Hewlett Packard
Enterprise

HPE CRAY MPI UPDATE

Steve Oyanagi

SC'21 ANL MPICH BOF
November 17, 2021

CRAY MPICH - XC

- Current XC Release – Cray MPICH 7.7.18
  - Supports proprietary Aries network interconnect on Cray XC systems
  - Based on ANL MPICH 3.2 (CH3), compliant with the MPI-3.1 Standard
  - Misc bug fixes and minor improvements

HPE CRAY MPI – CRAY EX/APOLLO

- New version based on ANL MPICH 3.4a2 (CH4), compliant with the MPI-3.1 Standard
- Current Cray EX/Apollo Release – HPE Cray MPI 8.1.11

FEATURE		
Network architectures	Slingshot: OFI (verbs; rxm) provider, UCX driver support	InfiniBand clusters
CPU architectures	Intel CPUs AMD CPUs	Intel CPUs AMD CPUs
GPU architectures	AMD GPUs NVIDIA GPUs Intel GPUs (Aurora - work in progress)	InfiniBand support of GPUs under investigation
Operating Systems	RHEL/CENTOS, SLES, and HPE COS	RHEL/CENTOS
Compilers	CCE/Cray, GNU, NVIDIA, AMD, and Intel	CCE/Cray, GNU, NVIDIA, AMD, and Intel
Launcher support	Slurm, PALS	Slurm, PALS

HPE CRAY MPI – CRAY EX/APOLLO

- HPE Cray MPI Features
  - XPMEM/CMA support providing optimized on-node single-copy transfers
  - Performance and scaling optimizations for many collectives
  - Robust support for multiple NICs per node
  - Optimized GPU support for NVIDIA and AMD GPUs (on-node and inter-node)
    - IPC and RDMA support for NVIDA and AMD GPUs
    - Efficient data movement algorithms between CPU- and GPU-attached memory regions
  - Support for hugepage memory allocations: standard Linux page sizes, Transparent Huge Pages
  - Flexible, intuitive rank re-ordering feature
  - MPI I/O performance enhancements and stats
  - Support for MPI singleton launches and single host applications (without networking resources)
  - MPICH ABI compatible with MPICH ABI Initiative MPI implementations like Intel MPI, MVAPICH2, ANL MPICH for ISV application support
  - Scalable Cray PMI implementation for fast launch/startup

FOCUS AREAS FOR 2022

- HPE Cray MPI for Cray EX/Apollo
  - Expanded Apollo system support
  - Merge to ANL MPICH 4.0
  - MPI-4.0 standard compliance
  - More collective optimizations
  - Improved GPU support
    - Investigating InfiniBand support of GPUs
    - Investigating GPU-NIC Async solutions with some GPU vendors
  - Slingshot-11 support
    - Traffic classes (quality of service)
    - Hardware Collectives
    - Hardware atomics
    - Hardware support for tag-matching and strong progression of MPI rendezvous protocols
    - Reduced memory footprint
    - Scalable startup
    - More tuning and scaling
  - Optimizations to support future processors and interconnects

THANK YOU

Steve Oyanagi
steven.oyanagi@hpe.com