX 1.3)
	Fixed in Version: 2.2 (UCX 1.4)
1949	**Description:** Fixed the issue where Hardware Tag Matching might not have functioned properly with UCX over DC transport. (Github issue: https://github.com/openucx/ucx/issues/1949)
	Keywords: UCX, Hardware Tag Matching, DC transport
	Discovered in Version: 2.0
	Fixed in Version: 2.1
-	**Description:** Fixed job data transfer from SD to libsharp.
	Keywords: Mellanox SHARP library
	Discovered in Release: 1.9
	Fixed in Release: 1.9.7
884482	**Description:** Fixed internal HCOLL datatype mapping.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
884508	**Description:** Fixed internal HCOLL datatype lower bound calculation.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
884490	**Description:** Fixed allgather unpacking issues.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406

Table 4 - Bug Fixes History (Sheet 4 of 5)

Internal Ref.	Issue
885009	**Description:** Fixed wrong answer in alltoallv.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
882193	**Description:** Fixed mcast group leak in HCOLL.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
-	**Description:** Added IN_PLACE support for alltoall, alltoallv, and allgatherv.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
-	**Description:** Fixed an issue related to multi-threaded MPI_Bcast.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
Salesforce: 316541	**Description:** Fixed a memory barrier issue in MPI_Barrier on Power PPC systems.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
Salesforce: 316547	**Description:** Fixed multi-threaded MPI_COMM_DUP and MPI_COMM_SPLIT hanging issues.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
894346	**Description:** Fixed Quantum Espresso hanging issues.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406
898283	**Description:** Fixed an issue which caused CP2K applications to hang when HCOLL was enabled.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.7.405
	Fixed in Release: 1.7.406

Table 4 - Bug Fixes History (Sheet 5 of 5)

Internal Ref.	Issue
906155	**Description:** Fixed an issue which caused VASP applications to hang in MPI_Allreduce.
	Keywords: HCOLL, FCA
	Discovered in Release: 1.6
	Fixed in Release: 1.7.406

5 Change Log History

5.1 HPC-X Toolkit Change Log History

Table 5 - HPC-X Toolkit Change Log History

Category	Description
Rev 2.2	
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) version 1.7 • HCOLL version 4.1 • UCX version 1.4
	Added support for Singularity containerization. For further information, please refer to HPC-X User Manual.
	"osc ucx" is no longer the default one-sided-component in OpenMPI.
	Removed KNEM library from HPC-X package. UCX will use the KNEM available in MLNX_OFED.
MXM Support	Open MPI and HCOLL are not compiled with MXM anymore. Both are compiled with UCX only and use it by default.
UCX	Added support for the following UCX features: • New API for establishing client-server connection. • Out-of-box support for Memory In Chip (MEMIC) on ConnectX-5 HCAs.
HPC-X Setup	Added support for HPC-X to work on Huawei ARM architecture.
HCOLL	Improved performance by utilizing zero-copy messaging for MPI Bcast.
Rev 2.1	
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • Open MPI version 3.1.x • Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) version 1.5 • HCOLL version 4.0 • MXM version 3.7 • UCX version 1.3 • OpenSHMEM v1.3 specification compliant

UCX	• UCX is now the default pml layer for Open MPI, default spml layer for OpenSHMEM, and default OSC component for MPI RMA.
	• Added the following UCX features:
	  • Added support for GPU memory in UCX communication libraries
	  • Added support for Multi-Rail protocol

MXM	The UD_RNDV_ZCOPY parameter is set to 'no' by default. This means that the zcopy mechanism for the UD transport is disabled when using the Rendezvous protocol.

Table 5 - HPC-X Toolkit Change Log History

Category	Description
HCOLL	• UCX is now the default p2p transport in HCOLL • Improved multi-threaded performance • Improved shared memory performance • Added support for Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) v1.5 • Added support for Mellanox SHARP software multi-channel/multi-rail capable algorithms • Improved Allreduce large message algorithm • Improved AlltoAll algorithm
Profiling IB verbs API (ibprof)	Removed ibprof tool from HPC-X toolkit.
UPC	Removed UPC from HPC-X toolkit.
Rev 2.0	
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • OpenMPI version 3.0.0 • Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) version 1.4 • HCOLL version 3.9 • UCX version 1.3

UCX	• UCX is now at GA level.
	• Added the following UCX features:
	  • **[ConnectX-5 only]** Added support for hardware Tag Matching with DC transport.
	  • **[ConnectX-5 only]** Added support for Out-of-order RDMA RC and DC to support adaptive routing with true RDMA.
	  • Hardware Tag Matching (See section *Hardware Tag Matching* in the User Manual)
	  • SR-IOV Support (See section *SR-IOV Support* in the User Manual)
	  • Adaptive Routing (AR) (See section *Adaptive Routing* in the User Manual)
	  • Error Handling (See section *Error Handling* in the User Manual)

Table 5 - HPC-X Toolkit Change Log History

Category	Description
HCOLL	• Added support for Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) v1.4 • Added support for NCCL on-host GPU based collectives. • Added support for Hierarchical GPU based allreduce using NCCL for scale-in and MXM/UCX for scale-out. • Improved shared memory performance for allreduce, barrier, and broadcast. Targeting high thread count systems, e.g. Power9. • Improved large message allreduce (multi-radix, zero-copy fragmentation, CPU vectorization.) • Added new and improved AlltoAllv algorithm - hybrid logarithmic pairwise exchange. • Added support for on-demand HCOLL memory. Improves HCOLL's memory footprint on high thread count system e.g. Power9. • Added a high performance multithreaded implementation to support MPI_THREAD_MULTIPLE applications. Designed specifically for high thread count systems, e.g. Power9. • HCOLL startup improvements.
Open MPI / OpenSHMEM	• Added support for Open MPI 3.0.0. • Added support for xpmem kernel module. • Added a high performance implementation of shmem_ptr() with UCX SPML. • Added a UCX allocator. The UCX allocator optimizes intra-node communication by allowing direct access to memories of processes on the same node. The UCX allocator can only be used with the UCX SPML. • Added a UCX one-sided component to support MPI RMA operations.
Rev 1.9.7	
Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)	Bug Fixes, see Section 4, “Bug Fixes History”, on page 11
Rev 1.9	
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • OpenMPI version 2.1.2a1 • Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) version 1.3.1 • HCOLL version 3.8.1652 • MXM version 3.6.3103 • UCX version 1.2.2947

Table 5 - HPC-X Toolkit Change Log History

Category	Description
UCX	Point-to-point communication API, with tag matching, remote memory access, and atomic operations. This can be used to implement MPI, PGAS, and Big Data libraries and applications- IB transport
	A cleaner API with lower software overhead which provides better performance especially for small messages.
	Support for multitude of InifiniBand transports and Mellanox offloads to optimize data transfer performance: • RDMA • DC • Out-of-order • HW tag matching offload • Registration cache • ODP
	Shared memory communications for optimal intra-node data transfer: • SysV • posix • knem • CMA • xpmem
MXM	Enabled Adaptive Routing for all the transport layers (UD/RC/DC).
	Memory registration optimization.
Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)	Improved the Out-of-the-box performance of Scalable Hierarchical Aggregation and Reduction Protocol (SHARP).
Shared memory	Improved the intranode performance of allreduce and barrier.
Configuration	Changed many default parameter setting in order to achieve best out-of-the-box experience for several applications including - CP2K, miniDFT, VASP, DL-POLY, Amber, Fluent, GAMES-UK, and LS-DYNA.
FCA	As of HPC-X v1.9, FCA v2.5 is no longer included in the HPC-X package.
	Improved AlltoAllv algorithm.
	Improved large data allreduce.
	Improved UCX BCOL.
OS architecture	Added support for ARM architecture.
Rev 1.8.2	
MXM	Updated MXM version to 3.6.2098 which includes memory registration optimization.
Rev 1.8	
Cross Channel (CC)	Added Cross Channel (CC) AlltoAllv
	Added CC zcpy Ring Bcas

Table 5 - HPC-X Toolkit Change Log History

Category	Description
Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)	Added Scalable Hierarchical Aggregation and Reduction Protocol (SHARP) non-blocking collectives
Shared memory POWER	Added shared memory POWER optimizations for allreduce
	Added shared memory POWER optimizations for Barrier
Mixed data types	Added support for mixed data types
Non-contiguous Bcast	Added support for non-contiguous Bcast with UMR or SGE in CC
UMR	Added UMR support in CC bcol
Unified Communication - X Framework (UCX)	A new acceleration library, integrated into the Open MPI (as a pml layer) and available as part of HPC-X. It is an open source communication library designed to achieve the highest performance for HPC applications.
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • HCOLL updated to v3.7. Open MPI updated to v2.10
FCA	FCA 2.x is no longer the default FCA used in HPC-X. As of HPC-X v1.8, FCA 3.x (HCOLL) is the default FCA used and it replaces FCA v2.x.
Bug Fixes	See Section 4, “Bug Fixes History”, on page 11
Rev 1.7	
MXM	Updated MXM version to 3.6
FCA Collective	Added Cross-Channel based Allgather, Bcast, 8-byte Allreduce.
FCA	Added MPI datatype support.
	Added optimizations for PPC platforms.
	Added support for multiple Mellanox SHARP technology leaders on a single host.
	Added support for collecting Mellanox SHARP technology usage statistics.
	Exposed cross-channel non-blocking collectives to the MPI level.
Rev 1.6	
MXM v3.5	See Section 5.3, “MXM Change Log History”, on page 23
IB-Router	Allows hosts that are located on different IB subnets to communicate with each other. This support is currently available when using the `'openib btl'` in Open MPI. **Note:** When using `'openib btl'`, RoCE and IB router are mutually exclusive. The Open MPI inside HPC-X 1.6 is not compiled with ib-router support, therefore it supports RoCE out-of-the-box.
FCA v3.5	See Section 5.2, “FCA Change Log History”, on page 21
Rev 1.5	

Table 5 - HPC-X Toolkit Change Log History

Category	Description
HPC-X Content	Updated the following communications libraries and acceleration packages versions: • Open MPI updated to v1.10 • UPC update to 2.22.0 • MXM updated to v3.4.369 • FCA updated to v3.4.799
MXM v3.4.369	See Section 5.3, “MXM Change Log History”, on page 23
FCA v3.4.799	See Section 5.2, “FCA Change Log History”, on page 21
Rev 1.4	
FCA v3.3	See Section 5.2, “FCA Change Log History”, on page 21
MXM v3.4	See Section 5.3, “MXM Change Log History”, on page 23
Rev 1.3	
MLNX_OFED	Added support for OFED Inbox drivers
CPU Architecture	Added support for PPC architecture
LID Mask Control (LMC)	Added support for multiple LIDs usage when the LMC in the fabric is higher than zero. MXM will use multiple LIDs to distribute traffic across multiple links and achieve better resource utilization.
Performance	Performance improvements for all transport layers.
Adaptive Routing	Enhanced support for Adaptive Routing for the UD transport layer. For further information, please refer to the HPC-X User Manual section *“Adaptive Routing for UD Transport”*.
UD zero copy	UD zero copy support on receiver side to achieve better bandwidth utilization and reduce CPU usage.

5.2 FCA Change Log History

Table 6 - FCA Change Log History

Category	Description
Rev 3.5	
FCA Collective	Added MPI Allgatherv and MPI reduce
FCA	Added support for Mellanox SHARP library (including SHARP allreduce, reduce and barrier)
	Enhanced scalability for CORE-Direct based collectives
	Added support for complex data types
Rev 3.4	

Table 6 - FCA Change Log History

Category	Description
General	UCX support
	Communicator caching scheme with eviction: improves jobstart and communicator creation time
Collectives	Collectives: Added Alltoallv and Alltoall small message algorithms.
Rev 3.3	
General	Ported to PowerPC
	Thread safety added
Collectives	Improved large message allreduce algorithm (Enabled by default)
	Beta version of network topology awareness (Enabled by default)
Rev 3.0	
Collectives	Offload collectives communication from MPI process onto Mellanox interconnect hardware
	Efficient collectives communication flow optimized to job and topology
MPI collectives	Significantly reduce MPI collectives runtime
MPI-3	Native support for MPI-3
Blocking and Non-blocking collectives	Support for blocking and nonblocking collectives
HCOLL	Supports hierarchical communication algorithms (HCOLL)
Collective algorithm	Supports multiple optimizations within a single collective algorithm
Performance	Increase CPU availability and efficiency for increased application performance
MPI libraries	Seamless integration with MPI libraries and job schedulers
Rev 2.5	
Multicast Group	Added MCG (Multicast Group) cleanup tool
Performance	Performance improvements
Rev 2.2	
Performance	Performance improvements
Dynamic offloading rules	Enabled dynamic offloading rules configuration based on the data type and reduce operations
Mixed MTU	Added support for mixed MTU
Rev 2.1.1	
AMD/Interlagos CPUs	Added support for AMD/Interlagos CPUs
Rev 2.1	
Core-Direct®	Added support for Mellanox Core-Direct® technology (enables offloading collective operations to the HCA.)

Table 6 - FCA Change Log History

Category	Description
Non-contiguous data lay-outs	Added support for non-contiguous data layouts
PGI compilers	Added support for PGI compilers

5.3 MXM Change Log History

Table 7 - MXM Change Log History

Category	Description
Rev 3.6	
General	Updated MXM version to 3.6
Rev 3.5	
Performance	Performance improvements
Rev 3.4.369	
Initialization	Job startup performance optimization
Supported Transports	DC enhancements and startup optimizations
Rev 3.4	
Supported Transports	Optimizations for the DC transport at scale
Rev 3.3	
LID Mask Control (LMC)	Added support for multiple LIDs usage when the LMC in the fabric is higher than zero. MXM will use multiple LIDs to distribute traffic across multiple links and achieve better resource utilization.
Adaptive Routing	Enhanced support for Adaptive Routing for the UD transport layer.
UD zero copy	UD zero copy support on receiver side to achieve better bandwidth utilization and reduce CPU usage.
Rev 3.2	
Atomic Operations	Added hardware atomic operations support in the RC and DC transport layers for 32bit and 64bit operands. This feature is set to ON by default. To disable it run: `oshrun -x MXM_CIB_USE_HW_ATOMICS=n ...` **Note:** If hardware atomic operations are disabled, the software atomic is used instead.
MXM API	Added two additional functions (`mxm_ep_wireup()` and `mxm_ep_power-down`)to the MXM API to allow pre-connection establishment for MXM (rather than on-demand). For further information, please refer to the HPC-X User Manual section *"MXM Performance Tuning"*.

Table 7 - MXM Change Log History

Category	Description
Event Interrupt	Added solicited event interrupt for the rendezvous protocol for potential performance improvement. For further information, please refer to the HPC-X User Manual section *"MXM Performance Tuning"*.
Performance	Performance improvements for the DC transport layer.
Adaptive Routing	Added Adaptive Routing for the UD transport layer. For further information, please refer to the HPC-X User Manual section *"Adaptive Routing for UD Transport"*.
Rev 3.0	
Service Level	Service Level support (at Alpha level)
Adaptive Routing	Adaptive Routing support in UD transport layers
Supported Transports	Dynamically Connected Transport (DC) (at GA level)
Performance	Performance optimizations
Rev 2.1	
Supported Transports	Dynamically Connected Transport (DC) (at Beta level)
	RC is currently fully supported
	KNEM support for Intra-node communication
Performance	Performance optimizations
Rev 2.0	
Reliable Connected	Added Reliable Connection (RC) support (at beta level)
MXM Binding	MXM process can be pinned to a specific HCA port. MXM supports the following binding policies: • static - user can specify process-to-port map • cpu affinity based - HCA port will be bound automatically based on process affinity
On-demand connection establishment	Added on-demand connection establishment between the processes
Performance	Performance tuning improvements
Rev 1.5	
MXM over Ethernet	Added Ethernet support
Multi-Rail	Added Multi-Rail support

5.4 HPC-X™ Open MPI/OpenSHMEM Change Log History

Table 8 - HPC-X™ Open MPI/OpenSHMEM Change Log History

Category	Description
Rev 2.2	
Performance	Added Sandy Bridge performance optimizations.
memheap	Allocated memheap using contiguous memory provided by the HCA.
ptmalloc allocator	Replaced the buddy memheap by the ptmalloc allocator.
multiple pSync arrays	Added the option of using multiple pSync arrays instead of barrier synchronization between collective routines (fcollect, reduction routines)
spml yoda	Optimized small size puts
Performance	Performance optimization
Memory footprint optimizations	Added memory footprint optimizations
Rev 1.8.2	
Acceleration Packages	Added support for new MXM, FCA, HCOLL versions
Job start optimization	Added job start optimization
Performance	Performance improvements