Experiences with INAM on SDSC Systems

Mahidhar Tatineni
Director, User Services, SDSC
SC23 OSU Booth (#1581) Talk
Nov 14, 2023

SDSC SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Outline

- Overview
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
  - Use case illustrating value of INAM
- Future work
  - Planned installation on Expanse
  - Expanse architecture
- Summary

Overview

- OSU InfiniBand Network Analysis and Monitoring (INAM) tool provides an open-source option to monitor IB networking

- 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

Outline

- Overview
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
  - Use case illustrating value of INAM
- Future work
  - Planned installation on Expanse
  - Expanse architecture
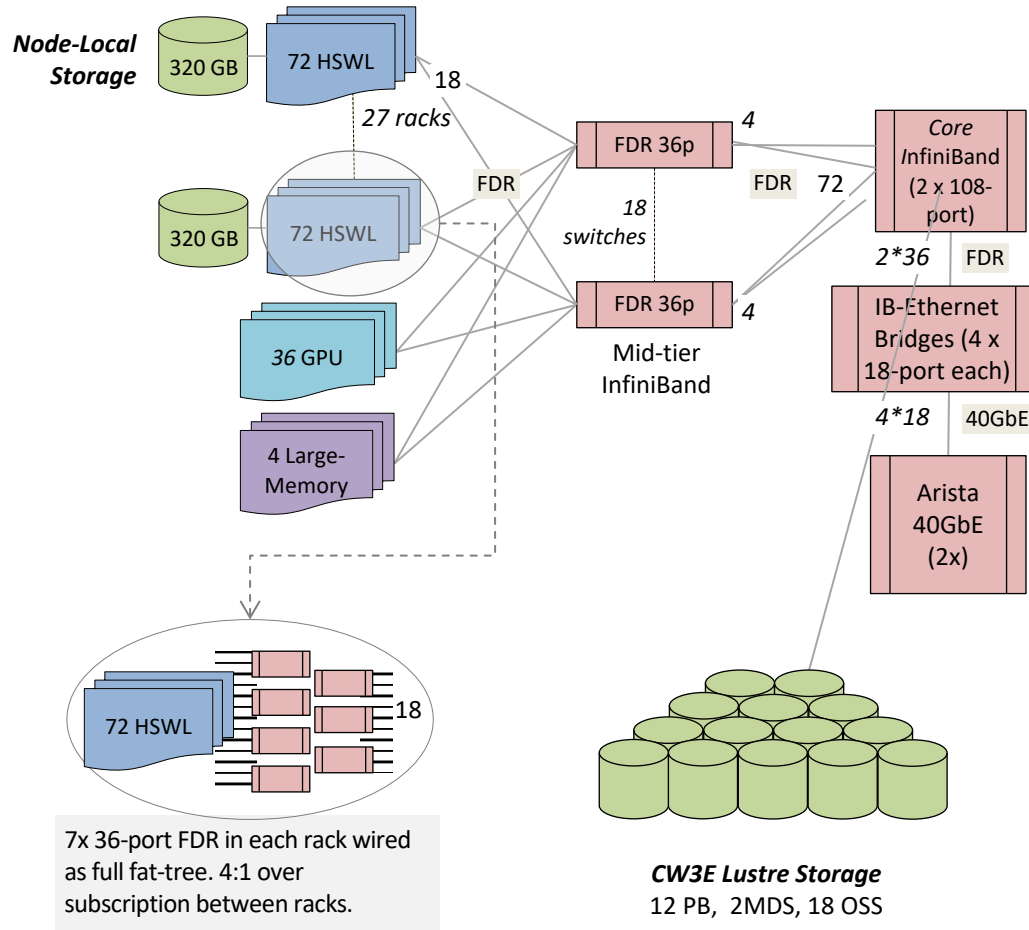- Summary

Comet Supercomputer

- *Center For Western Weather and Water Extremes (CW3E)* has made exclusive use 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

Node-Local Storage

320 GB

72 HSWL

18

27 racks

320 GB

72 HSWL

FDR

36 GPU

4 Large-Memory

FDR 36p — 4

18 switches

FDR 36p — 4

Mid-tier InfiniBand

FDR — 72

Core InfiniBand (2 x 108-port)

*2*36* — FDR

IB-Ethernet Bridges (4 x 18-port each)

*4*18* — 40GbE

Arista 40GbE (2x)

Internet 2

Research and Education Network Access Data Movers

Juniper 100 Gbps

Arista 40GbE (2x)

Data Mover *Nodes*

72 HSWL — 18

7x 36-port FDR in each rack wired as full fat-tree. 4:1 over subscription between racks.

CW3E Lustre Storage
12 PB, 2MDS, 18 OSS
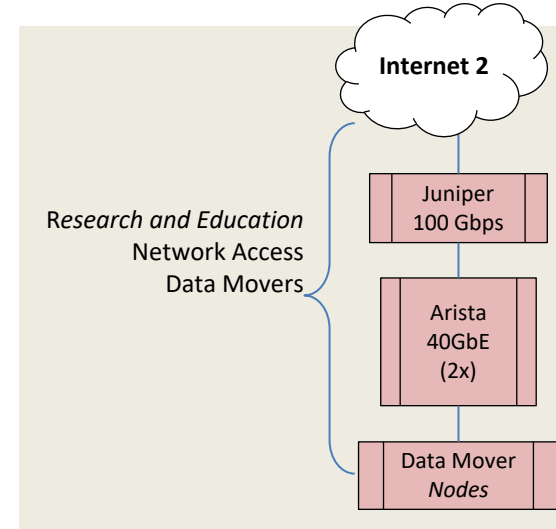
Additional Support Components
(not shown for clarity)
Ethernet Mgt Network (10 GbE)
NFS Servers for Home Directories
Virtual Image Repository
Gateway/Portal Hosting Nodes
*Login Nodes*
Rocks Management Nodes

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.
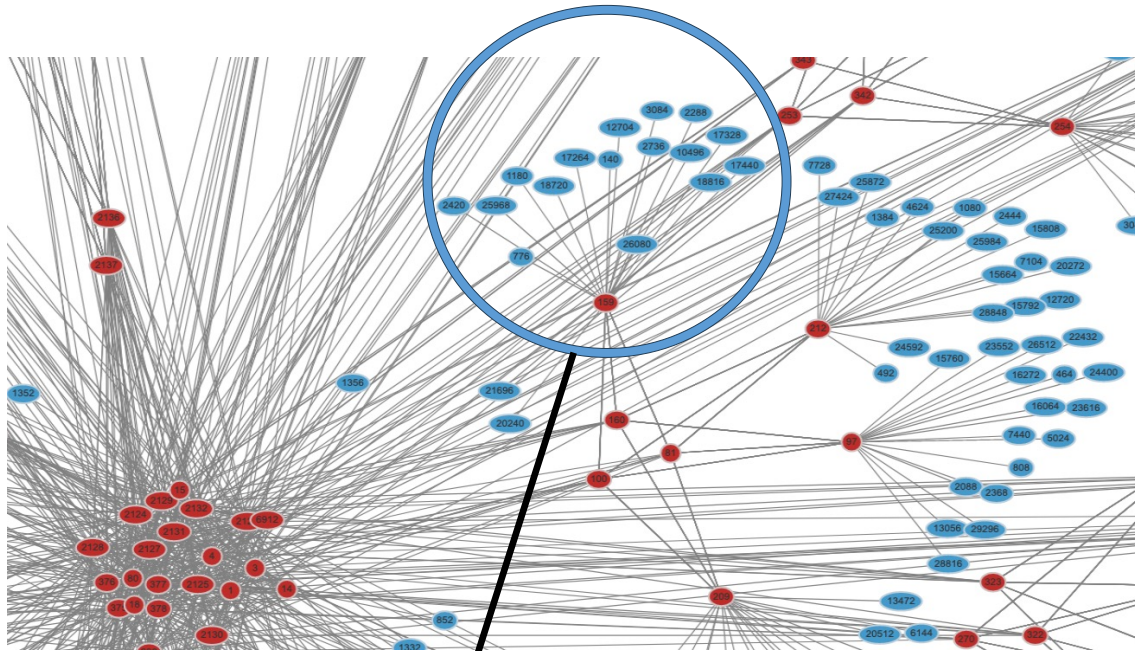
Full System Network View from INAM
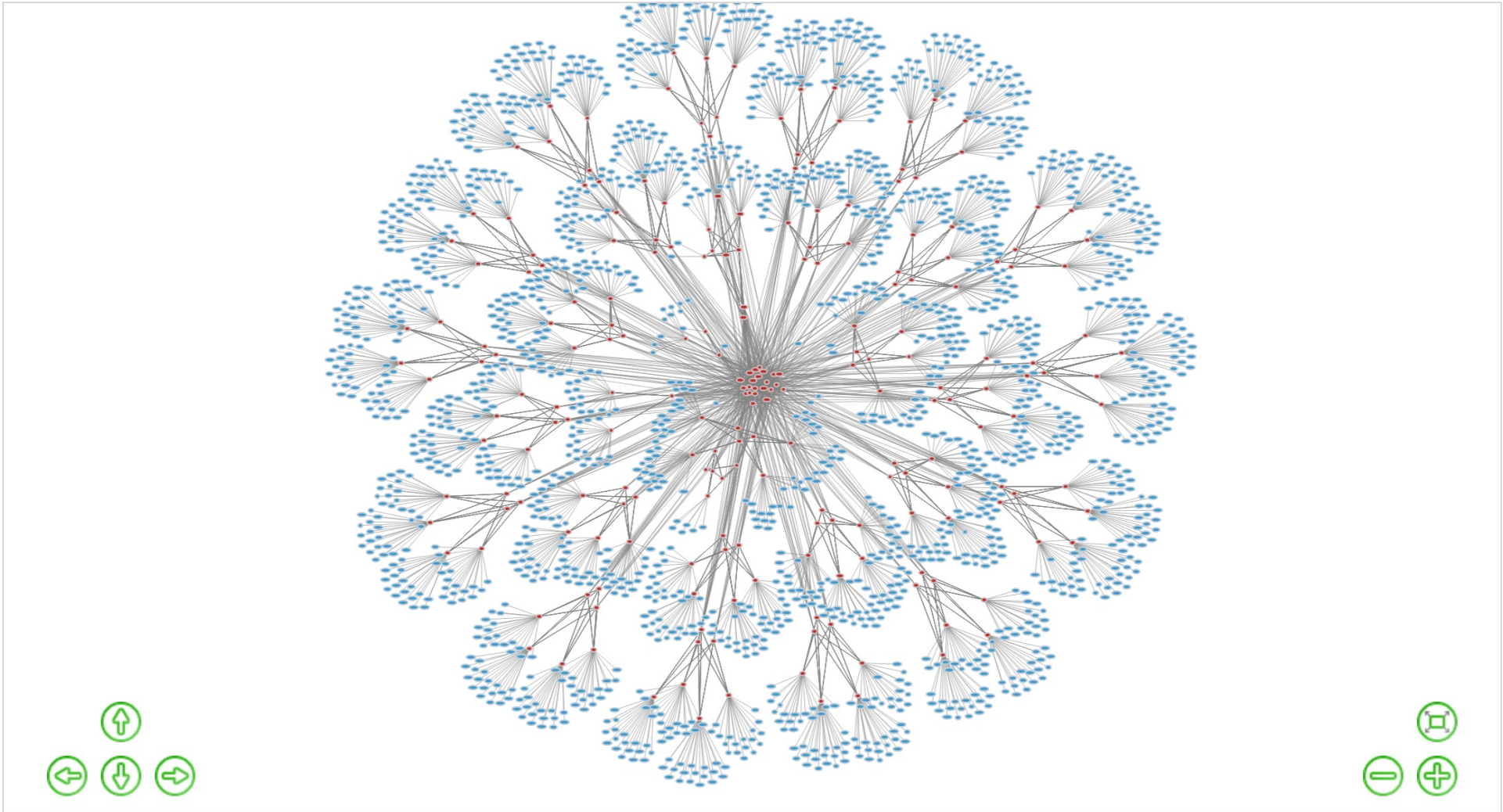
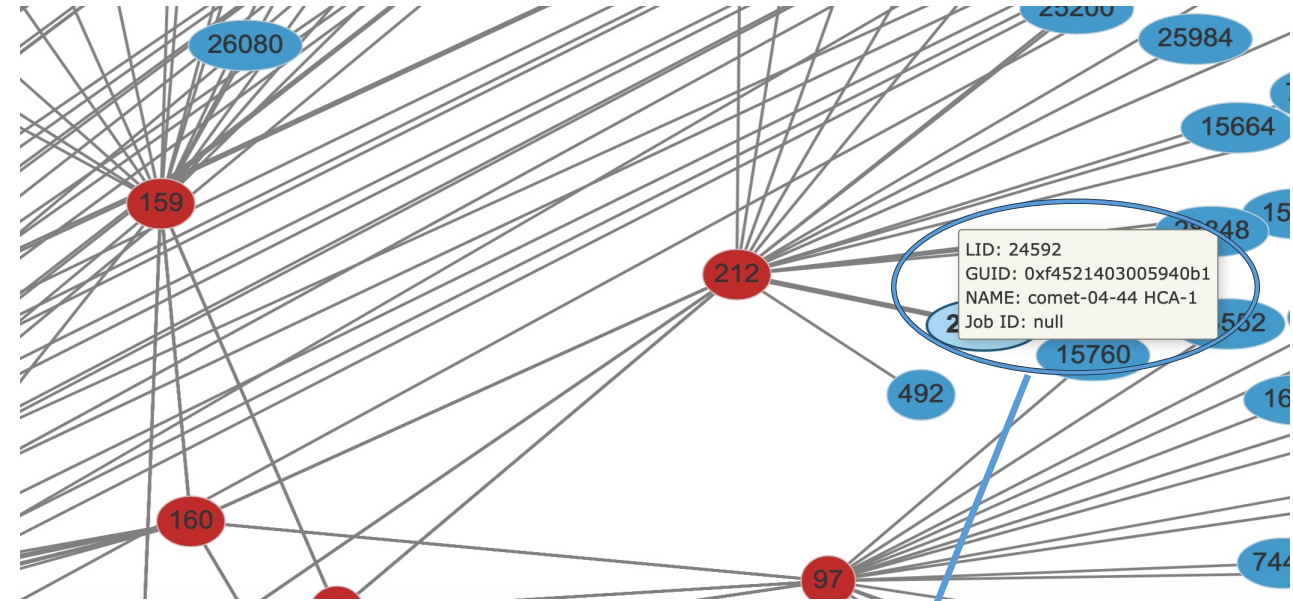# Rack Level Network View from INAM
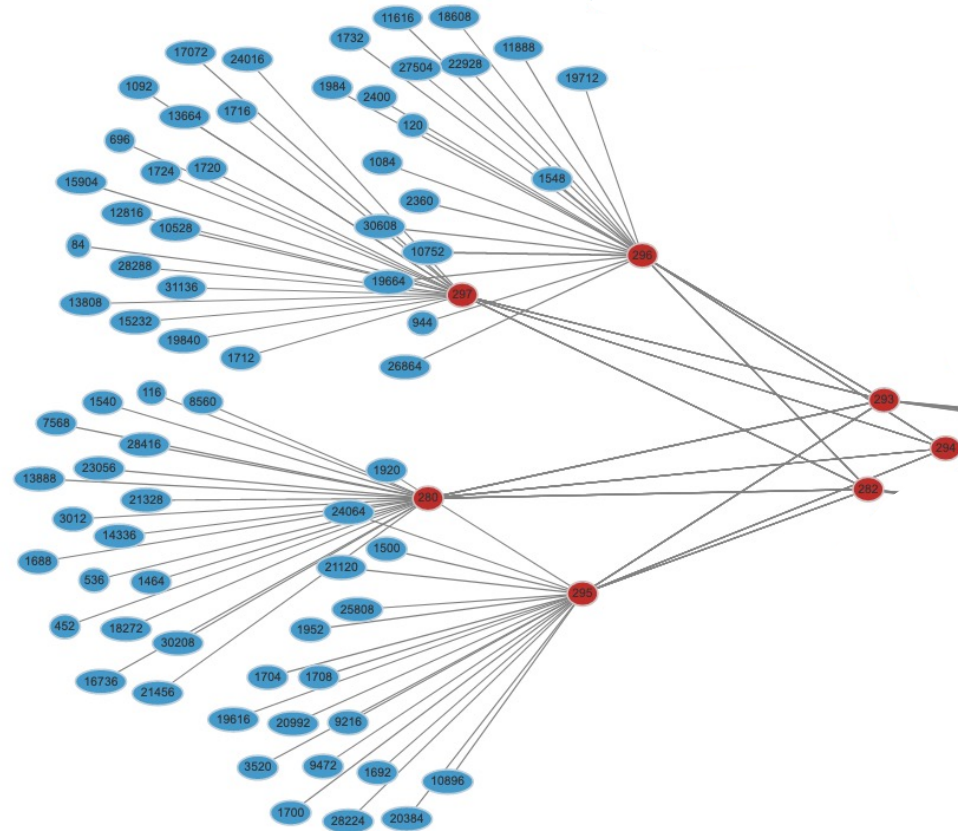
- *18 nodes per switch*
- *4 switches at first level that feed into 3 switches in the next tier*
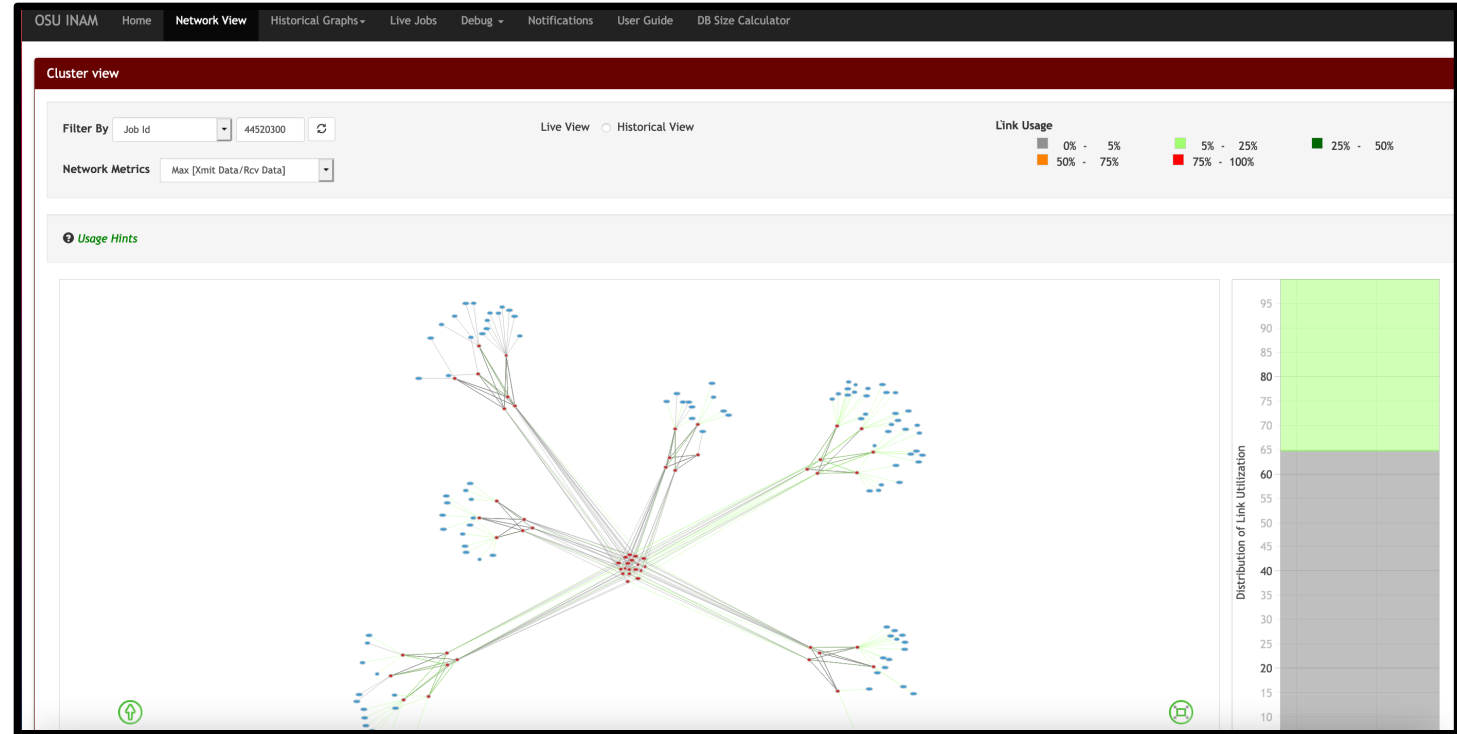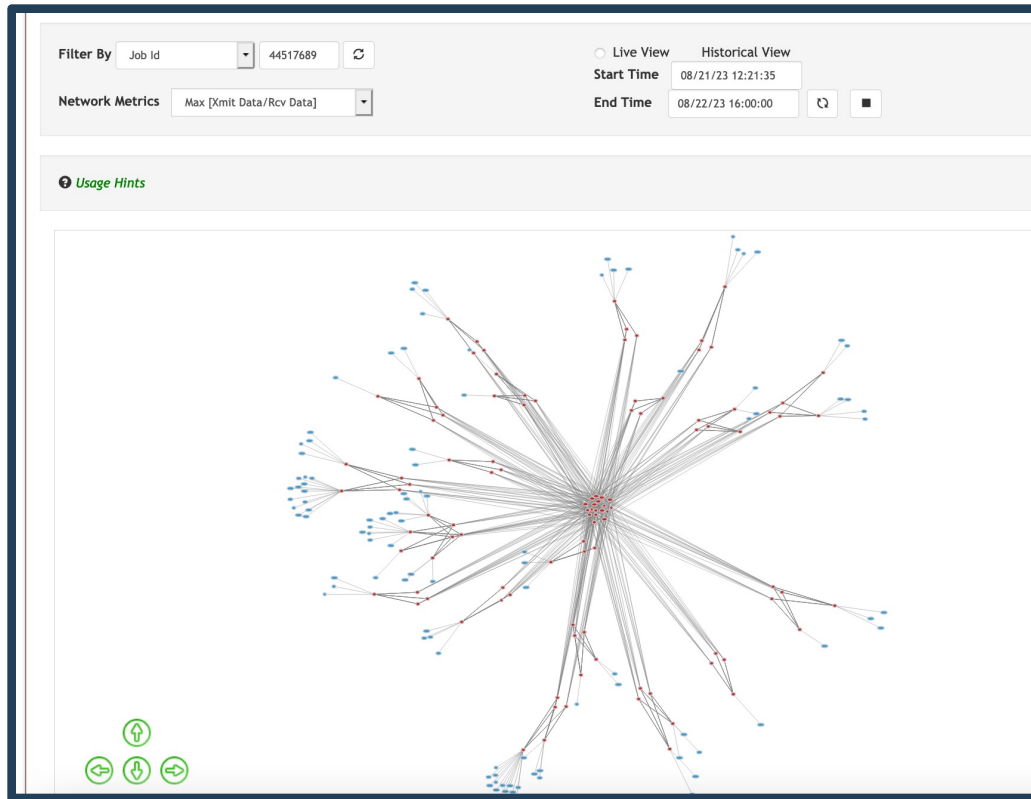
Network Views from INAM

Single Switch with 18 nodes

Node LID, GUID, Name info

LID: 24592
GUID: 0xf4521403005940b1
NAME: comet-04-44 HCA-1
Job ID: null

Job level view from INAM

Notifications from INAM

Notifications from INAM

Notifications can identify issues

In Port Counters , linkusage exceeded threshold value - 95.0.

Link information

IB-R24-L3-01 (0xf452140300f635a0):36 - comet-dr1-L04 (0xf4521403006ae2d0):2

comet-dr1-L04 (0xf4521403006ae2d0):23 - comet-dr1-S01 (0xe41d2d0300002500):23

MF0;comet-dr1:SX6506/L06/U1 (0xf4521403008b7120):16 - cw3e-oss-3-5 HCA-1 (0x0c42a10300825a84)

MF0;comet-dr1:SX6506/L06/U1 (0xf4521403008b7120):17 - cw3e-oss-3-3 HCA-1 (0x0c42a1030079c610)

comet-dr1-L01 (0xf4521403006ae330):17 - cw3e-oss-2-1 HCA-1 (0x0c42a10300825ad4)

comet-dr1-L02 (0xf4521403006ae090):13 - cw3e-oss-2-5 HCA-1 (0x0c42a10300825a64)

comet-dr1-L01 (0xf4521403006ae330):11 - cw3e-oss-1-3 HCA-1 (0x0c42a10300825ac8)

comet-dr1-L02 (0xf4521403006ae090):7 - cw3e-oss-2-2 HCA-1 (0x0c42a10300825a60)

comet-dr1-L01 (0xf4521403006ae330):1 - cw3e-oss-1-1 HCA-1 (0x0c42a10300825a94)

IB-R09-L3-02 (0xf452140300f74d50):35 - comet-dr1-L04 (0xf4521403006ae2d0):3

comet-dr1-S02 (0xe41d2d0300002520):2 - comet-dr1-L01 (0xf4521403006ae330):26

comet-dr1-L04 (0xf4521403006ae2d0):28 - comet-dr1-S02 (0xe41d2d0300002520):22

…

…

comet-dr1-L02 (0xf4521403006ae090):9 - cw3e-oss-2-3 HCA-1 (0x0c42a103006b1810)

comet-dr1-L01 (0xf4521403006ae330):13 - cw3e-oss-1-4 HCA-1 (0x0c42a10300825a74)

comet-dr1-L01 (0xf4521403006ae330):9 - cw3e-oss-1-2 HCA-1 (0x0c42a10300825a4c)

comet-dr1-L03 (0xf4521403006ae390):23 - comet-dr1-S01 (0xe41d2d0300002500):17

MF0;comet-dr1:SX6506/L06/U1 (0xf4521403008b7120):13 - cw3e-oss-3-6 HCA-1 (0x0c42a1030079f000)

MF0;comet-dr1:SX6506/L06/U1 (0xf4521403008b7120):14 - cw3e-oss-3-4 HCA-1 (0x0c42a103006b1804)

-OSU INAM

Connections from mid-tier to director switch

Connections from leaf to spine in director switch

Connections from director switch to Lustre OSSs

Notifications from INAM

- These notifications were received when the intensive near real time (NRT) forecast simulations were being run.

- The I/O patterns and load were saturating leaf to spine links in the director switch and also some of the director switch to OSS links were saturating.

- This was impacting I/O performance for the NRT runs and in some cases that lead to missed targets (time bound).

- Based on information from the INAM notifications and network monitoring data, decision was made to move the OSS links into the mid-tier switches.

- This removes the leaf to spine bottlenecks and also balanced out the traffic on the OSS links.

Comet Network Architecture (original)
InfiniBand compute, Ethernet Storage
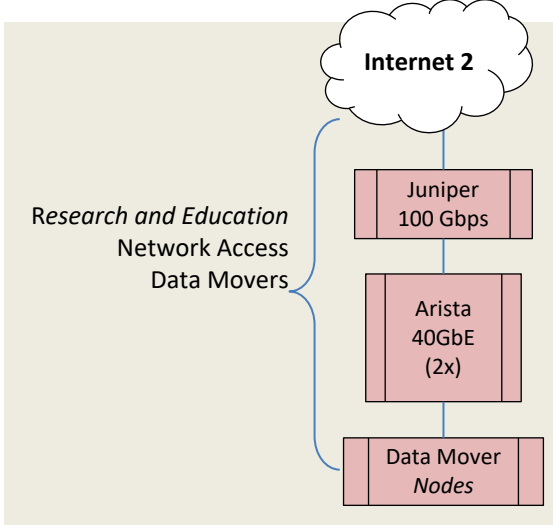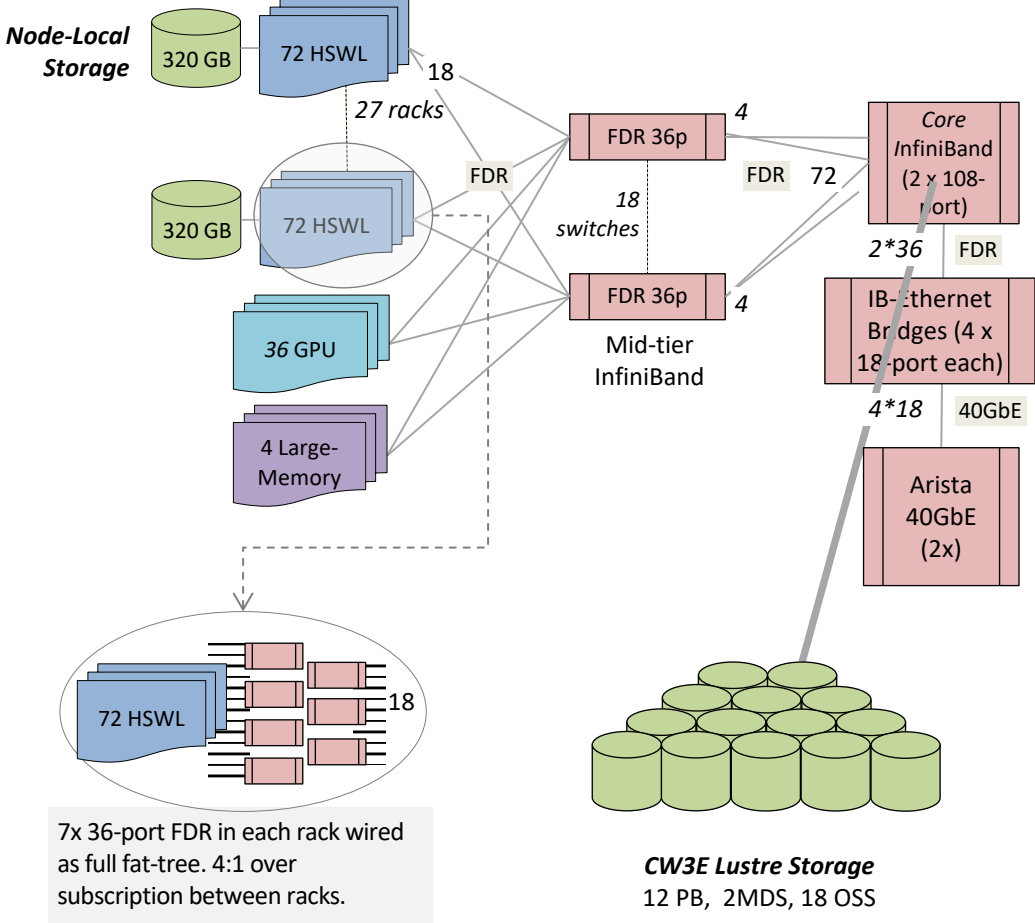
Node-Local Storage

320 GB

72 HSWL

18

27 racks

320 GB

72 HSWL

FDR

FDR 36p

4

18 switches

FDR

72

Core InfiniBand (2 x 108-port)

2*36

FDR

36 GPU

FDR 36p

4

Mid-tier InfiniBand

IB-Ethernet Bridges (4 x 18-port each)

4 Large-Memory

4*18

40GbE

Arista 40GbE (2x)

72 HSWL

18

7x 36-port FDR in each rack wired as full fat-tree. 4:1 over subscription between racks.

CW3E Lustre Storage
12 PB, 2MDS, 18 OSS

Internet 2

Research and Education Network Access Data Movers

Juniper 100 Gbps

Arista 40GbE (2x)

Data Mover Nodes

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

Comet Network Architecture (New)
InfiniBand compute, Ethernet Storage
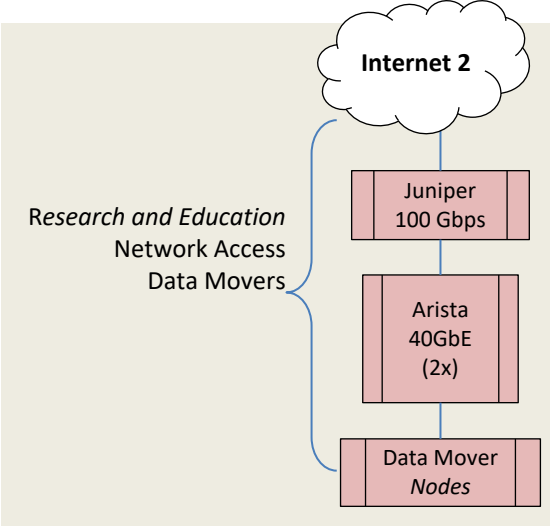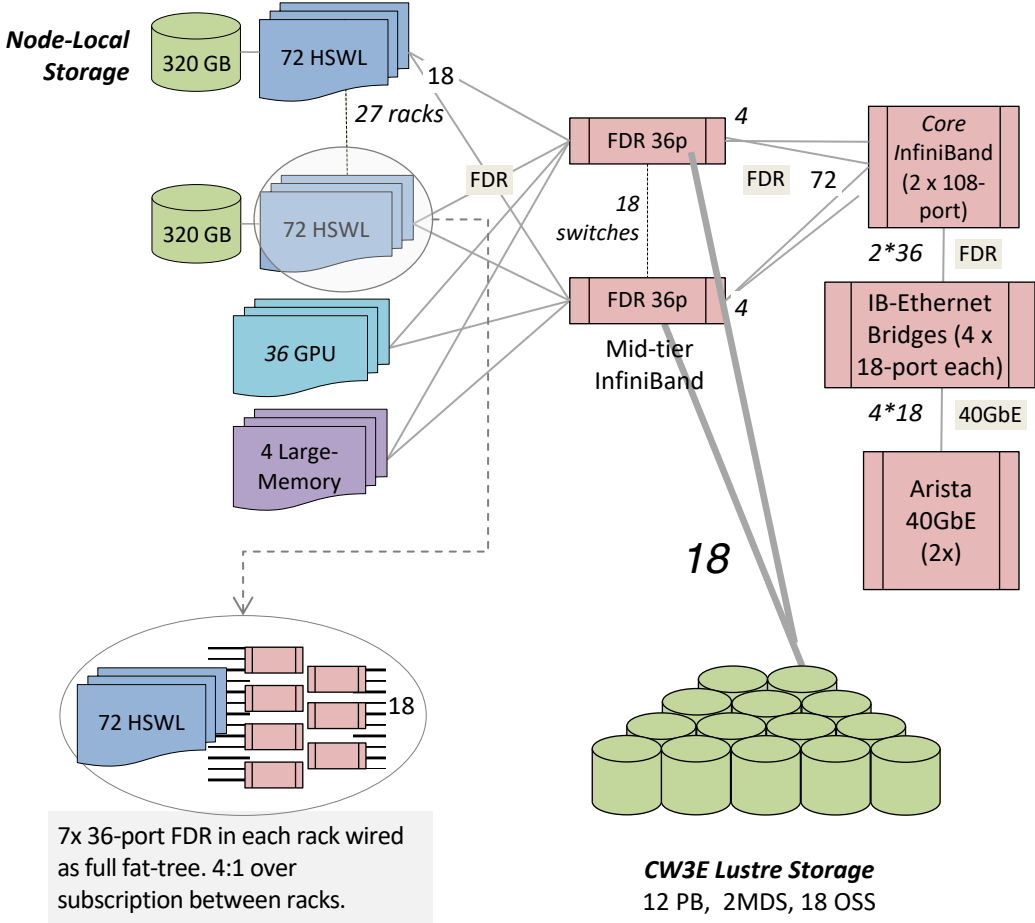
Node-Local Storage

320 GB

72 HSWL — *27 racks* — 18

320 GB

72 HSWL — FDR

36 GPU

4 Large-Memory

FDR 36p — 4

18 switches

FDR 36p — 4

Mid-tier InfiniBand

Core InfiniBand (2 x 108-port)

FDR — 72

2*36 — FDR

IB-Ethernet Bridges (4 x 18-port each)

4*18 — 40GbE

Arista 40GbE (2x)

72 HSWL — 18

7x 36-port FDR in each rack wired as full fat-tree. 4:1 over subscription between racks.

18

CW3E Lustre Storage
12 PB, 2MDS, 18 OSS

Internet 2

Research and Education Network Access Data Movers

Juniper 100 Gbps

Arista 40GbE (2x)

Data Mover *Nodes*

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

Notifications after recabling

Link information

IB-R16-L1-02 (0xf4521403008fcbf0):2 - IB-R16-L2-05 (0xf452140300918b40):8
IB-R16-L1-02 (0xf4521403008fcbf0):36 - comet-16-37 HCA-1 (0xf4521403005bff01)
IB-R20-L1-02 (0xf452140300969400):2 - IB-R20-L2-05 (0xf452140300969100):8
IB-R28-L1-02 (0xf4521403006db860):7 - IB-R28-L2-06 (0xf4521403008dce90):7
IB-R28-L1-04 (0xf4521403008dcf10):15 - IB-R28-L2-07 (0xf4521403008dca90):21
IB-R11-L1-02 (0xf45214030075bfe0):26 - comet-11-47 HCA-1 (0xf4521403005c0c01)
IB-R28-L1-02 (0xf4521403006db860):4 - IB-R28-L2-05 (0xf4521403008d7980):10
B-R16-L1-02 (0xf4521403008fcbf0):13 - IB-R16-L2-07 (0xf4521403009010f0):7
IB-R20-L1-02 (0xf452140300969400):13 - IB-R20-L2-07 (0xf452140300969800):7
IB-R28-L1-04 (0xf4521403008dcf10):26 - comet-28-11 HCA-1 (0xf4521403005936f1)
IB-R08-L2-06 (0xf452140300968400):34 - IB-R24-L3-07 (0xf452140300f67570):10
IB-R19-L1-02 (0xf452140300953000):9 - IB-R19-L2-06 (0xf452140300952e00):9
…

…
IB-R19-L1-02 (0xf452140300953000):1 - IB-R19-L2-05 (0xf4521403009528a0):7
IB-R07-L1-01 (0xf452140300f631a0):6 - IB-R07-L2-05 (0xf452140300f65ef0):6
IB-R23-L1-03 (0xf4521403006f1f00):31 - comet-23-24 HCA-1 (0xf452140300596971)
IB-R19-L1-02 (0xf452140300953000):22 - comet-19-51 HCA-1 (0xf452140300592531)
IB-R07-L1-01 (0xf452140300f631a0):12 - IB-R07-L2-06 (0xf452140300f66570):6
IB-R12-L1-01 (0xf45214030075c0e0):18 - IB-R12-L2-07 (0xf45214030075c960):6
IB-R24-L3-05 (0xf452140300f65df0):36 - comet-dr1-L04 (0xf4521403006ae2d0):10

-OSU INAM

Connections within rack level switches

Connections from nodes to in rack switches

Connections from mid-tier to director switch

The OSS links and director switch leaf/spine links no longer show up in the saturation lists

Outline

- Overview
- INAM on Comet
  - System architecture
  - Network and Job Level Views
  - Notifications
  - Use case illustrating value of INAM
- Future work
  - Planned installation on Expanse
  - Expanse architecture
- Summary

Future Work: INAM on Expanse

- Expanse features a similar architecture to Comet with 3:1 oversubscription of links between racks.

- Single switch at rack level (unlike Comet which had 7).

- Will be using the ClickHouse database server and the C++ client. Expected to perform better than the setup on Comet.

- Expanse system is managed using Bright Cluster Manager (BCM). Look into install of INAM using BCM or develop a compatible process.

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

Scalable Compute Unit
Non-blocking fabric
56 CPU nodes
4 GPU nodes

60 HDR 100 to nodes
10x HDR 200 to L2

10 GbE switch for mgmt.
And external connectivity

14 Chassis
4 nodes/chassis
56 compute nodes
7,168 CPU cores

4x4 GPU nodes
10,240 Tensor Cores
256 CPU cores

System Layout
1 row 7 SSCU
1 row 6 SSCU + Core Mgmt. rack

Level 2 IB Fabric
- 5 HDR switches (1 shown)
- 2x HDR200 down to each of the 13 non-blocking SSCUs
- 3 uplinks each to Performance Storage switch
- 2 uplink each to Cloud Storage switch

Performance Storage
12PB Lustre
7 HA OSS pairs
4 NVMe HA Metadata Servers

Performance Storage (12 PB)

2 links to each OSS 1 link to each MDT
3 links to each L2 switch

Object Storage
7 PB Ceph
32 storage servers

Object Storage (7 PB)

32 links to each storage server

3 links to each L2 switch

Core Management Rack
Service nodes
Core switches
Home File System

2x 10GbE
5x40-port HDR
2 login nodes
NFS server
2 hosting nodes
Cluster mgmt. nodes
4 large memory nodes
2TB/node
64 CPU cores/node

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
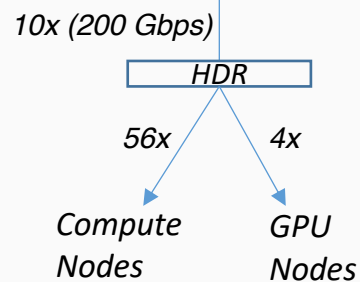- 1 TB NVME

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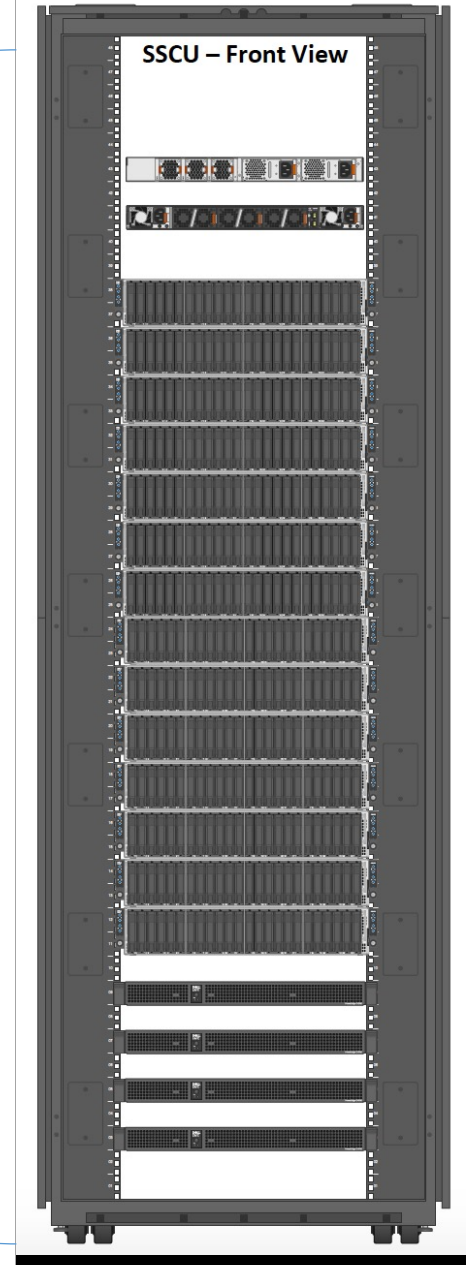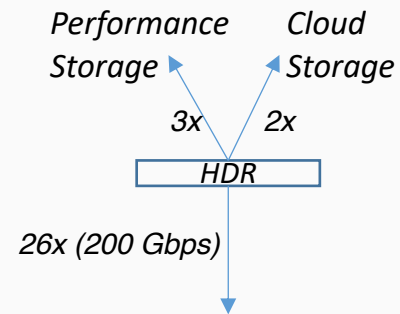- 1.6 TB NVMe
- Intel CPUs

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

1 HDR Switch/SSCU

10x (200 Gbps)

HDR

56x   4x

Compute   GPU
Nodes   Nodes

5 Level 2 switches

Performance   Cloud
Storage   Storage

3x   2x

HDR

26x (200 Gbps)

SSCU – Front View

SDSC SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Summary

- INAM installed and in production use on Comet.
- Used for network health/performance monitoring and to identify sources of congestion.
- Successfully identified hotspots and bottlenecks that were causing Lustre filesystem performance issues under heavy NRT simulation loads.
- Recabled and confirmed performance improved significantly and I/O bottlenecks mitigated.
- Expanse install in planning phase. Will be using the ClickHouse database server and the C++ client.