Performance Studies of MVAPICH2 Libraries on AWS and Oracle Clouds

9th MVAPICH User Group (MUG) Conference '21

Shulei Xu
The Ohio State University
xu.2452@osu.edu

https://twitter.com/mvapich
Presentation Outline

• AWS EC2 with EFA
• OCI HPC Workloads
• Work in Progress & Future Plans
MVAPICH2 Libraries on AWS EFA

• Scalable and reliable MPI library support on AWS EC2 instances with Elastic Fabric Adapter (EFA) support

• Current Optimization:
  – Utilizing larger MTU Size
  – Re-ordering data packets
  – Performance tuning on x86 systems
  – Automatic XPMEM Kernel detection

• Support to new ARM instance type with AWS Graviton ARM CPU.

• Experiment Setups:
  – Instance Type: c6gn.16xlarge
  – CPU: AWS Graviton (64 cores per node)
  – OS: Amazon Linux 2
  – Parallel Cluster 2.7.1
  – Versions of MPI Libraries:
    • MVAPICH2-X-AWS 2.3
    • OpenMPI 4.1.0
  – Benchmarks:
    • OSU Microbenchmarks 5.8
MVAPICH2 Libraries on AWS EFA

- Newly added support and optimizations to AWS EC2 HPC instances with ARM architecture
- Point-to-point latency performance on c6gn instances:
MVAPICH2 Libraries on AWS EFA

- Collective Performance on 32 AWS c6gn.16xlarge instances

![Allreduce - 64-ppn](chart1)

![Scatter – 64-ppn](chart2)
OCI HPC Workloads

• MVAPICH2 MPI libraries recently added scalable and reliable performance support on OCI HPC Clouds
  − Based on BM.HPC2 instance shape & Oracle Linux 7 HPC Image
  − Dedicated performance tuning
  − Automatic XPMEM kernel module detection

• Experiment setups:
  − Instance Shape: BM.HPC2.36
  − OCI Cluster Network version: oci-hpc v2.6.3
  − Versions of MPI Libraries:
    • MVAPICH2-X 2.3
    • HPCx-2.8.1 (Built-in module of OCI HPC Image)
    • Intel MPI 2021
MPI-level Performance on OCI HPC System

• Point-to-point performance evaluation

[Graphs showing inter-node latency and bandwidth for different message sizes, comparing MVAPICH2-X and IntelMPI.]
MPI-level Performance on OCI HPC System

• Collective performance evaluation on 8 BM.HPC2 instances
Application Level Performance on OCI HPC System

![Graph showing total execution time for different processes on different MPI implementations. The graph indicates that MVAPICH2-X is 4x faster than other implementations.]
Work in Progress & Future Plans

• **AWS:**
  - Further optimization in larger scale collective operations on AWS ARM instances
  - Finalizing our optimized design based on GDR and RDMA feature on P4D instances with A100 GPUs

• **Oracle:**
  - Further performance optimization for other OCI HPC instance shapes
  - More user-friendly deployment for MVAPICH2 MPI library stacks
Thank You!

xu.2452@osu.edu

Network-Based Computing Laboratory

http://nowlab.cse.ohio-state.edu/

Follow us on

https://twitter.com/mvapich

The High-Performance MPI/PGAS Project
http://mvapich.cse.ohio-state.edu/

The High-Performance Big Data Project
http://hibd.cse.ohio-state.edu/

The High-Performance Deep Learning Project
http://hidl.cse.ohio-state.edu/