Experiences with MVAPICH2 and INAM on SDSC Systems

Mahidhar Tatineni
Director, User Services, SDSC

MVAPICH2 User Group (MUG) Meeting
August 23, 2023

Outline

- Overview
- MVAPICH2 and MVAPICH2-GDR on Expanse
  - System architecture
  - Applications
  - Benchmark results
- MVAPICH2 on Voyager
  - System architecture
  - Kubernetes based usage of MVAPICH2
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
- Summary

Overview

- SDSC operates and manages software for several HPC and AI systems with a large userbase spanning many scientific domains.
- MVAPICH2 and MVAPICH2-GDR have been an integral part of many SDSC HPC systems over the past decade including Trestles, Gordon, Comet, and Expanse.
- MVAPICH2 is being tested on machines with innovative architectures as well. This includes the Voyager heterogenous system, designed for AI workloads, which is currently in its 3-year testbed phase.
- INAM is in use on Comet to monitor network health and to identify sources of congestion.

Outline

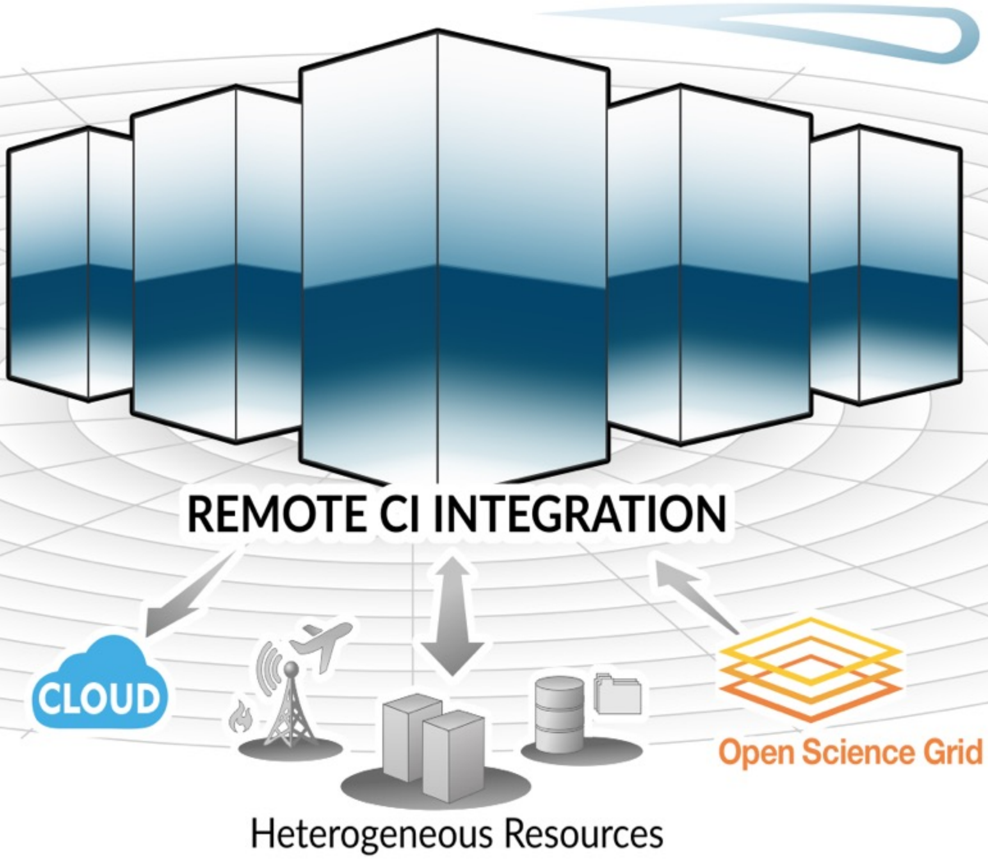
- Overview
- MVAPICH2 and MVAPICH2-GDR on Expanse
  - System architecture
  - Applications
  - Benchmark results
- MVAPICH2 on Voyager
  - System architecture
  - Kubernetes based usage of MVAPICH2
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
- Summary

EXPANSE

COMPUTING WITHOUT BOUNDARIES
5 PETAFLOP/S HPC and DATA RESOURCE

HPC RESOURCE
13 Scalable Compute Units
728 Standard Compute Nodes
52 GPU Nodes: 208 GPUs
4 Large Memory Nodes

DATA CENTRIC ARCHITECTURE
12PB Perf. Storage: 140GB/s, 200k IOPS
Fast I/O Node-Local NVMe Storage
7PB Ceph Object Storage
High-Performance R&E Networking



LONG-TAIL SCIENCE
Multi-Messenger Astronomy
Genomics
Earth Science
Social Science

INNOVATIVE OPERATIONS
Composable Systems
High-Throughput Computing
Science Gateways
Interactive Computing
Containerized Computing
Cloud Bursting

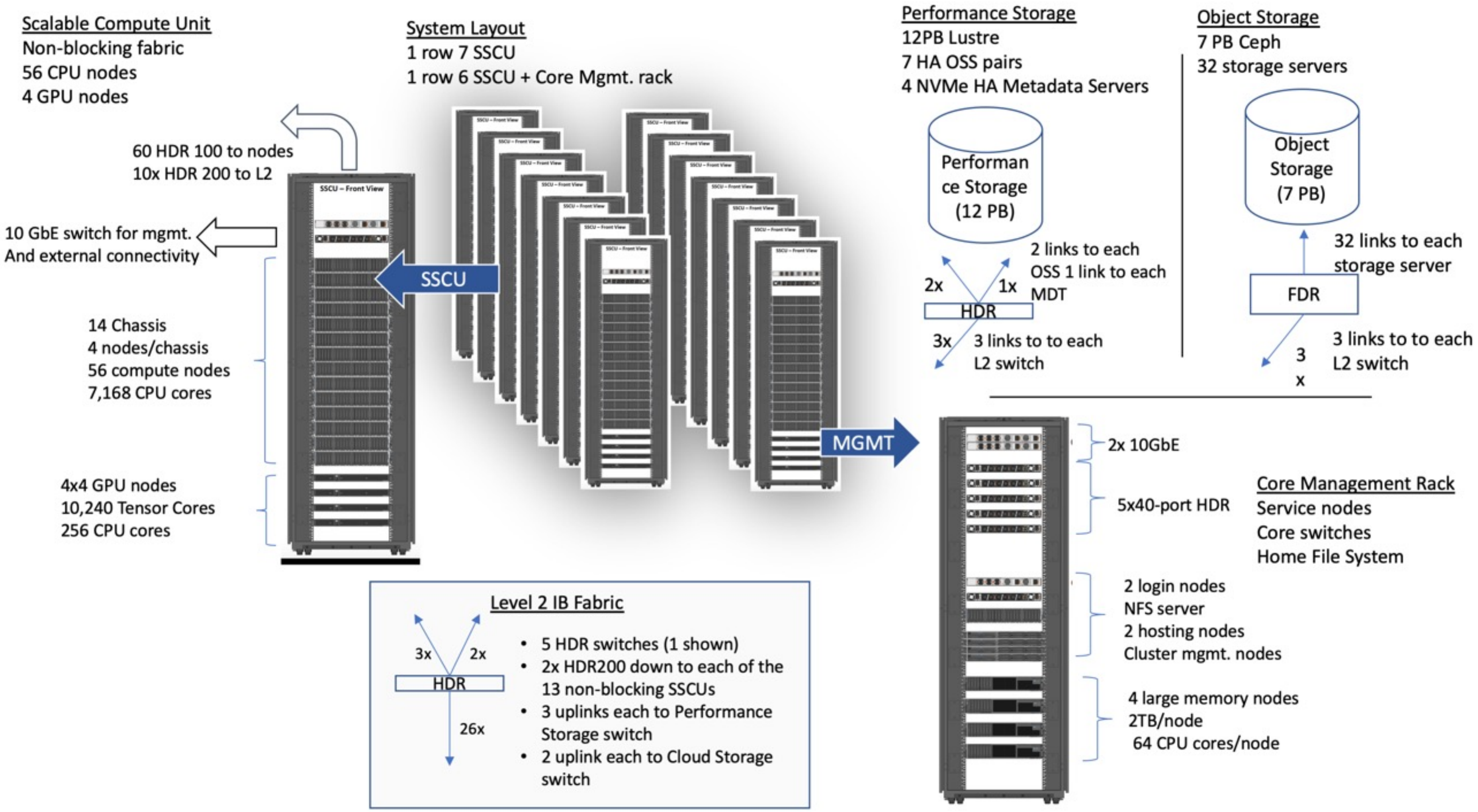
NSF Award # 1928224

PIs: Mike Norman (PI), Ilkay Altintas, Amit Majumdar, Mahidhar Tatineni, Shawn Strande

Expanse is a heterogeneous architecture designed for high performance, reliability, flexibility, and productivity

System Summary

- *14 SDSC Scalable Compute Units (SSCU)*
- *784 x 2s Standard Compute Nodes*
- *100,352 Compute Cores*
- *200 TB DDR4 Memory*
- *56x 4-way GPU Nodes w/NVLINK*
- *224 V100s*
- *4x 2TB Large Memory Nodes*
- *HDR 100 non-blocking Fabric*
- *12 PB Lustre High Performance Storage*
- *7 PB Ceph Object Storage*
- *1.2 PB on-node NVMe*
- *Dell EMC PowerEdge*
- *Direct Liquid Cooled*



The SSCU is Designed for the Long Tail Job Mix, Maximum Performance, Efficient Systems Support, and Efficient Power and Cooling

Standard Compute Nodes

- *2x AMD EPYC 7742 @2.25 GHz*
- *128 Zen2 CPU cores*
- *PCIe Gen4*
- *256 GB DDR4*
- *1 TB NVME*

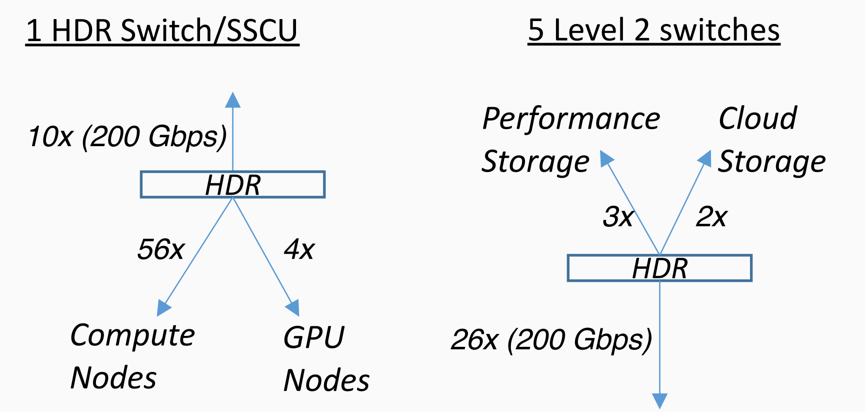
GPU Nodes

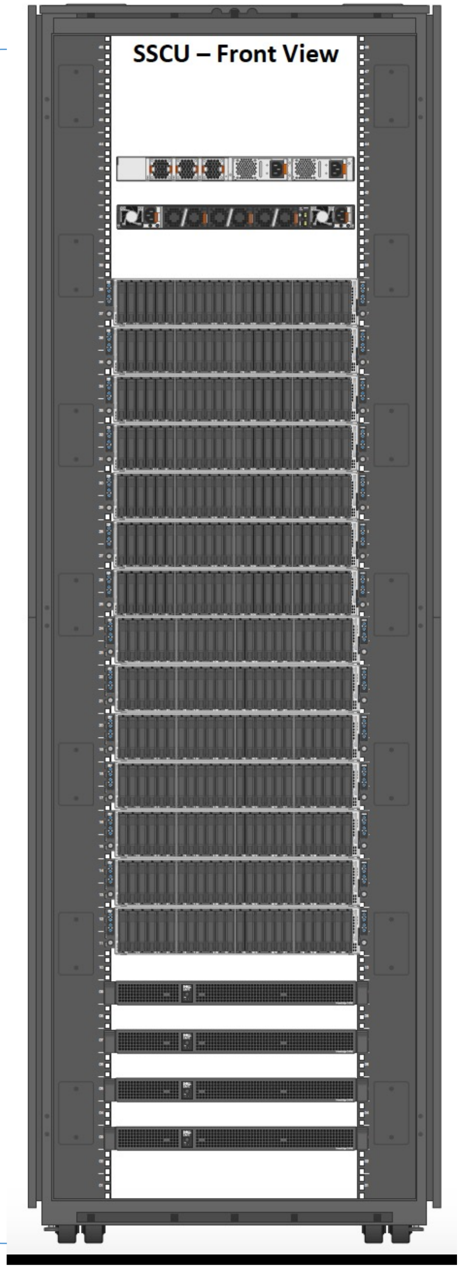
- *4x NVIDIA V100/follow-on*
- *10,240 Tensor Cores*
- *32 GB GDDR*
- *1.6 TB NVMe*
- *Intel CPUs*

SSCU Components

- *56x CPU nodes*
- *7,168 Compute Cores*
- *4x GPU nodes*
- *1x HDR Switch*
- *1x 10GbE Switch*
- *HDR 100 non-blocking fabric*
- *Wide rack for serviceability*
- *Direct Liquid Cooling to CPU nodes*

Non-blocking Interconnect





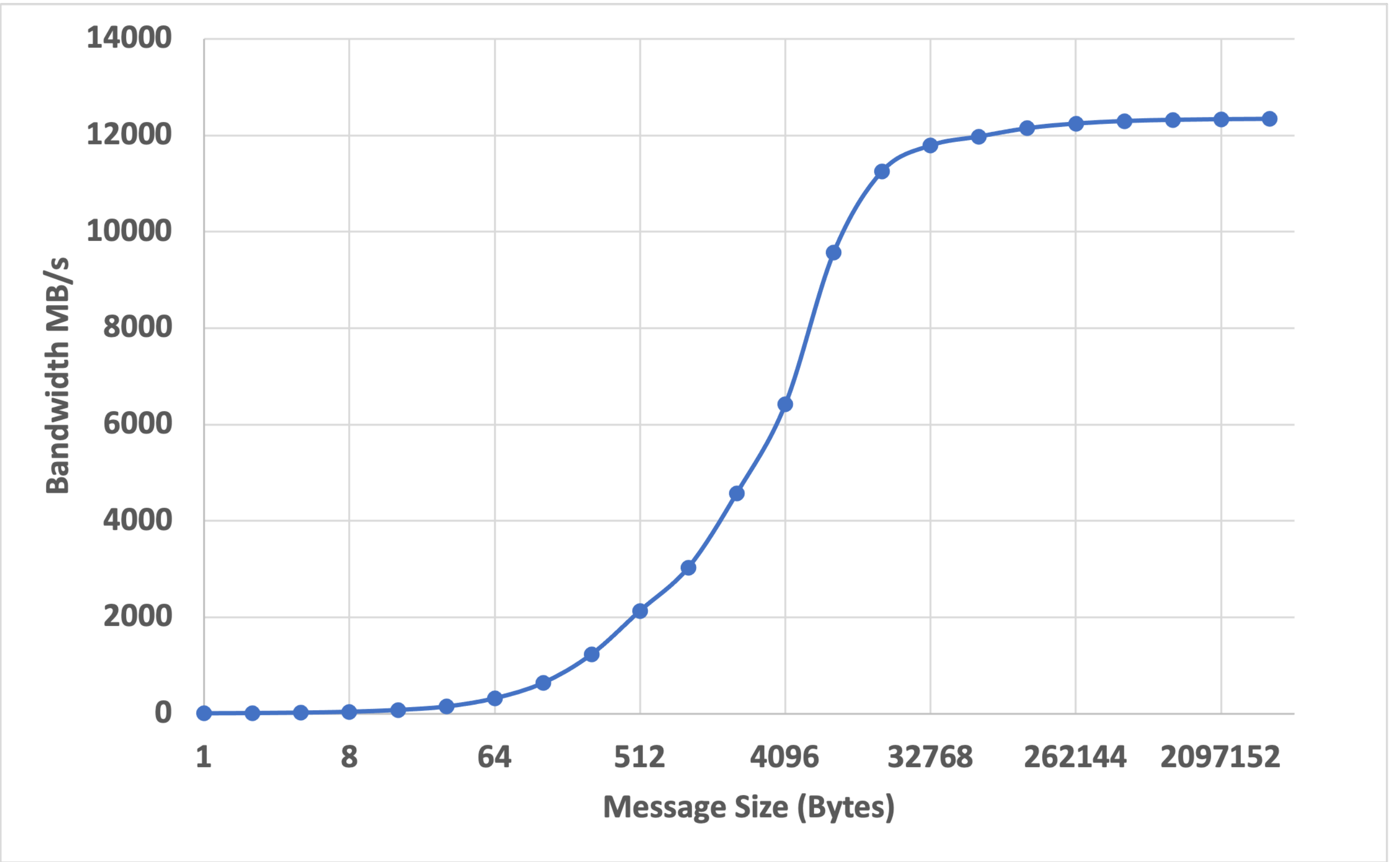
Expanse GPU Node Architecture

- 4 V100 32GB SMX2 GPUs
- 384 GB RAM, 1.6 TB PCIe NVMe
- 2 Intel Xeon 6248 CPUs
- Topology:

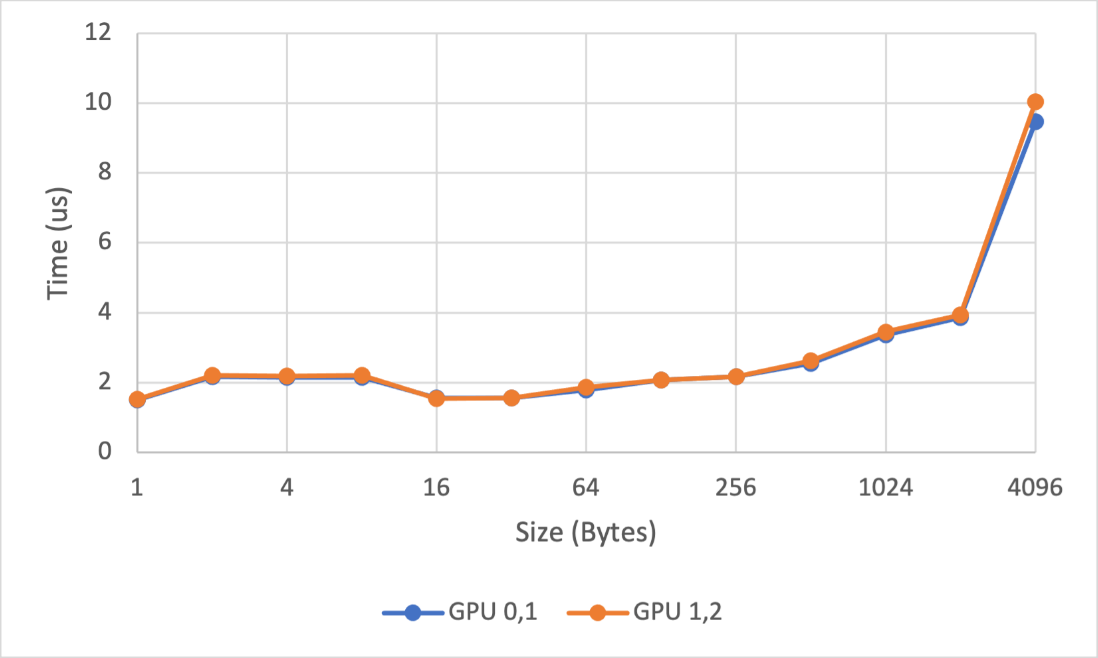
```
        GPU0    GPU1    GPU2    GPU3    mlx5_0  CPU Affinity
GPU0     X      NV2     NV2     NV2     SYS     0-0,4-4,8-8,12-12,16-16,20-20,24-24,28-28,32-32,36-36
GPU1    NV2      X      NV2     NV2     SYS     0-0,4-4,8-8,12-12,16-16,20-20,24-24,28-28,32-32,36-36
GPU2    NV2     NV2      X      NV2     SYS     1-1,5-5,9-9,13-13,17-17,21-21,25-25,29-29,33-33,37-37
GPU3    NV2     NV2     NV2      X      SYS     1-1,5-5,9-9,13-13,17-17,21-21,25-25,29-29,33-33,37-37
mlx5_0  SYS     SYS     SYS     SYS      X

Legend:

  X    = Self
  SYS  = Connection traversing PCIe as well as the SMP interconnect between NUMA nodes (e.g., QPI/UPI)
  NODE = Connection traversing PCIe as well as the interconnect between PCIe Host Bridges within a NUMA node
  PHB  = Connection traversing PCIe as well as a PCIe Host Bridge (typically the CPU)
  PXB  = Connection traversing multiple PCIe bridges (without traversing the PCIe Host Bridge)
  PIX  = Connection traversing at most a single PCIe bridge
  NV#  = Connection traversing a bonded set of # NVLinks
```

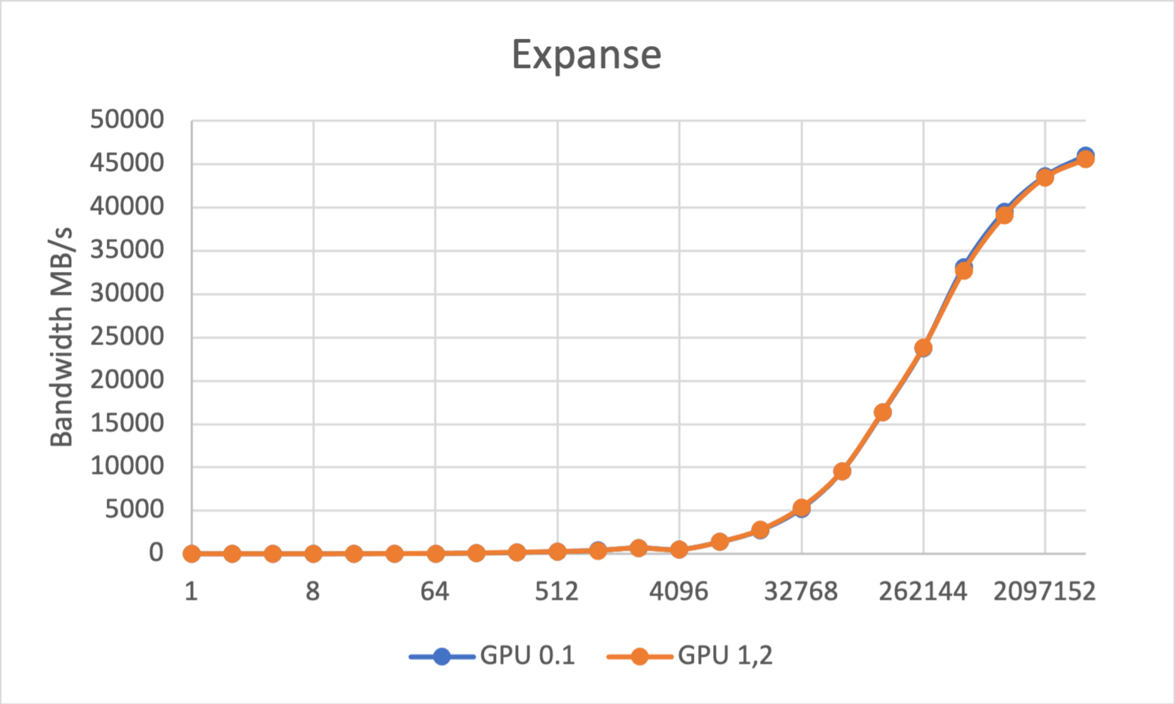
OSU Bandwidth Benchmark (osu_bw): Inter-node test

MVAPICH2 version 2.3.7



OSU Latency and Bandwidth (osu_latency, osu_bw) Benchmark Intra-node, V100 nodes on Expanse

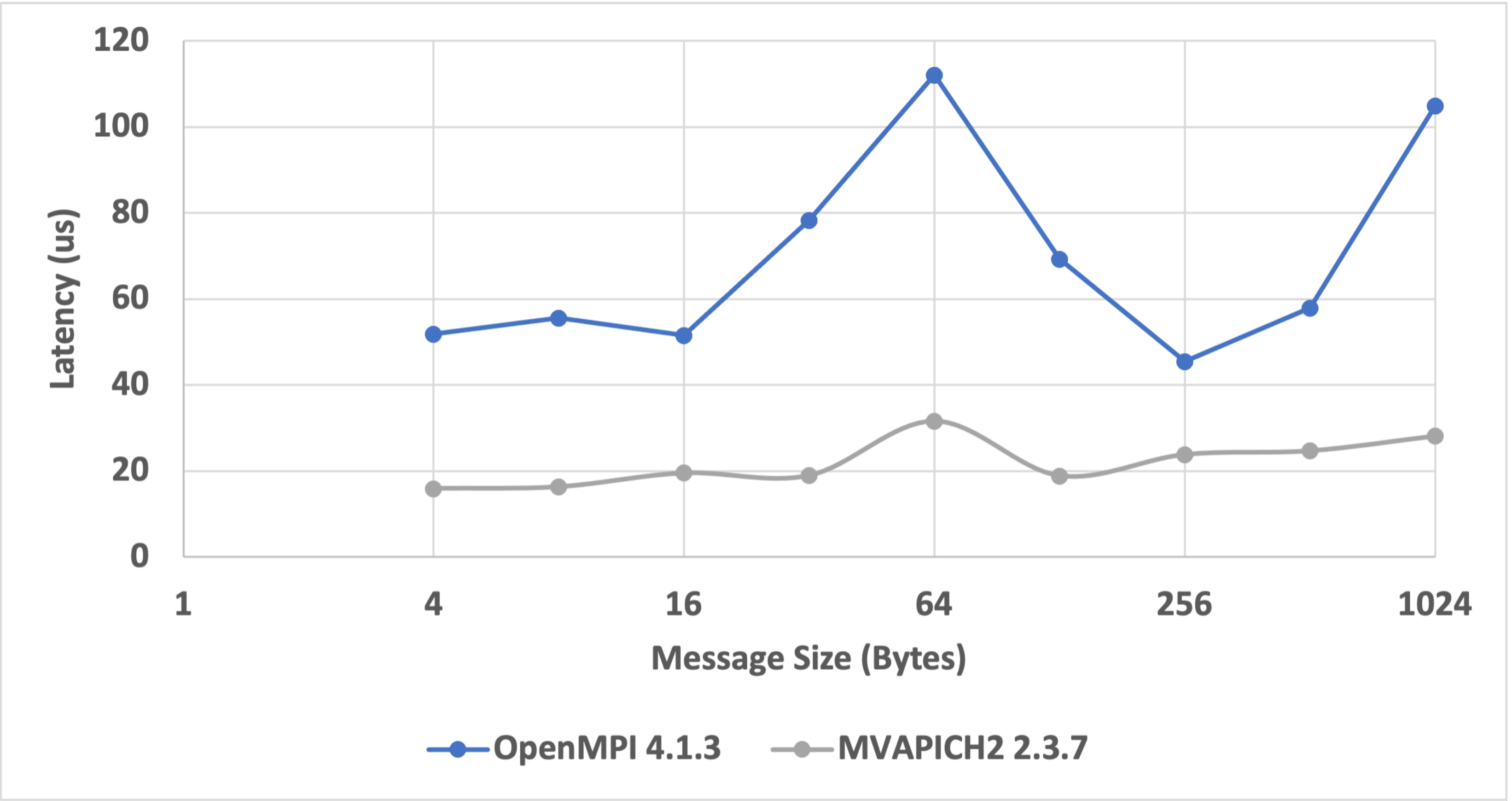




- *Expanse - V100 nodes*
- *Latency between GPU 0 , GPU 1: 1.51 μs*
- *Latency between GPU 1 , GPU 2: 1.53 μs*
- *MVAPICH2 GDR 2.3.6, GCC 8.3.1*

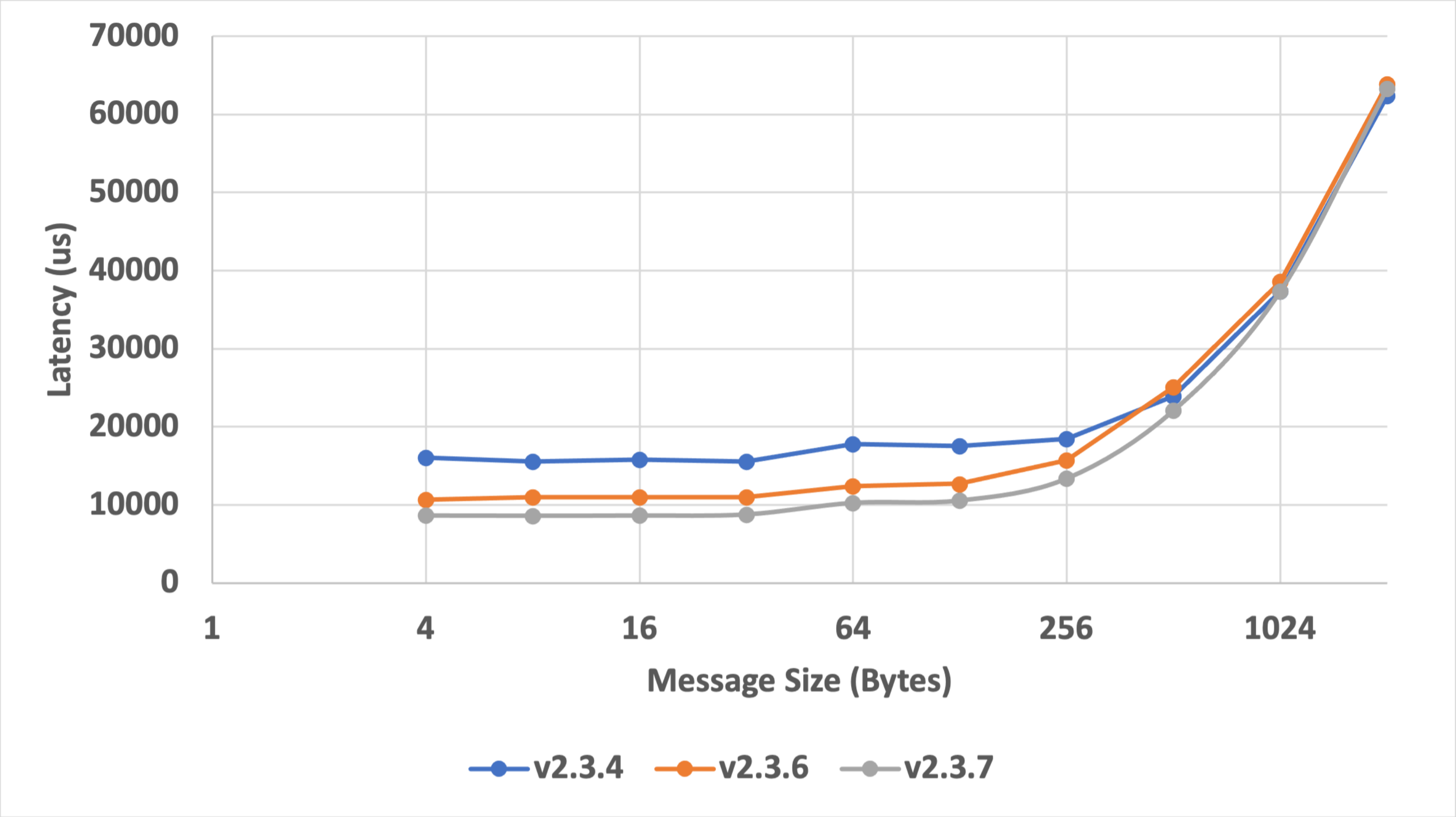
OSU Allreduce benchmark (osu_allreduce): 4096 cores

MVAPICH2 version 2.3.7, OpenMPI 4.1.3



OSU Alltoallv benchmark (osu_alltoallv): 2048 cores

MVAPICH2 version 2.3.4 through 2.3.7

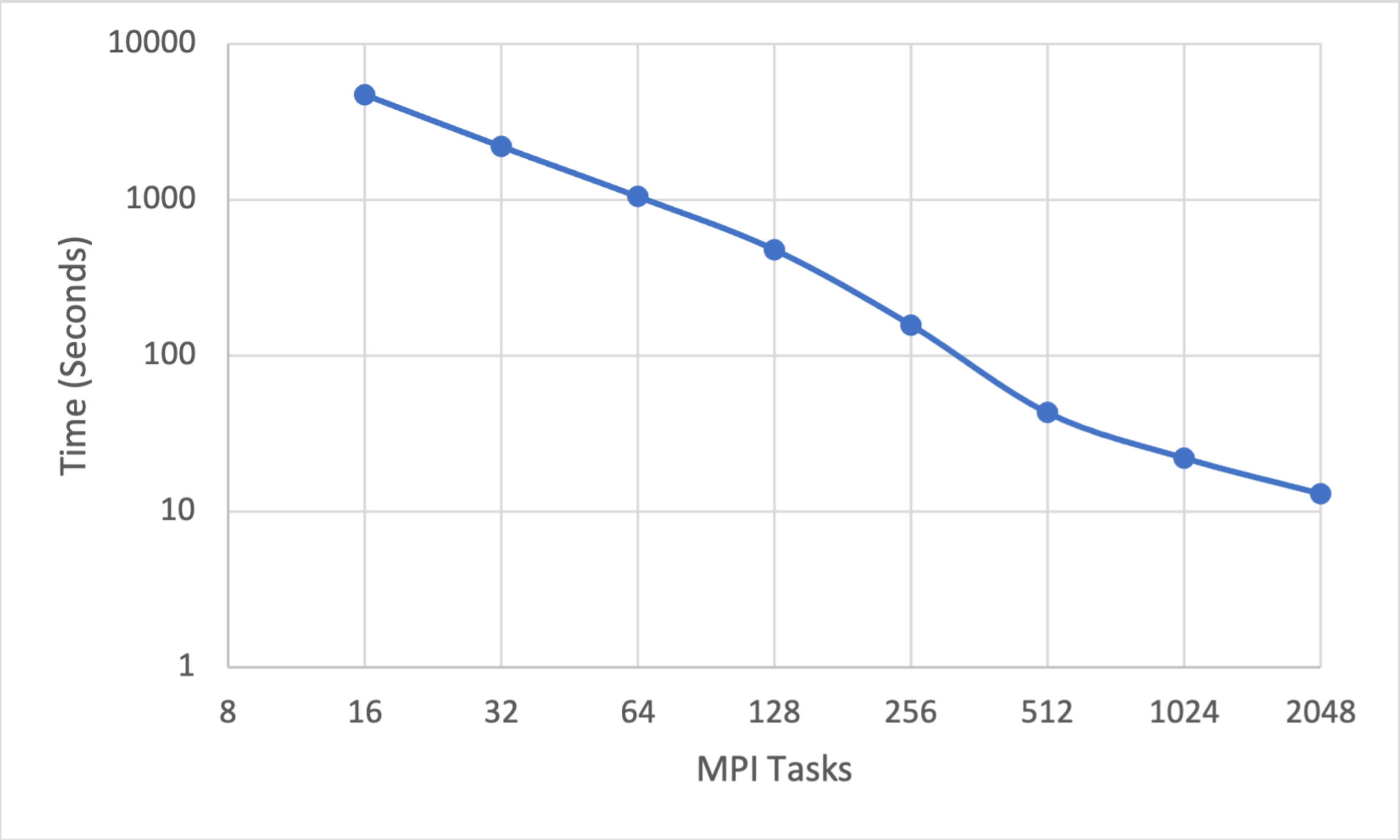


MVAPICH2 based Application Installs on Expanse

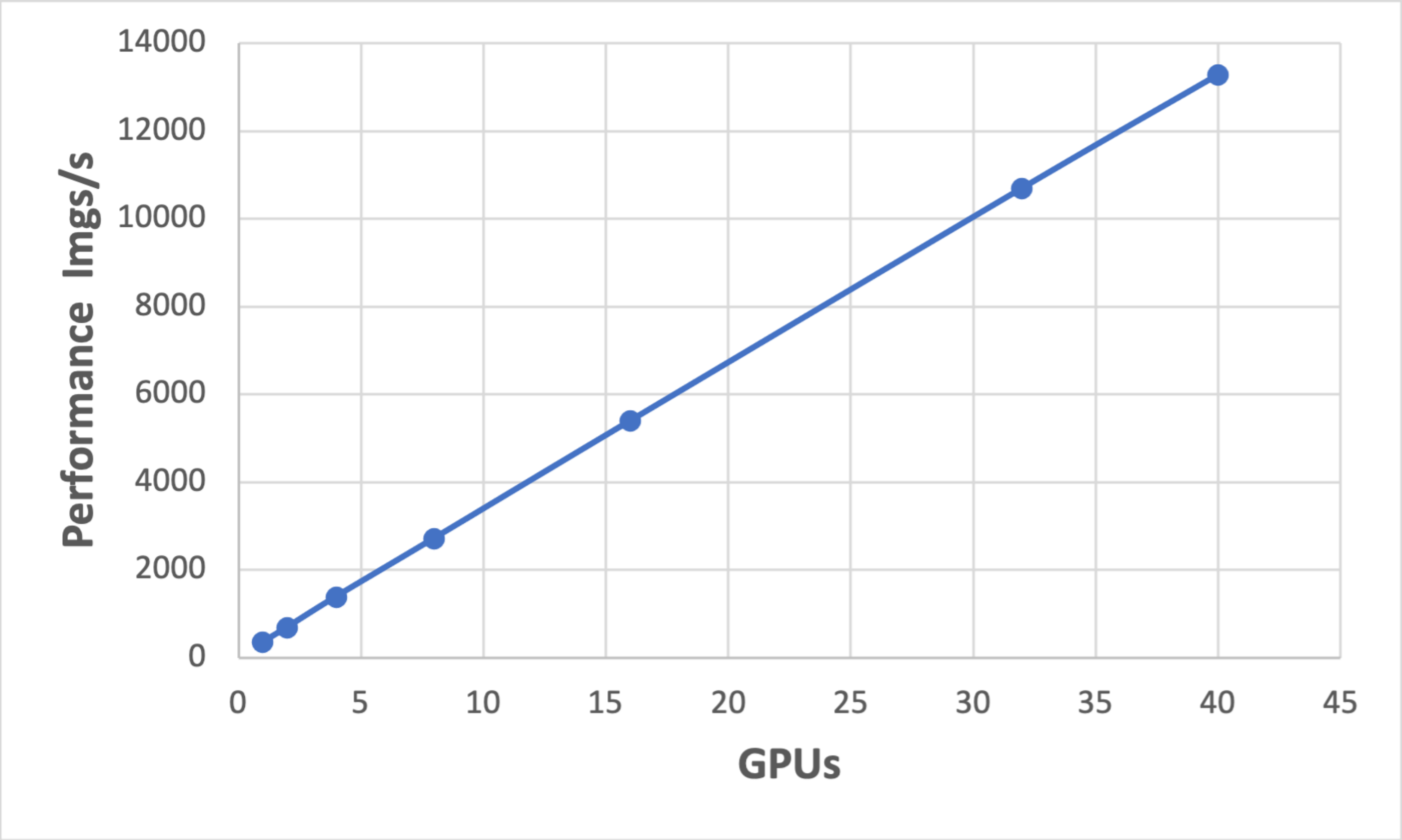
Application	Description
RAxML	Code for sequential and parallel Maximum Likelihood based inference of large phylogenetic trees
Q-Chem	Commercial package for comprehensive ab initio quantum chemistry software for accurate predictions of molecular structures,
AMBER	Suite of biomolecular simulation programs
LAMMPS	Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) Classical molecular dynamics code with a focus on materials modeling
NAMD	Parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems
ABINIT	Open-source software suite to calculate the optical, mechanical, vibrational, and other observable properties of materials
NEURON	Simulation environment for modeling individual and networks of neurons
TensorFlow w/ Horovod	Open-source platform for machine learning
PyTorch w/ Horovod	Open-source machine learning framework

Additionally, libraries installed: e.g. hypre, fftw, hdf5, netcdf, netcdf, ncview

NEURON Benchmark:
Large-scale model of olfactory bulb: 10,500 cells, 40K timesteps



ResNet50 Benchmark: MVAPICH2-GDR 2.3.7, Horovod, V100 GPUs on Expanse



AWP-ODC Performance on Expanse

Expanse (640x320x2048)	Teraflop/s	Time (sec/step)
gcc+mvapich2/2.3.7-gdr (2 nodes, 8 GPUs, A100) (non-cuda aware)	3.65	0.0361
gcc+mvapich2/2.3.7-gdr (2 nodes, 8 GPUs, A100)	4.03	0.0326
gcc+mvapich2/2.3.7-gdr (2 nodes, 8 GPUs, V100) (non-cuda aware)	2.40	0.0548
gcc+mvapich2/2.3.7-gdr (2 nodes, 8 GPUs, V100)	3.83	0.0342
Expanse (640x640x2048)	Teraflop/s	Time (sec/step)
gcc+mvapich2/2.3.7-gdr (4 nodes, 16 GPUs, V100) (non-cuda aware)	4.42	0.0599
gcc+mvapich2/2.3.7-gdr (4 nodes, 16 GPUs, V100)	6.09	0.0433

Credit: Dr. Yifeng Cui, Qinghua Zhou
Full results on MUG 2023 talk by Yifeng

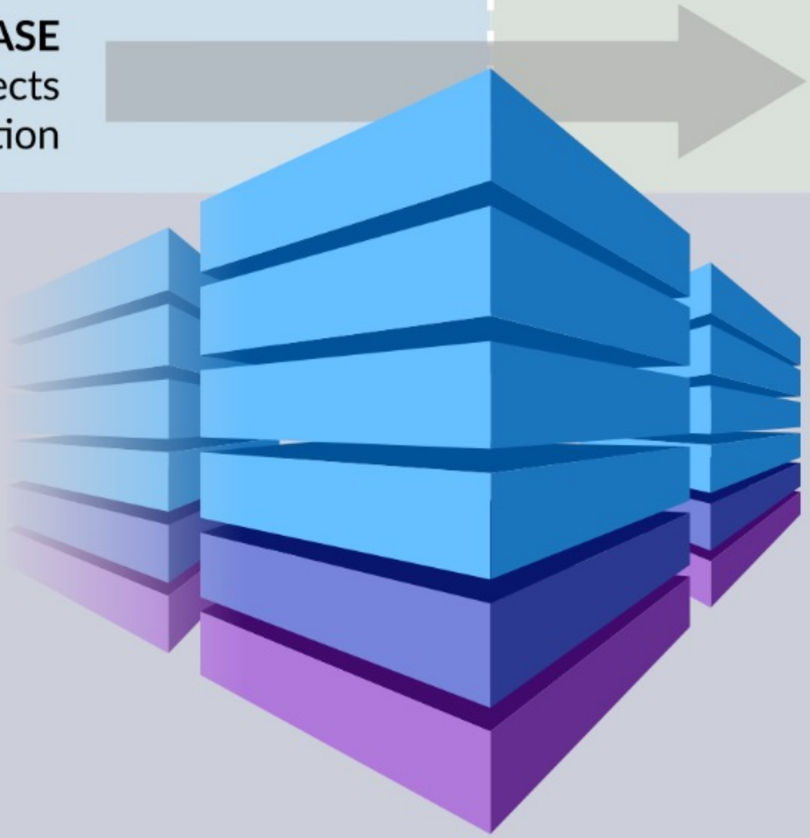
Outline

- Overview
- MVAPICH2 and MVAPICH2-GDR on Expanse
  - System architecture
  - Applications
  - Benchmark results
- MVAPICH2 on Voyager
  - System architecture
  - Kubernetes based usage of MVAPICH2
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
- Summary

VOYAGER
EXPLORING AI PROCESSORS in SCIENCE and ENGINEERING

3-YEAR TESTBED PHASE
Focused Select Projects
Workshops, Industry Interaction

2-YEAR ALLOCATIONS PHASE
NSF Allocations to the Broader Community
User Workshops



INNOVATIVE AI RESOURCE
Specialized Training Processors
Specialized Inference Processors
High-Performance Interconnect
X86 Standard Compute nodes
Rich Storage Hierarchy

OPTIMIZED AI SOFTWARE
Community Frameworks
Custom user-developed AI Applications
PyTorch, Tensorflow

IMPACT & ENGAGEMENT
Large-Scale Models
AI Architecture Advancement
Improved Performance of AI Applications
External Advisory Board of AI & HPC Experts
Wide Science & Engineering Community
Advanced Project Support & Training
Accelerating Scientific Discovery
Industrial Engagement

Category II System, NSF Award # 2005369

PI: Amit Majumdar (SDSC); Co PIs: Rommie Amaro (UCSD), Javier Duarte (UCSD), Mai Nguyen (SDSC), Robert Sinkovits (UCSD)

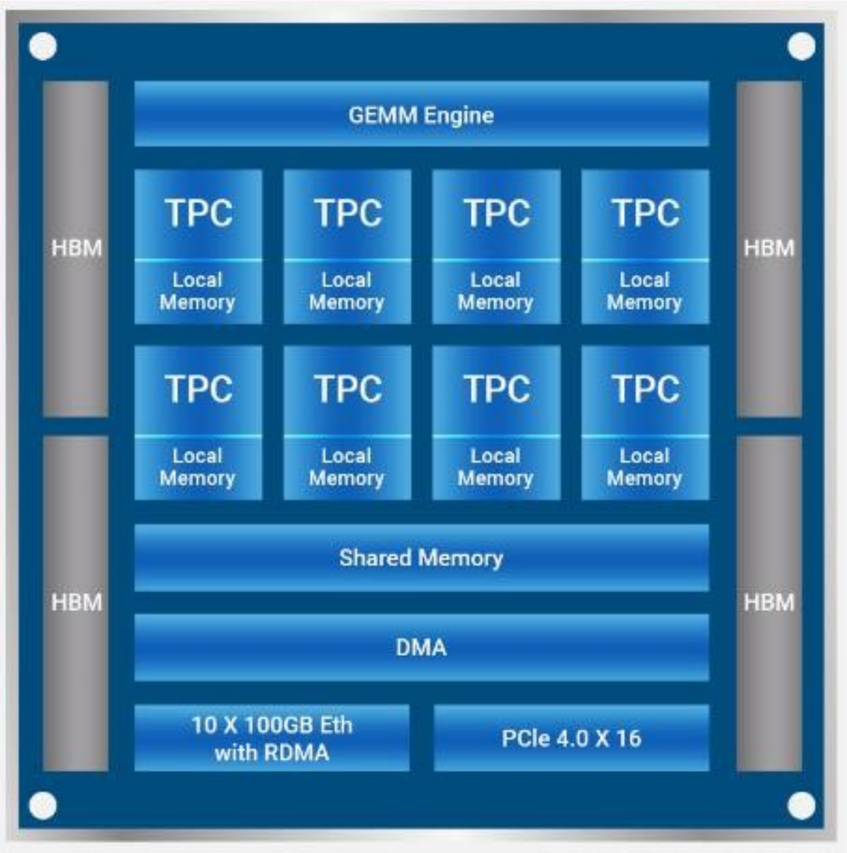
Voyager is a heterogeneous system designed to support complex AI workflows

- 42x Intel Habana Gaudi training nodes, each with 8 training processors (336 in total); all-to-all network between processors on a node
- Gaudi processors feature specialized hardware units for AI, HBM2, and on-chip high-speed Ethernet
- 2x Goya inference nodes, each with 8 Goya processors (16 in total)
- 36x Intel x86 processors compute nodes for general purpose computing and data processing
- 400 GbE interconnect using RDMA over Converged Ethernet
- 3 PB Storage system connected vis 25GbE. Deployed as Ceph, but open to others
- 324 TB HFS; connectivity to compute via 25GbE

System Component	*Configuration*
INTEL GAUDI TRAINING NODES	
Node count	42
Training processors/node	8
Host x86 processors/node	2
Memory/node	512 GB DDR4
Memory/training processor	32 GB HBM2
Local NVMe	6.4 TB
INTEL GOYA INFERENCE NODES	
Node count	2
Inference processors/node	8
Host x86 processors/node	2
Memory/node	512 GB DDR4
Memory/inference processor	16 GB DDR4
Local NVMe	3.2 TB
STANDARD COMPUTE NODES	
Node count	36
x86 processors/node	2
Memory capacity	384 GB
Local NVMe	3.2 TB
STORAGE SYSTEM	
High performance storage: HDD:NVMe	3 PB:140 TB
High performance filesystems	Ceph, Lustre
Home filesystem storage: HDD:NVMe	324 TB: 12.4 TB
File system	NFS

Gaudi: Architected for performance and efficiency

- *Fully programmable Tensor Processing Cores (TPC) with tools & libraries*
- *Configurable Matrix Math Engine (GEMM)*
- *Multi-stage memory hierarchy with 32GB HBM2 memory*
- *Integrated 10 x 100 Gigabit Ethernet for multi-chip scale-out training*

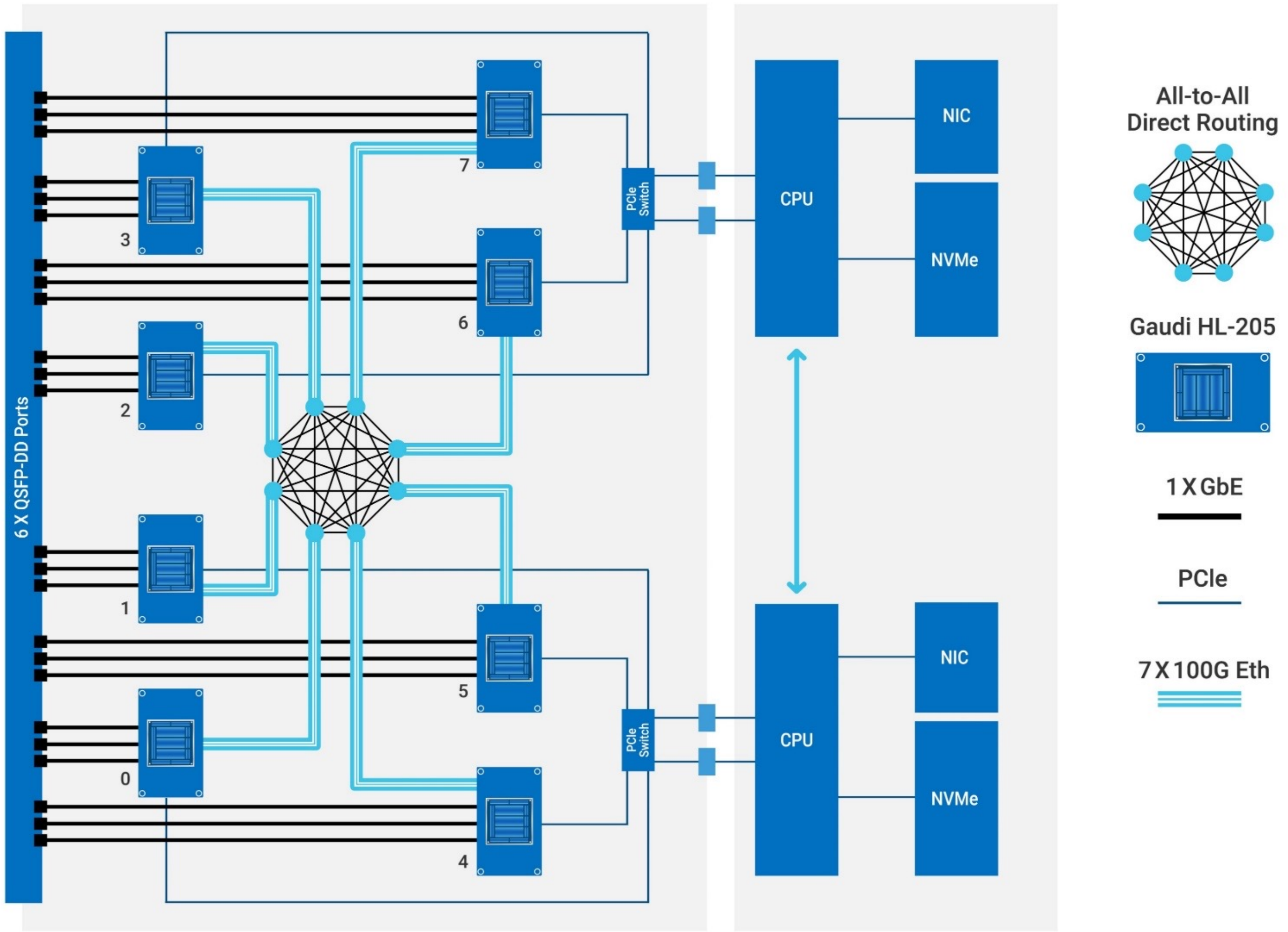


Voyager's three distinct networks support application performance, data movement, and systems management

- 400 GbE for scale-out training
  - Six connections from each Gaudi node to a single 230 Tbps Arista 7808 non-blocking switch
- 25 GbE Bonded Control and data network
  - Every node has a bonded 50GbE (2 X 25GbE) connection.
- 1 GbE out-of-band management network for IPMI and other traffic.

Gaudi servers supports all-to-all connectivity

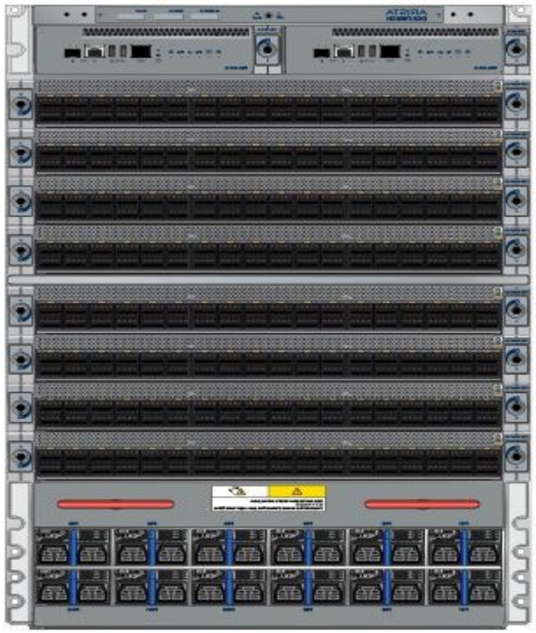
- 8 Gaudi OCP OAM cards
- 24 x 100GbE RDMA RoCE for scale-out
- Non-blocking, all-2-all internal interconnect across Gaudi AI processors
- Separate PCIe ports for external Host CPU traffic



Example of Integrated Server with eight Gaudi AI processors, two Xeon CPU and multiple Ethernet Interfaces

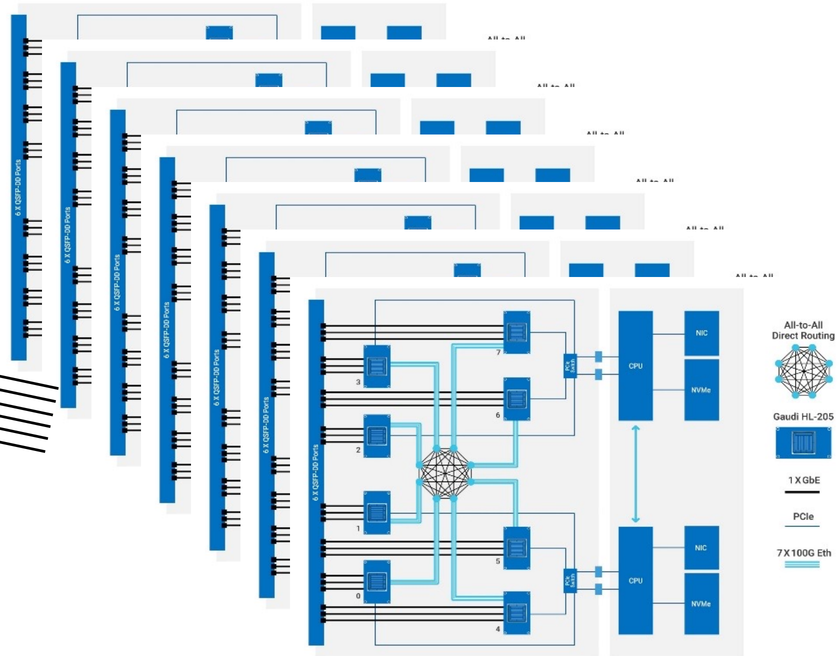
Gaudi design enables highly efficient scaling

- *Natively integrated RoCE on Gaudi processor*
- *6x Quad-100 GbE per node (8x Gaudi)*
- *7808 Arista 400 GbE switch*



7803 Arista 400GbE

6x 400GbE per Gaudi Node

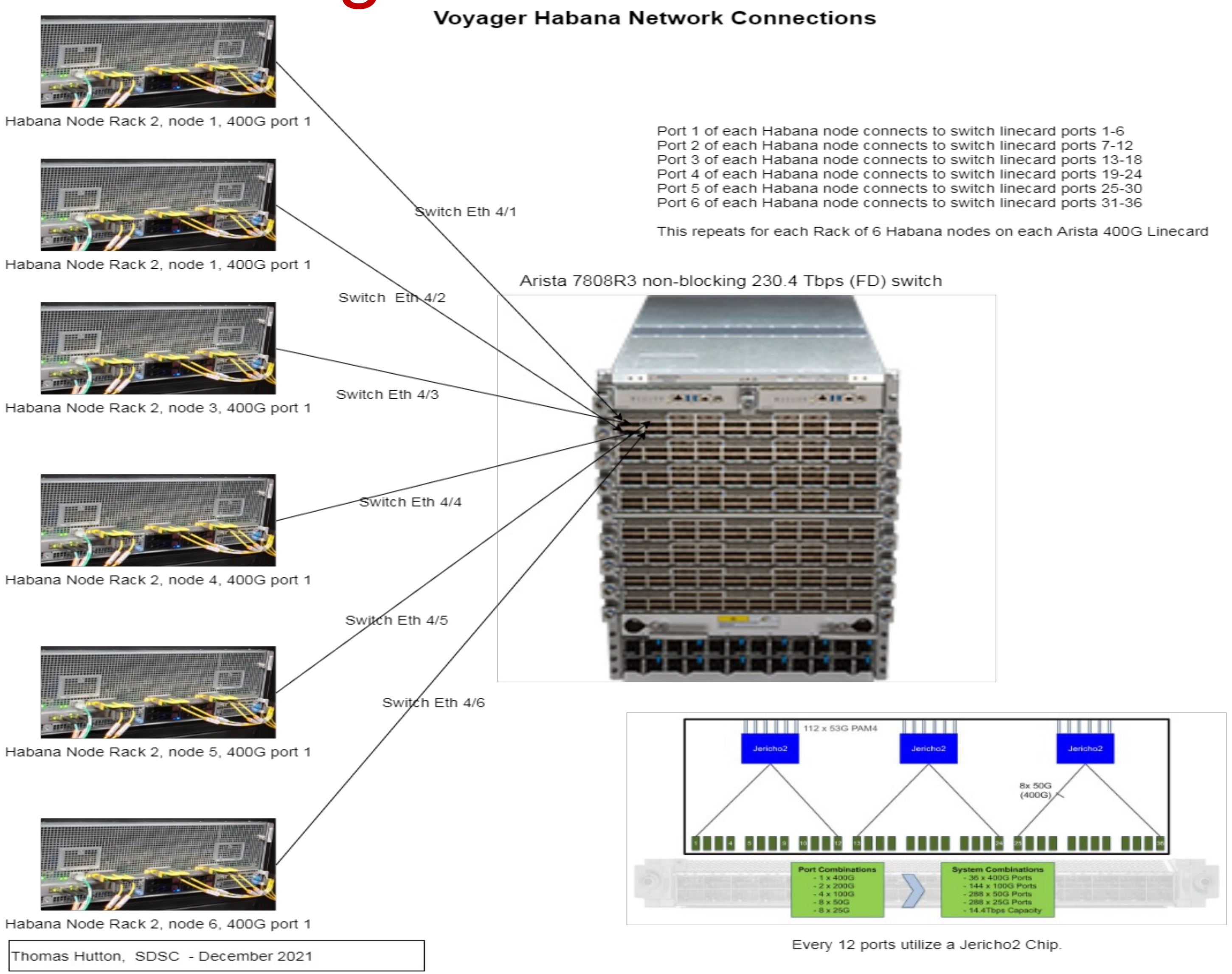


6x Gaudi nodes per rack

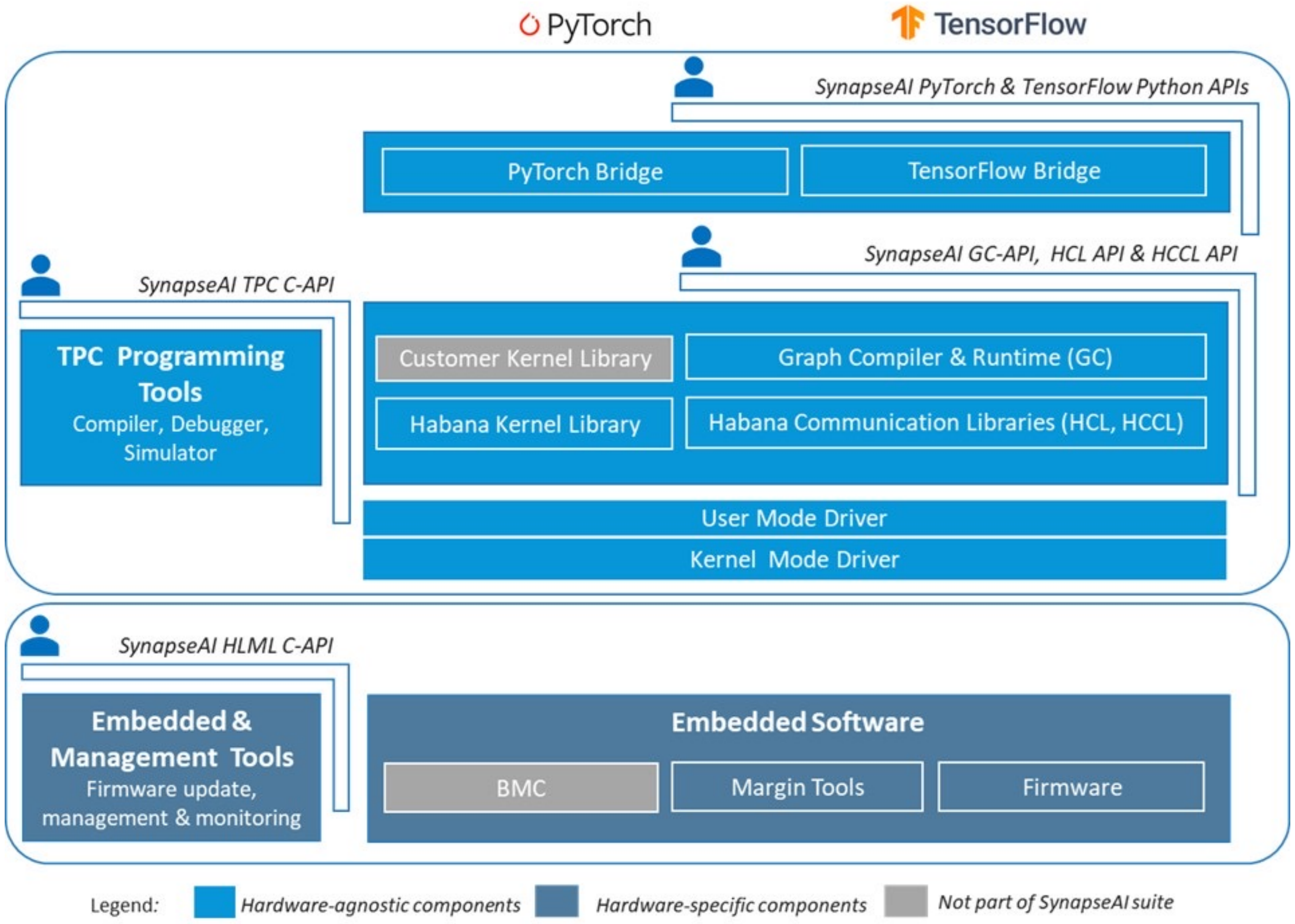
400GbE scale-out networking

6x400 connections from each Gaudi node to Arista 400GbE non-blocking switch

200 Tb/s of bandwidth (42x6x400)



Software Stack Overview: Synapse AI

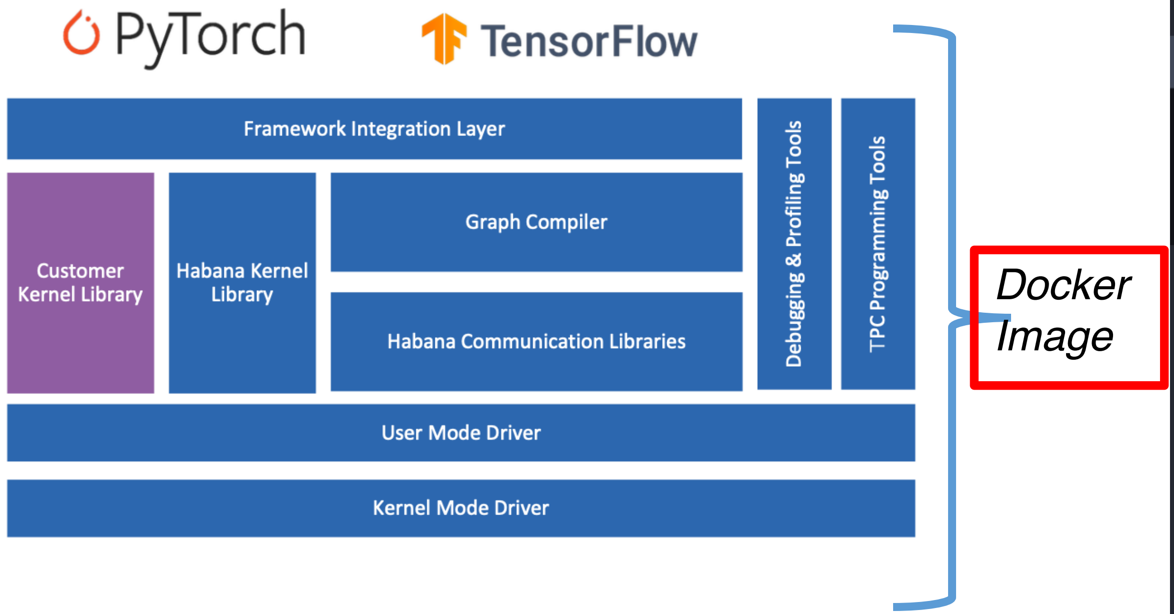


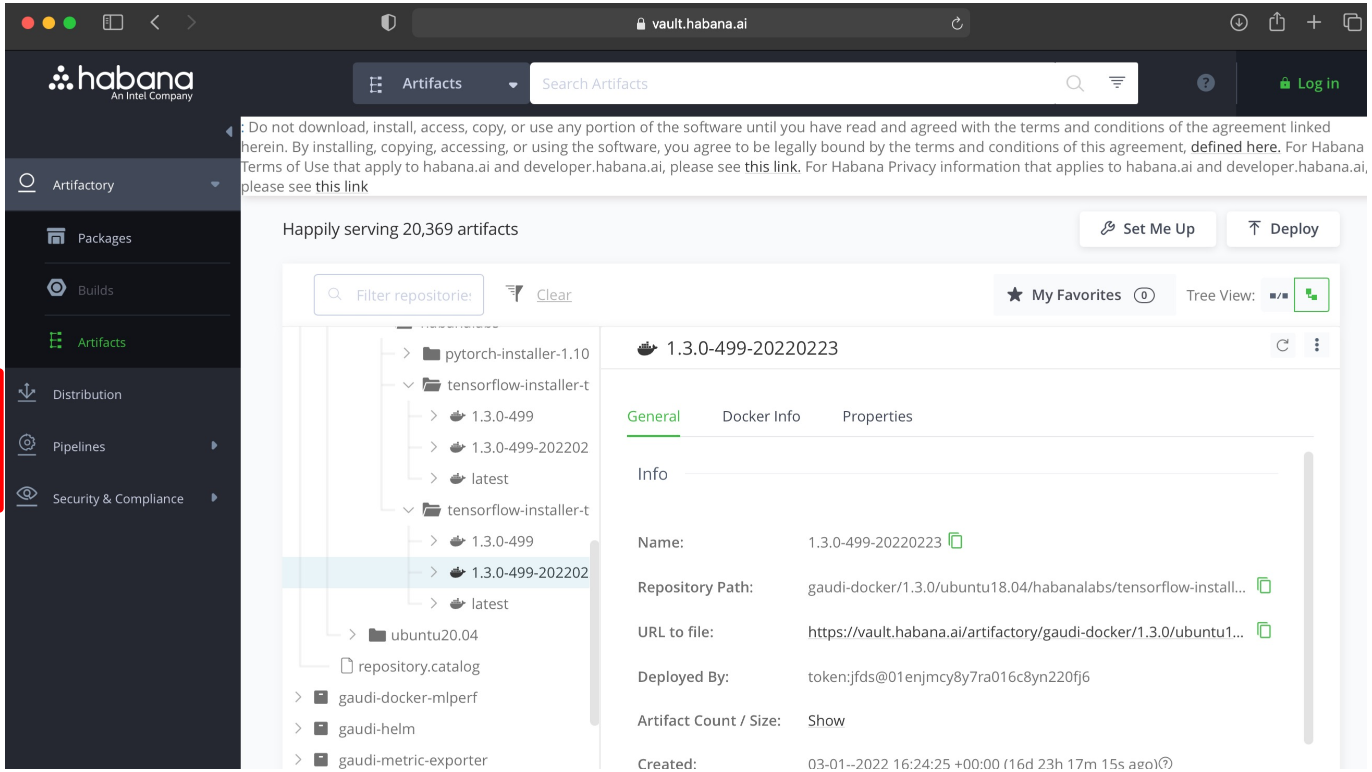
Reference: *https://docs.habana.ai/en/latest/Gaudi_Overview/Gaudi_Overview.html#synapseai-software-suite*

- *Train deep learning models on Gaudi with minimal code changes*
- *Natively integrated with TensorFlow & PyTorch*
- *Popular computer vision and NLP models*
- *Habana Developer Site & GitHub*
- *Support with reference models, kernel*
- *libraries, docs & "how-to" guides*
- *Advanced users can write their own custom software kernels*

Containerized Software Stack

https://vault.habana.ai





Reference: *https://docs.habana.ai/en/latest/Gaudi_Overview/Gaudi_Overview.html#synapseai-software-suite*

Experience using MVAPICH2 on Voyager

- Dr. Panda's group provided the MVAPICH2 complied install binaries and libraries. The install is done to host path location mounted within a Kubernetes pod making it easily usable by others.
- MPI jobs on Voyager are run using the Kubeflow MPI operator. The MVAPICH2 install can be bind mounted into the right location in the pods:

```
volumeMounts:
- name: home
  mountPath: /home/mahidhar
- name: mv2install
  mountPath: /usr/local/apps/mvapich2gdr
```

- With this mount addition, the install is usable by all users with host access to the install location.

MVAPICH2 on Voyager

```yaml
    replicas: 2
    template:
      spec:
        restartPolicy: Never
        serviceAccountName: mahidhar
        nodeSelector:
          brightcomputing.com/node-category: 'gaudi'
        hostNetwork: false
        volumes:
        - name: home
          hostPath:
            path: /home/mahidhar
            type: Directory
        - name: scratch
          emptyDir: {}
        - name: mv2install
          hostPath:
            path: /home/mahidhar/mvapich2gdr
            type: Directory
        containers:
        - name: i70-ubuntu2004
          image: vault.habana.ai/gaudi-docker/1.7.0/ubuntu20.04/habanalabs/tensorflow-installer-tf-cpu-2.8.3:lates
          imagePullPolicy: Always
          resources:
            requests:
              cpu: 48
              memory: 256Gi
              habana.ai/gaudi: 8
              hugepages-2Mi: 96000Mi
            limits:
              cpu: 96
              memory: 368Gi
              habana.ai/gaudi: 8
              hugepages-2Mi: 96000Mi
          volumeMounts:
          - name: home
            mountPath: /home/mahidhar
          - name: scratch
            mountPath: /scratch
          - name:  mv2install
            mountPath: /usr/local/apps/mvapich2gdr
```

MVAPICH2 Install location on host

Container image with Synapse AI software

MVAPICH2 mount location in Kubernetes worker pods

MVAPICH2 on Voyager

```
~/shared/MPI# /usr/local/apps/mvapich2gdr/bin/mpirun_rsh -np 8 -hostfile /home/mahidhar/hosts MV2_ENABLE_GPU=1 MV2_IS_PT2PT
=0 LD_PRELOAD=/usr/local/apps/mvapich2gdr/lib/libmpi.so /usr/local/apps/mvapich2gdr/libexec/osu-micro-benchmarks/mpi/collec
tive/osu_allreduce -d synp
 * Starting OpenBSD Secure Shell server sshd
   ...done.

# OSU MPI-SYNP Allreduce Latency Test v7.0
# Size       Avg Latency(us)
4                      79.67
8                      75.30
16                     74.61
32                     75.79
64                     76.80
128                    75.29
256                    77.00
512                    74.93
1024                   72.27
2048                   71.40
4096                   72.53
8192                   72.66
16384                  77.78
32768                  73.96
65536                  74.68
131072                 75.96
262144                 75.52
524288                 72.74
1048576                73.39
```

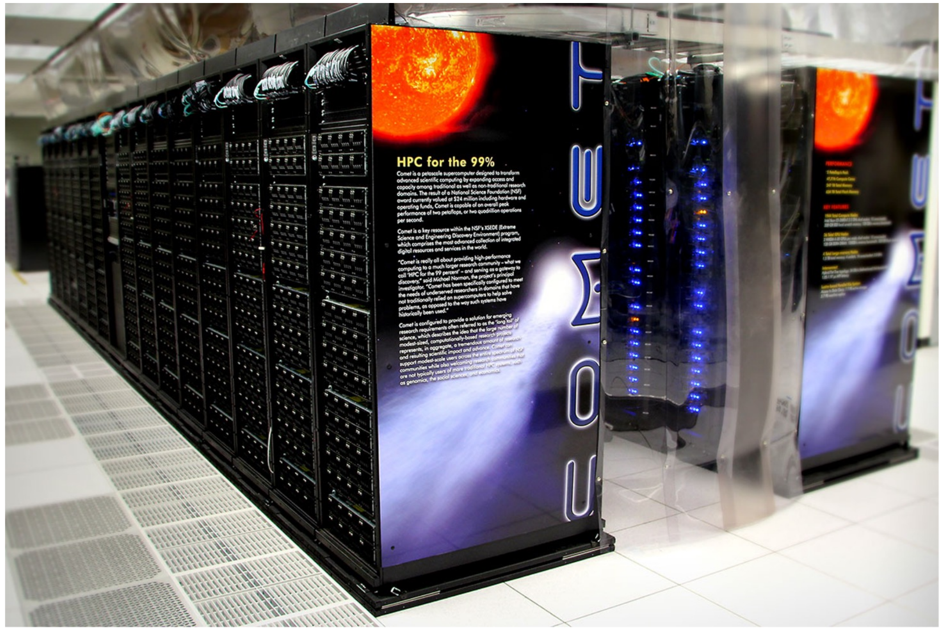
Credit: Chen-Chun , OSU ; Full details in Chen-Chun's MUG 2023 talk

Outline

- Overview
- MVAPICH2 and MVAPICH2-GDR on Expanse
  - System architecture
  - Applications
  - Benchmark results
- MVAPICH2 on Voyager
  - System architecture
  - Kubernetes based usage of MVAPICH2
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
- Summary

Comet Supercomputer

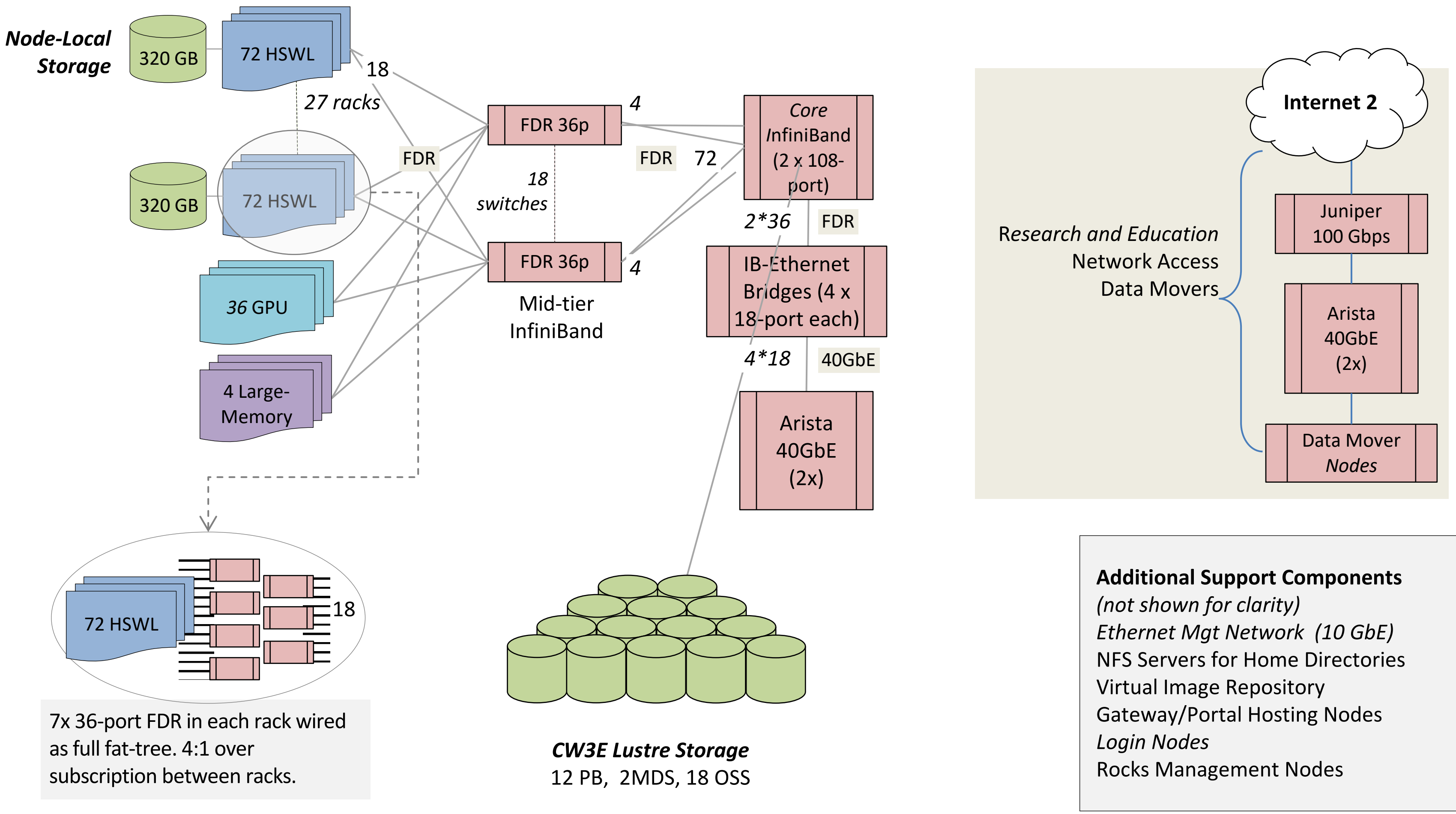
- *Center For Western Weather and Water Extremes (CW3E) has made e*xclusive *u*se of Comet starting *July* 2021 after it was retired from NSF XSEDE/ACCESS service
- Comet is being used for West-WRF ensemble runs during wet-seasons for near real-time forecasts. In addition, it used by CW3E researchers for several research projects.
- *Managed by the* San Diego Supercomputer Center (SDSC)
- Represents over 1 billion core hours (SUs) of computing over the period of 2.5 years from *July* 2021 through December 2023



Highlights: ~440 M CPU hours / year, ~1.2 M GPU hours / year, >12 PB of storage

Comet Network Architecture

InfiniBand compute, Ethernet Storage



Additional Support Components
(not shown for clarity)
Ethernet Mgt Network (10 GbE)
NFS Servers for Home Directories
Virtual Image Repository
Gateway/Portal Hosting Nodes
Login Nodes
Rocks Management Nodes

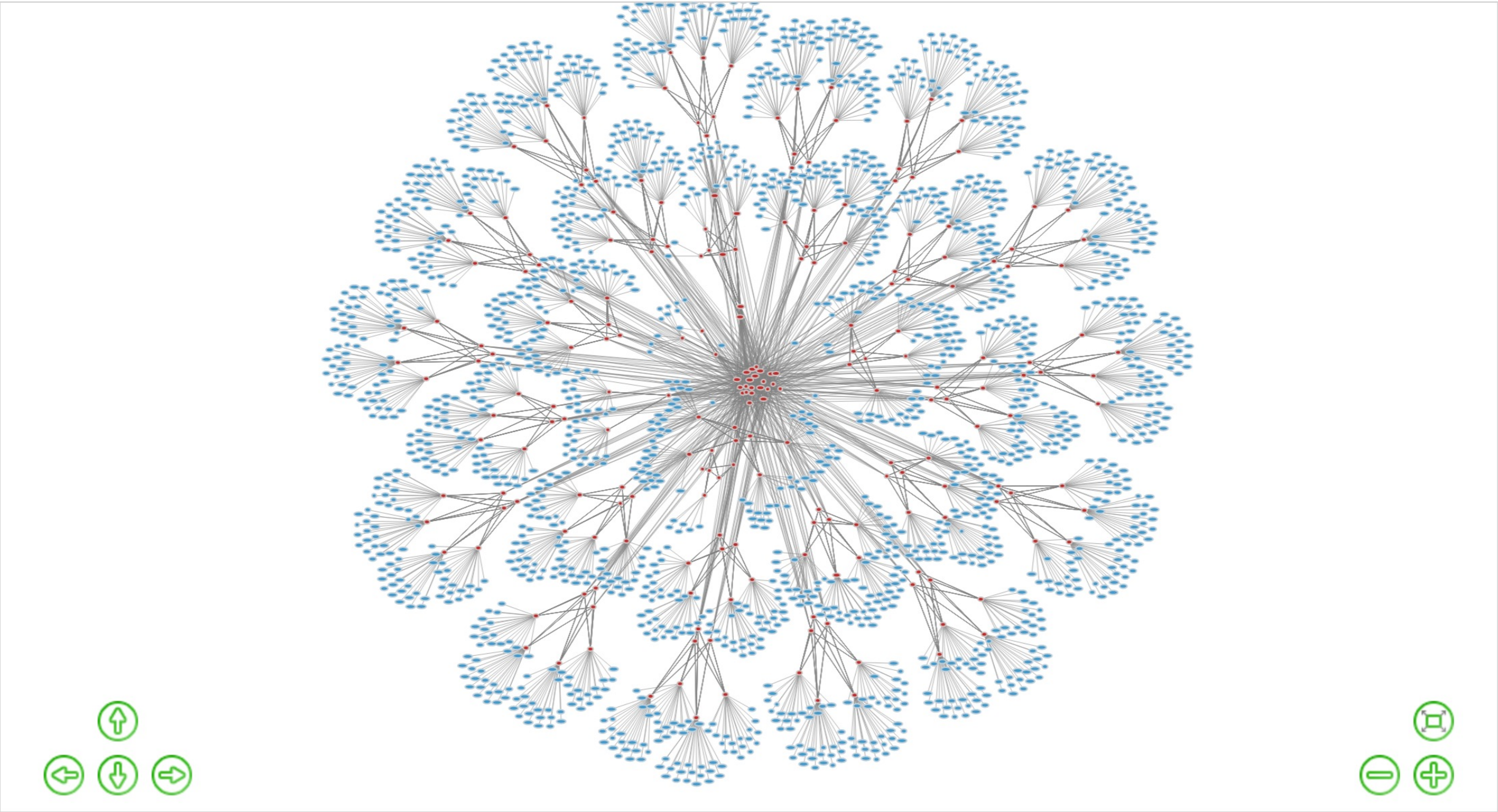
Motivation for INAM on Comet

- CW3E workloads can involve both IO intensive and communication intensive workloads. Sometimes this leads to performance issues. The goal is to use both historical and live data to troubleshoot such issues.
- Monitoring of network fabric health, utilization with notification thresholds set for errors and utilization.
- Enable Lustre traffic monitoring to get a handle on the IO aspect of workloads
- Use MPI information from jobs in combination with Lustre info to evaluate impact on network

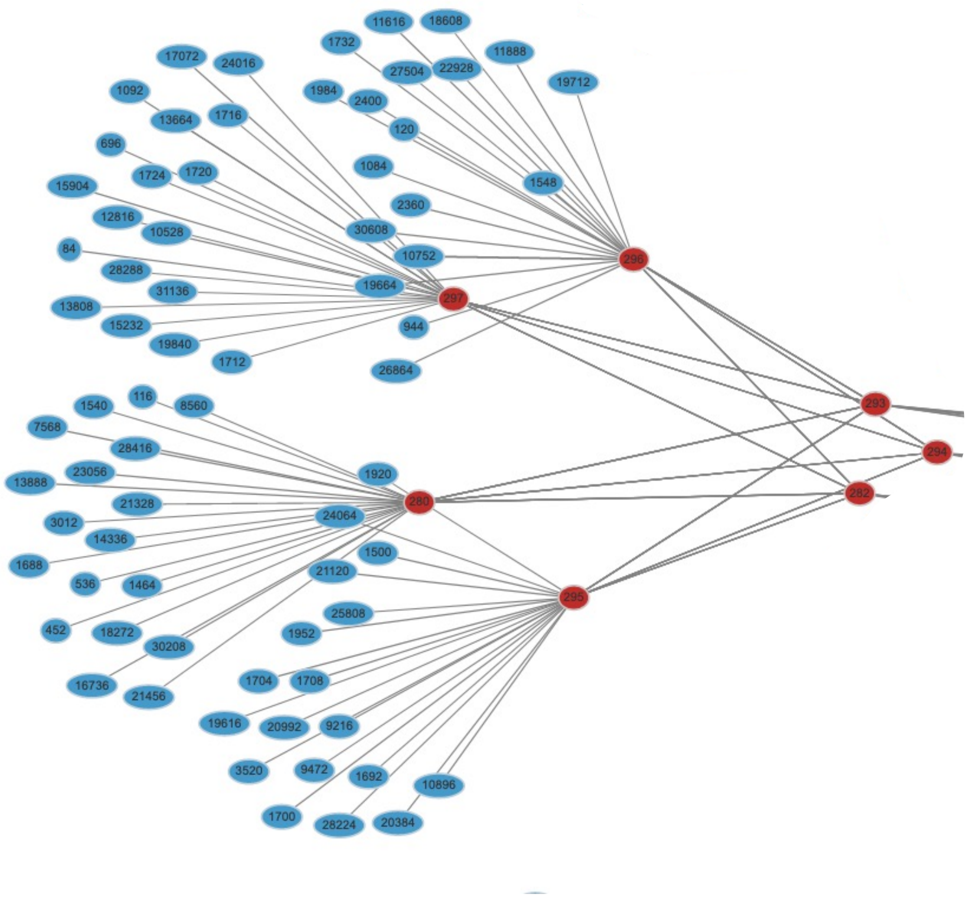
INAM Experience on Comet

- OSU INAM group helped build a custom INAM RPM for Comet OS and OFED stack.
- There was an issue parsing the switch map file. Worked ok after we removed all of the blank lines in our map file. Suggestion to fix as the map file can be parsed as is by other tools.
- At present the install is setup for internal use. Will need user/auth setup additions to enable end user access.
- PhantomJS took some time to build up cached info initially, but web service has been good since.
- We will be installing INAM for Expanse as well. Plan to evaluate use of ClickHouse DB backend.

Full System Network View from INAM

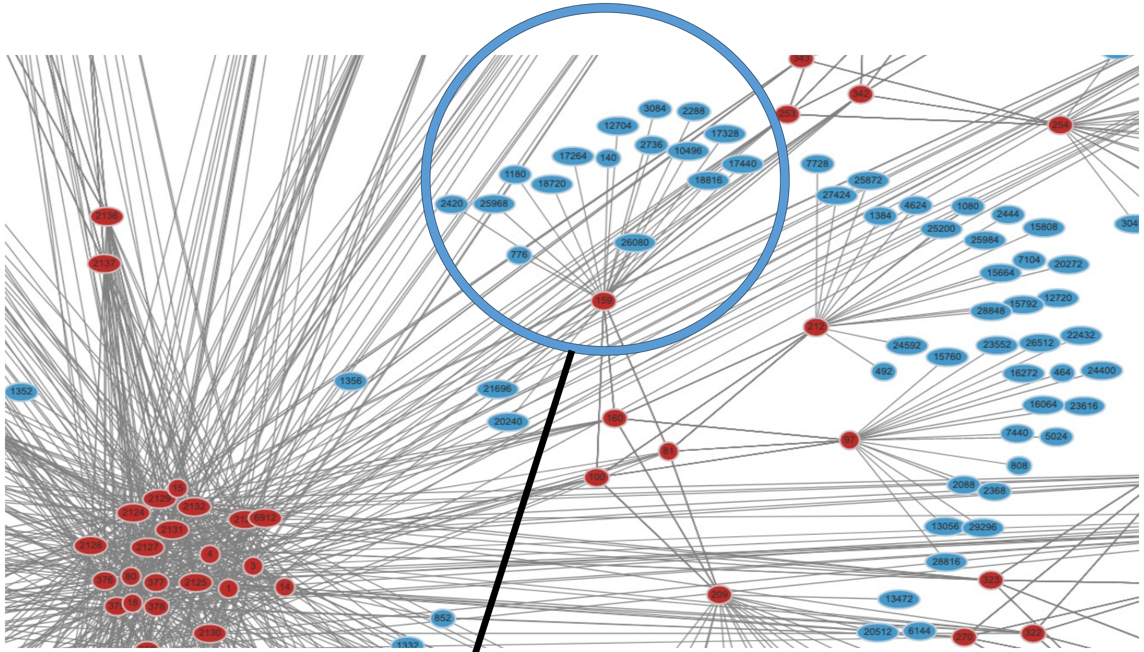


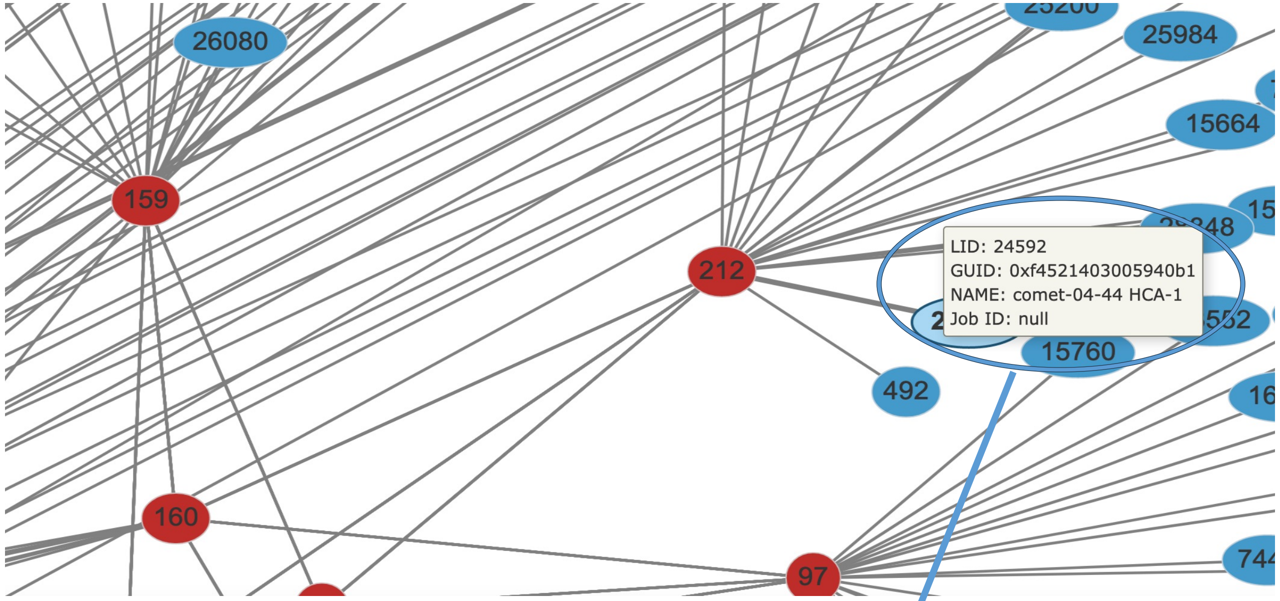
Rack Level Network View from INAM



- *18 nodes per switch*
- *4 switches at first level that feed into 3 switches in the next tier*

Network Views from INAM





Single Switch with 18 nodes

Node LID, GUID, Name info

Notifications from INAM

OSU INAM | Home | Network View | Historical Graphs | Live Jobs | Debug | **Notifications** | User Guide | DB Size Calculator

Notifications & Criteria

Notifications

Search

Notification ID	Criteria ID	Category	Metric	Condition	Threshold Value	Time	Additional Information	Action
2	1	Port Counters	Symbol errors	Greater than	50	22 Aug 2023 15:33:50	MF0;switch-928eee:SX6036/U1 (0x248a070300f25580):11 - comet-34-08 HCA-1 (0x248a0703005c39e1), MF0;sw...	
1	3	Port Counters	Link utilization (percentage)	Greater than	95	22 Aug 2023 09:25:57	comet-dr1-L01 (0xf4521403006ae330):7 - cw3e-oss-1-5 HCA-1 (0x0c42a1030079f00c)	

Showing 1 to 2 of 2 rows

Notification Criteria

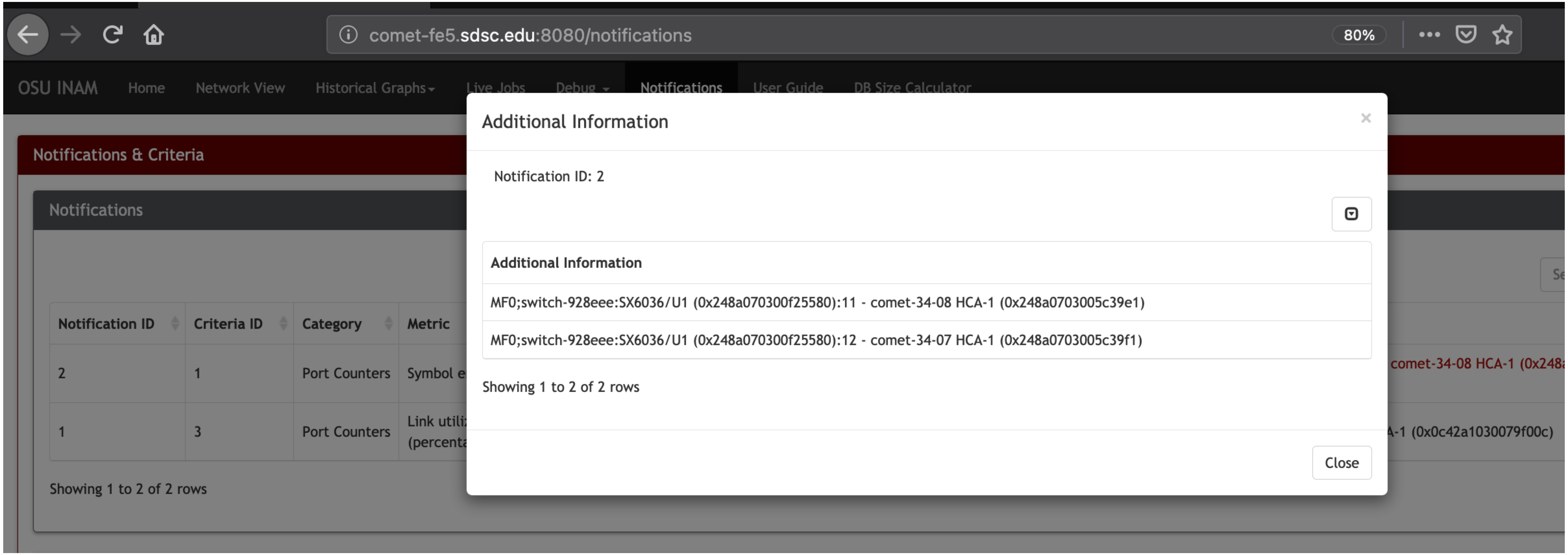
Add Notification Criteria +

Search

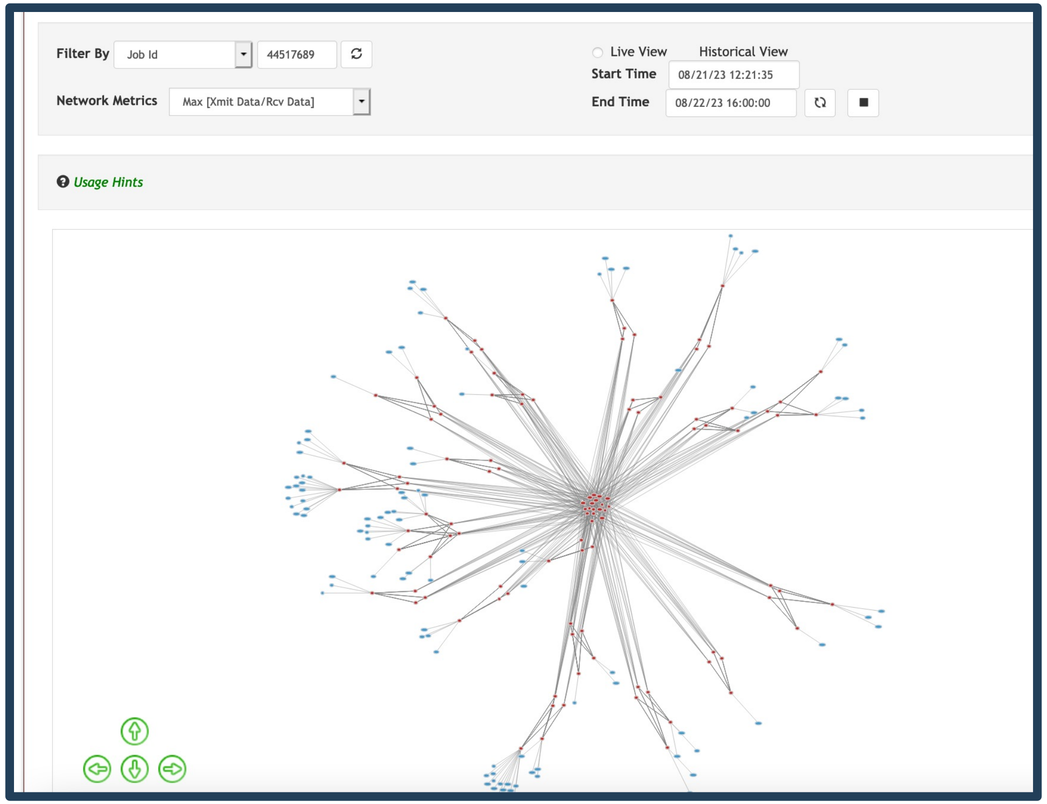
Criteria ID	Category	Metric	Comparison	Threshold Value	Is Recurring	Action
1	Port Counters	Symbol errors	Greater than	50	✔	
2	Port Counters	Link downed	Greater than	1	✔	
3	Port Counters	Link utilization (percentage)	Greater than	95	✖	

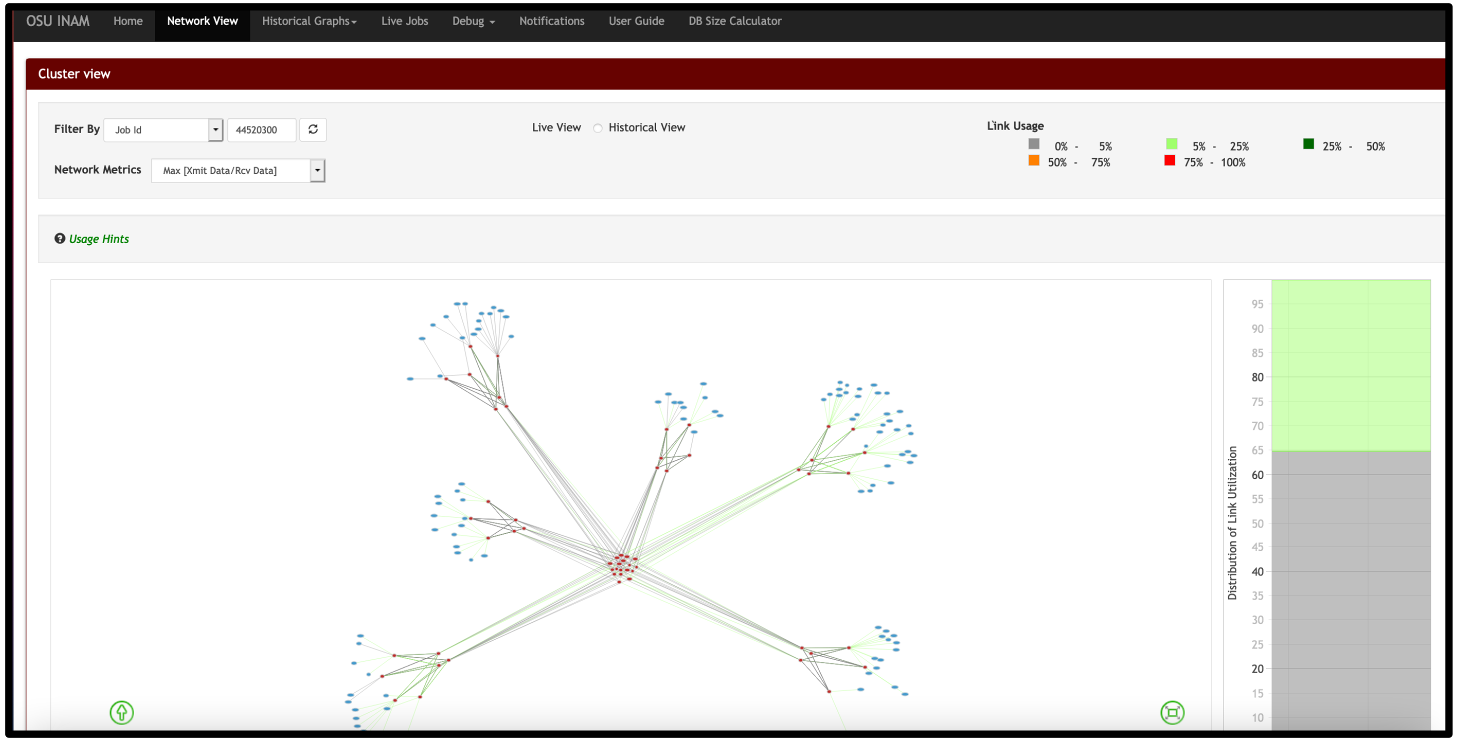
Showing 1 to 3 of 3 rows

Notifications from INAM



Job level view from INAM





Summary

- MVAPICH2 and MVAPICH2-GDR used as the performant option for several application on Expanse.
- Voyager - Heterogenous system designed for AI workloads. Currently in 3-year testbed phase. MVAPICH2 development and testing ongoing.
- INAM installed and in use on Comet. Will be used for network health/performance monitoring and to identify sources of congestion. An install is planned on Expanse as well.
- Upcoming work includes use of MVAPICH2 on the Prototype National Research Platform (PNRP) which is a nationally distributed, open system that features CPUs, FP32- and FP64- optimized GPUs, and FPGAs.

Thank you to our collaborators, partners, users, and the SDSC team!

XSEDE

Extreme Science and Engineering
Discovery Environment

AEON COMPUTING

Mellanox TECHNOLOGIES

Ilkay Altintas
Haisong Cai
Amit Chourasia
Trevor Cooper
Jerry Greenberg
Eva Hocks
Tom Hutton
Christopher Irving
Marty Kandes
Amit Majumdar
Dima Mishin
Sonia Nayak
Mike Norman
Wayne Pfeiffer
Scott Sakai
Fernando Silva
Bob Sinkovits
Subha Sivagnanam
Michele Strong
Shawn Strande
Mahidhar Tatineni
Mary Thomas
Nicole Wolter
Frank Wuerthwein

EXPANSE

COMPUTING WITHOUT BOUNDARIES

Voyager would not be possible without a dedicated team of professionals and experts

Rommie Amaro
Haisong Cai*
Trevor Cooper
Chris Cox*
Javier Duarte
Tom Hutton*
Christopher Irving*
Marty Kandes
Amit Majumdar
Tim McNew*
Dmitry Mishin
Mai Nguyen

Susan Rathbun
Paul Rodriguez
Scott Sakai
Manu Shantharam
Robert Sinkovits
Fernando Silva*
Shawn Strande
Tom Tate*
Mahidhar Tatineni
Mary Thomas
Cindy Wong
Nicole Wolter

Supermicro Team
Habana Team
Arista

*A special thanks to the HPC Systems Group and the Data Center staff who performed onsite work under COVID restrictions to get the system ready for production.