

Large Scale System Monitoring and Analysis on Blue Waters Using OVIS

SAND2014-3594C

Mike Showerman, Jeremy Enos, Joseph Fullop[§],
Paul Casella[‡],
Nichamon Naksinehaboon, Narate Taerat, Tom Tuckert[†],
Jim Brandt , Ann Gentile, Ben Allan*

[§] NCSA, Urbana-Champaign, IL

[‡] Cray Inc., Seattle, WA

[†] Open Grid Computing, Austin, TX

*Sandia National Laboratories, Albuquerque, NM

[§] This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (award number ACI 1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

* Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

2014-3075C

Outline

- Motivation for Continuous Whole System Monitoring
- Data of Interest
- Monitoring Requirements
- Overview of OVIS Data Collection and Transport
- Enhancements to Meet Requirements
- Application Impact Testing & Results
- A Look at the Data
- Conclusions & Future Work

Motivation

Gain insight into resource utilization/bottlenecks (e.g. network bandwidth/hotspots, file system utilization/contention, memory utilization)

- Debugging
- Anomaly detection
- Historical comparison
- Intelligent job placement

Data of Interest

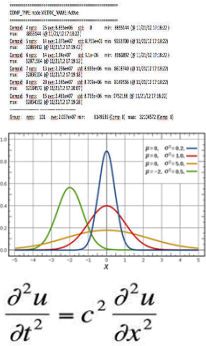
- High Speed Network Performance Counters
 - Traffic
 - Contention
 - Link Status
- Lustre File System Statistics
- LNet traffic
- CPU load
- Memory being used

Blue Waters Monitoring Requirements

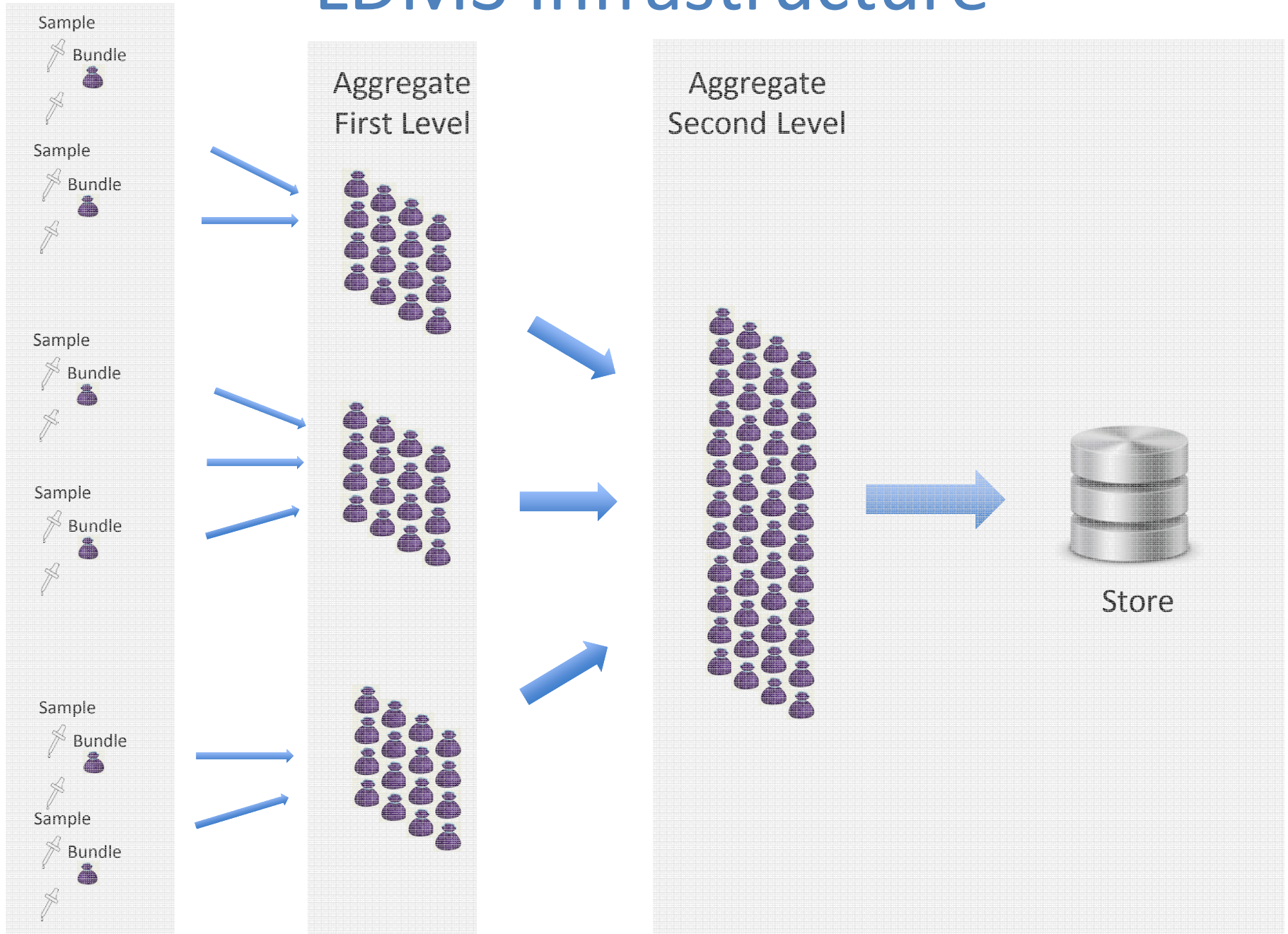
- Need to collect High Speed Network performance metrics to understand network contention and impact on applications
- Would like to collect at one minute intervals
- All data collection synchronized to provide “snapshots” of the system
- Quantify monitoring impact on large scale applications

OVIS Functional Components

- Lightweight Distributed Metric Service (LDMS)
 - Sample, aggregate, transport, and store data
- Analysis
- Modeling
- Visualization
- Notification and Feedback

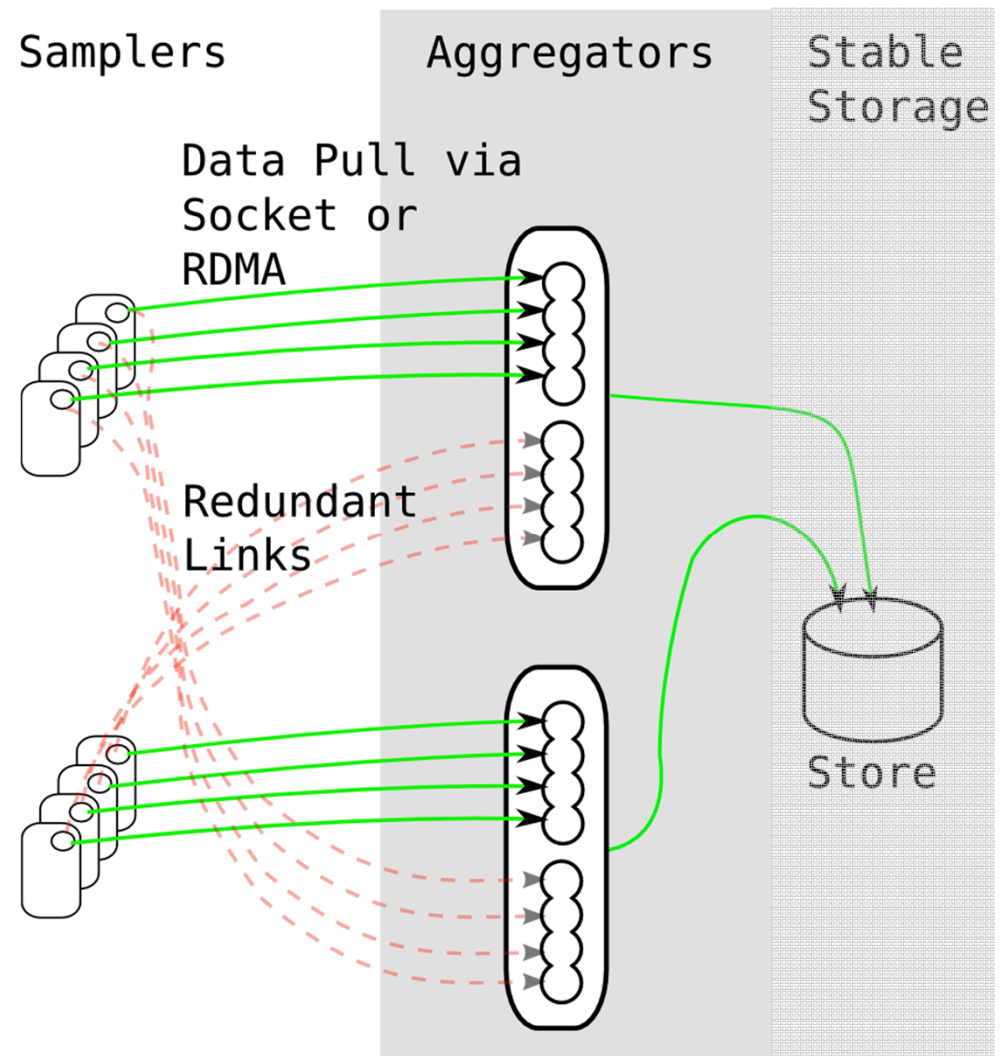


LDMS Infrastructure



Generic LDMS Configuration

- Samplers collect data and bundle into metric sets
- Aggregators pull metric set data from Samplers over Socket or RDMA connections
 - Redundant inactive links can be defined for fast failover
- Aggregators can be daisy chained to provide hierarchy and/or network transition
- Aggregators can load storage plugins and push data to stable storage in a variety of formats



LDMS Functional Overview

- Data is bundled into “Metric Sets” – this is the granularity of storage and query
- Metric Sets have associated Data and Meta-data and include generation numbers for both
 - Meta-data is only transmitted during initial setup and when change occurs
- Run-time plugin add, start, stop
 - Add new collection components
 - Start collection – begin scheduling data collection and make data visible to queries
 - Stop collection – stop scheduling data collection, last data set still visible to queries – no CPU overhead associated with this as no collection scheduled
 - Modify collection frequency – change the length of time between collection on a per data set basis
- RDMA over Gemini transport is utilized for Blue Waters

Metric Set Memory

Metric Meta Data

- Generation Number

Metric Descriptor

- Name
- Component ID
- Type
- Offset

Metric Descriptor

- Name
- Component ID
- Type
- Offset

Metric Descriptor

- Name
- Component ID
- Type
- Offset



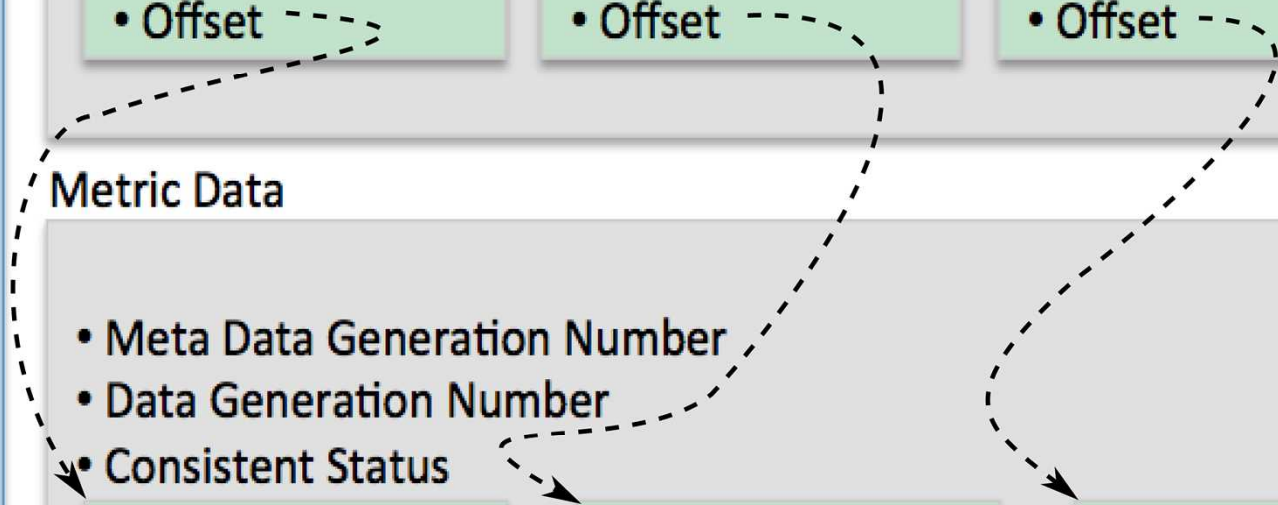
Metric Data

- Meta Data Generation Number
- Data Generation Number
- Consistent Status

Value

Value

Value



LDMS metric set Example (meta data)

```
# ldms_ls -h nid00044 -x ugni -p 412 -v
```

```
nid00044/cray_system_sampler_r: consistent, last update: Wed Apr 09 08:55:20  
2014 [727us]
```

METADATA -----

Size : 13560

Inuse : 7144

Metric Count : 130

GN : 131

DATA -----

Timestamp : Wed Apr 09 08:55:20 2014 [727us]

Consistent : TRUE

Size : 1088

Inuse : 1088

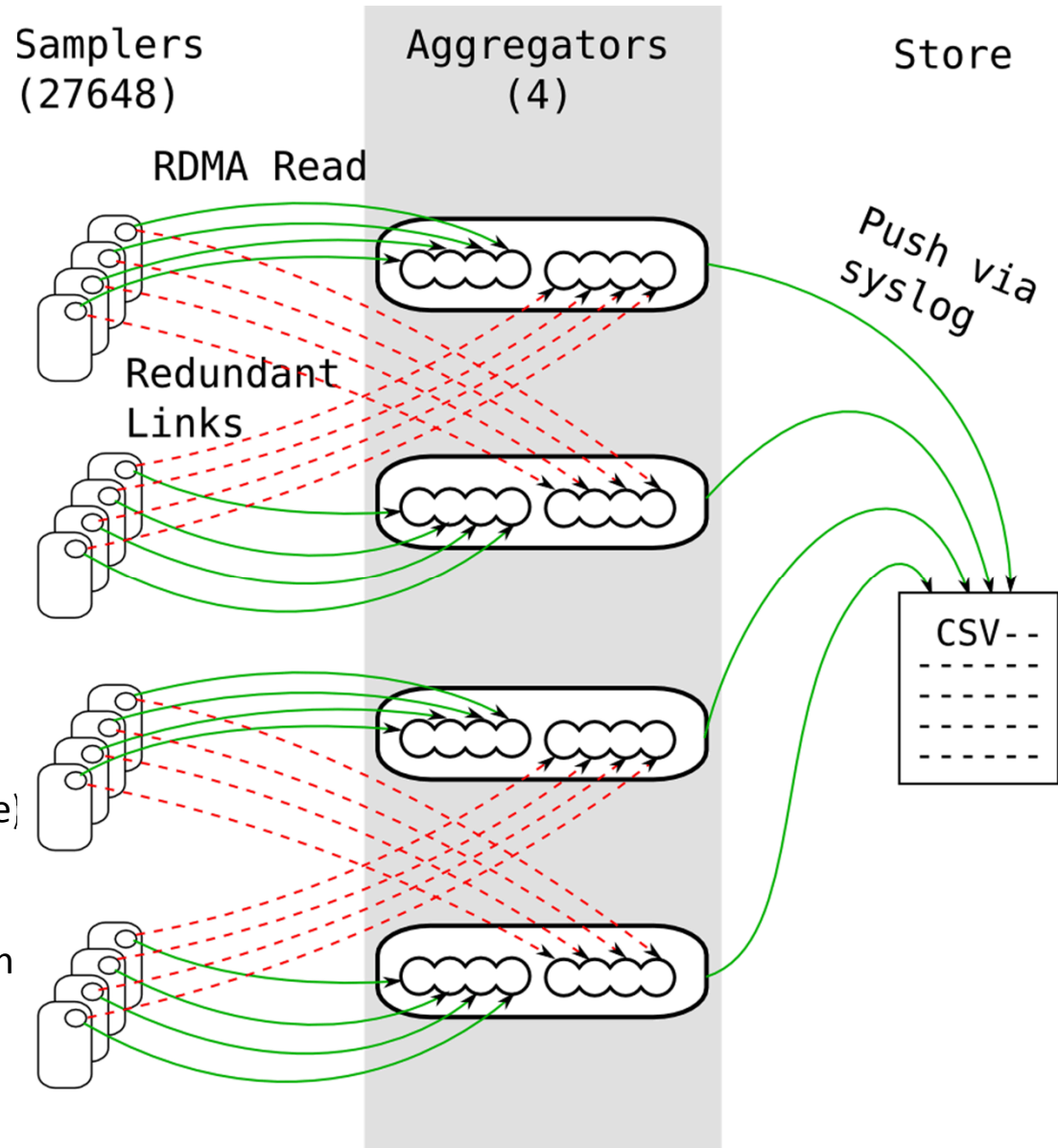
GN : 1735

LDMS metric set Example (data)

```
# ldms_ls -h nid00044 -x ugni -p 412 -l
nid00044/cray_system_sampler_r: consistent, last update: Wed Apr 09 08:52:40 2014 [726us]
U64 1      nettopo_mesh_coord_X
U64 1      nettopo_mesh_coord_Y
U64 6      nettopo_mesh_coord_Z
U64 3265901109447  X-_traffic (B)
U64 21509840670687 Y-_traffic (B)
U64 53884897461291 Z+_traffic (B)
U64 89887627257   X-_packets (1)
U64 475674895649  Y-_packets (1)
U64 1333216704813 Z+_packets (1)
U64 40775903446   X-_inq_stall (ns)
U64 711117651410  Y-_inq_stall (ns)
U64 544039347642  Z+_inq_stall (ns)
U64 48          X-_sendlinkstatus (1)
U64 24          Y-_sendlinkstatus (1)
U64 24          Z+_sendlinkstatus (1)
U64 191         X-_SAMPLE_GEMINI_LINK_BW (B/s)
U64 306         Y-_SAMPLE_GEMINI_LINK_BW (B/s)
U64 344         Z+_SAMPLE_GEMINI_LINK_BW (B/s)
U64 1           X-_SAMPLE_GEMINI_LINK_USED_BW (% x10e6)
U64 2           Y-_SAMPLE_GEMINI_LINK_USED_BW (% x10e6)
U64 2           Z+_SAMPLE_GEMINI_LINK_USED_BW (% x10e6)
U64 19          X-_SAMPLE_GEMINI_LINK_PACKETSIZE_AVE (B)
U64 19          Y-_SAMPLE_GEMINI_LINK_PACKETSIZE_AVE (B)
U64 19          Z+_SAMPLE_GEMINI_LINK_PACKETSIZE_AVE (B)
U64 0           X-_SAMPLE_GEMINI_LINK_INQ_STALL (% x10e6)
U64 0           Y-_SAMPLE_GEMINI_LINK_INQ_STALL (% x10e6)
U64 0           Z+_SAMPLE_GEMINI_LINK_INQ_STALL (% x10e6)
U64 13071017859520 totaloutput_optA
U64 1551040415605 read_bytes#stats.snx11024
U64 111681033094  write_bytes#stats.snx11024
U64 33185713      open#stats.snx11024
U64 33459578      close#stats.snx11024
U64 200           loadavg_latest(x100)
U64 203           loadavg_5min(x100)
U64 2             loadavg_running_processes
U64 217           loadavg_total_processes
U64 32069868      current_freemem
U64 180128670     SMSG_ntx
U64 84138092941   SMSG_tx_bytes
U64 179201767     SMSG_nrx
U64 62591572089   SMSG_rx_bytes
U64 2463841       RDMA_ntx
U64 166910425701  RDMA_tx_bytes
U64 5995457       RDMA_nrx
U64 265128956892  RDMA_rx_bytes
U64 207633071910  ipogif0_rx_bytes
U64 116299863623  ipogif0_tx_bytes
```

Blue Waters Configuration

- All metric sets identical independent of node
 - 194 metrics
- Sample period
 - 60 seconds (normal)
 - 1 second (high)
- Each aggregator primary for 6912 nodes
 - Pull model using RDMA read
- Each aggregator secondary for 6912 nodes
 - RDMA connection established
- In event of failover aggregator collects from 13824 nodes
- Data is pushed to store (MySQL database) using syslog-ng
- One day data set for 60 second collection period contains ~35 million data points per metric and 6.8 billion data points overall



Blue Waters Related Enhancements

- Synchronization
- Minimize Image Footprint
- Node type independent metric set
- Single Metric Set
 - Single Time Attribution
- Storage
 - CSV
 - Split sec and fraction with comma

Synchronous Collection

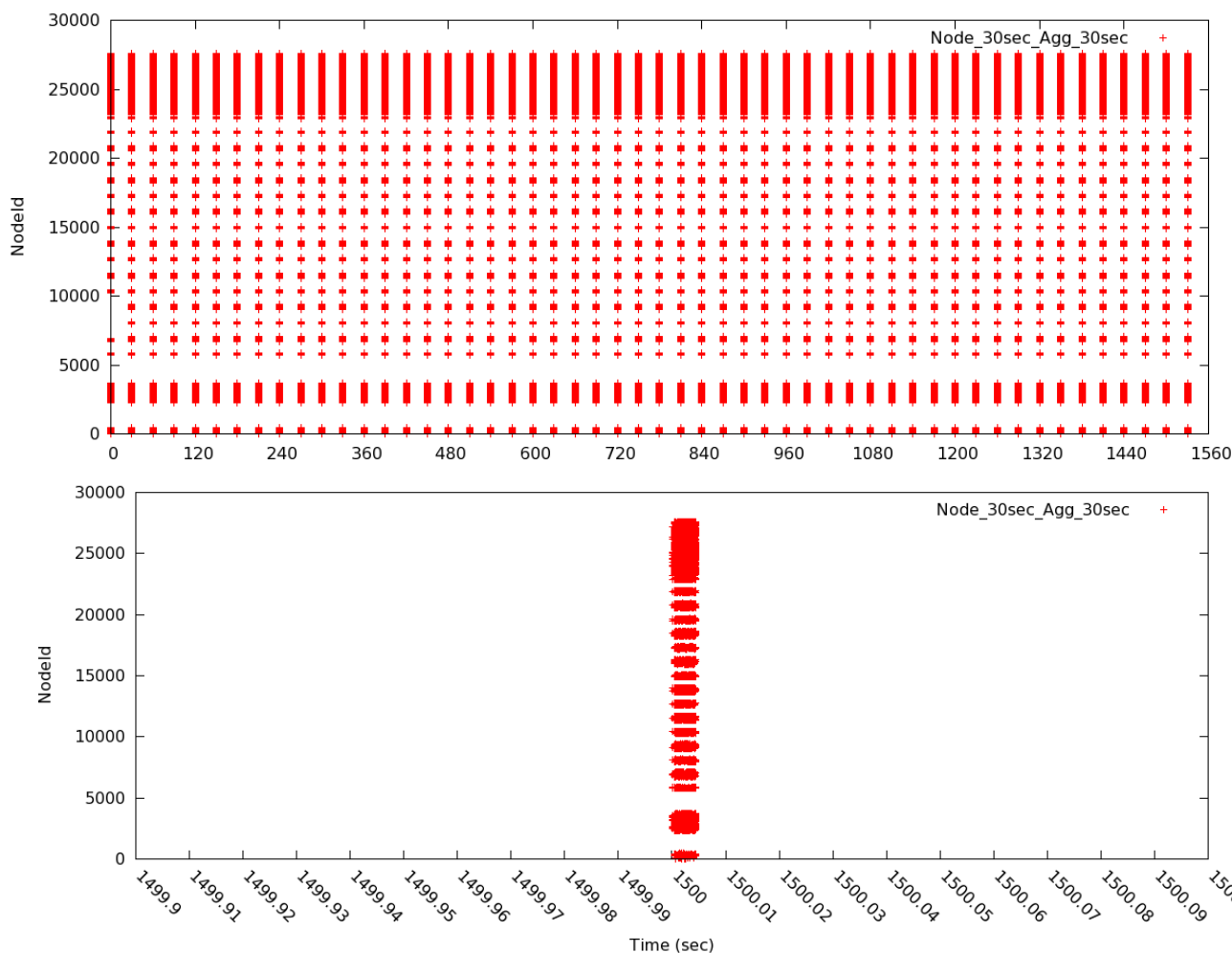
Synchronized collection across all nodes:

- Enables a coherent system snapshot
- Asynchronous* option spreads network load

Synchronous:

- Variance in collection timestamps $\sim 4\text{ms}$

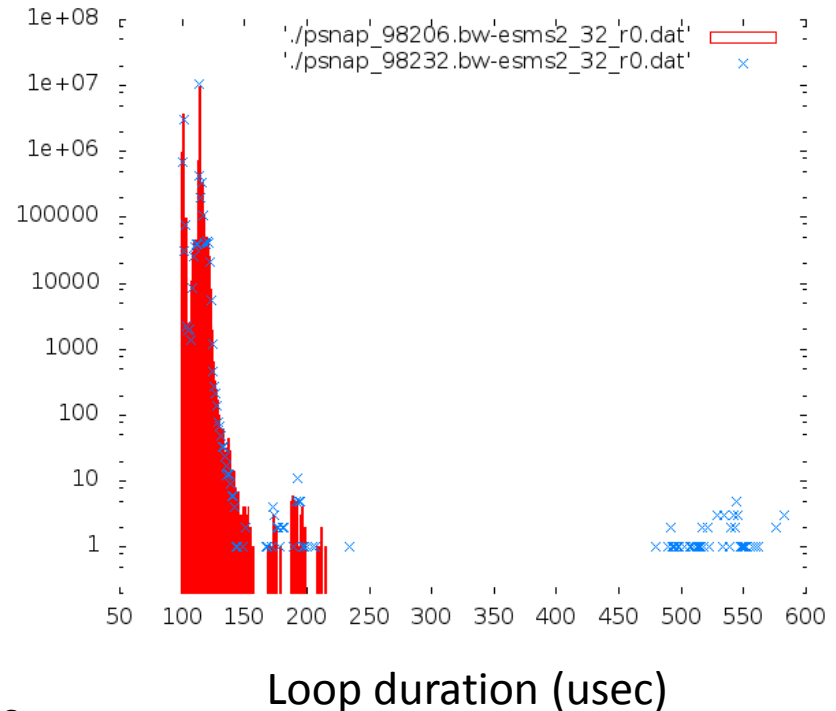
Note: Clock skew not accounted for



Collection occurrences over 10000 nodes on Blue Waters

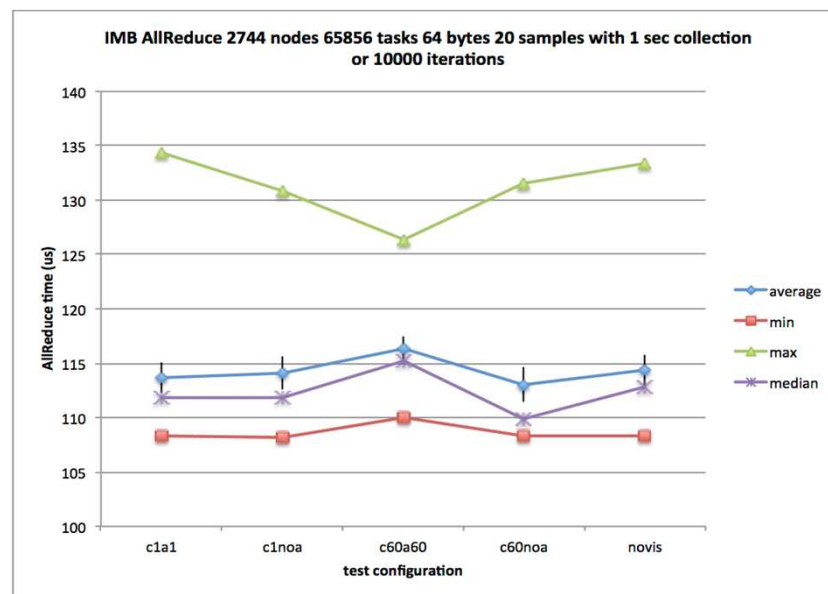
Impact Testing: Benchmarks

- PSNAP
 - No sampling (red)
 - 1 sec sampling (blue)
 - 60/16M points shifted by sampling time of ~450 usec
 - *Effect on app mitigated by synchronized sampling*
- Cray's LinkTest
 - 10,000 iterations of 8kB messages.
 - The no sampling result is 1.7427 msec/packet
 - Sampling result is 20 nanoseconds shorter
 - *No statistical significance*



Impact Testing: Applications

- Intel MPI Benchmark
 - *No correlation of performance with sampling*
- MILC
 - 2774 node run 50 steps
 - 5 phases + Step time
 - *No statistically significant impact*



MILC/C G	novis	c60noa	c60a60	c1noa	c1a1
Ave	5.20e-3	5.21e-3	5.20e-3	5.20e-3	5.19e-3
Min	5.00e-3	5.20e-3	5.00e-3	5.01e-3	5.00e-3
Max	5.43e-3	5.44e-3	5.44e-3	5.45e-3	5.41e-3

Impact Testing: Applications

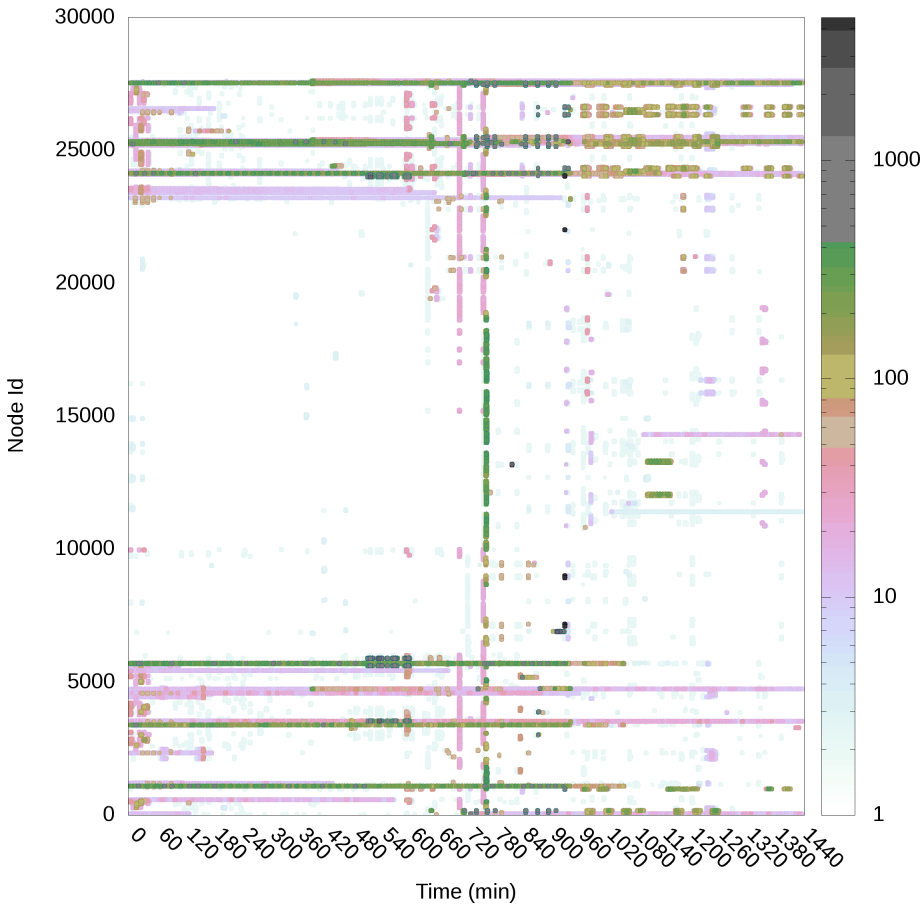
- SNL MiniGhost
 - Instrumented for runtime, communication time, time which includes the barrier
 - 8192 nodes, 3 reps
 - *No statistically significant impact*

Total Runtime	novis	c1a1
Rep1	98.5	92.3
Rep2	95.3	90.2
Rep3	91.8	90.8

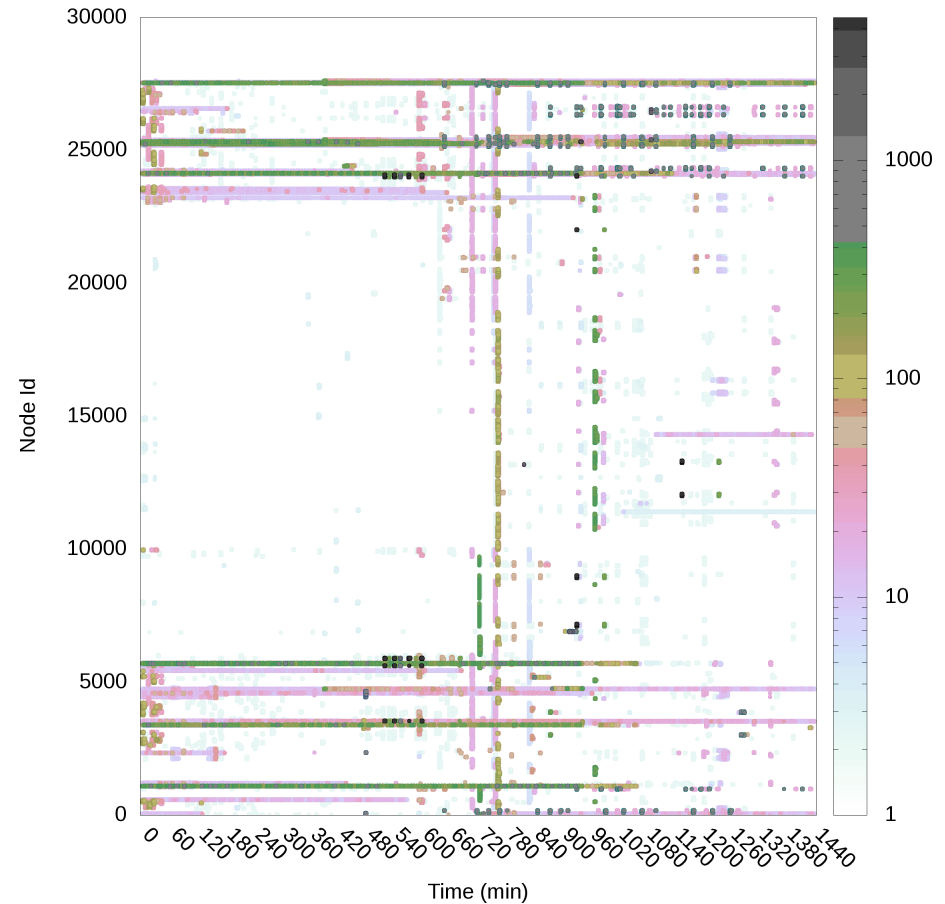
A Look at the Data

Lustre Opens/Closes

snx11001: Opens over 1 min interval

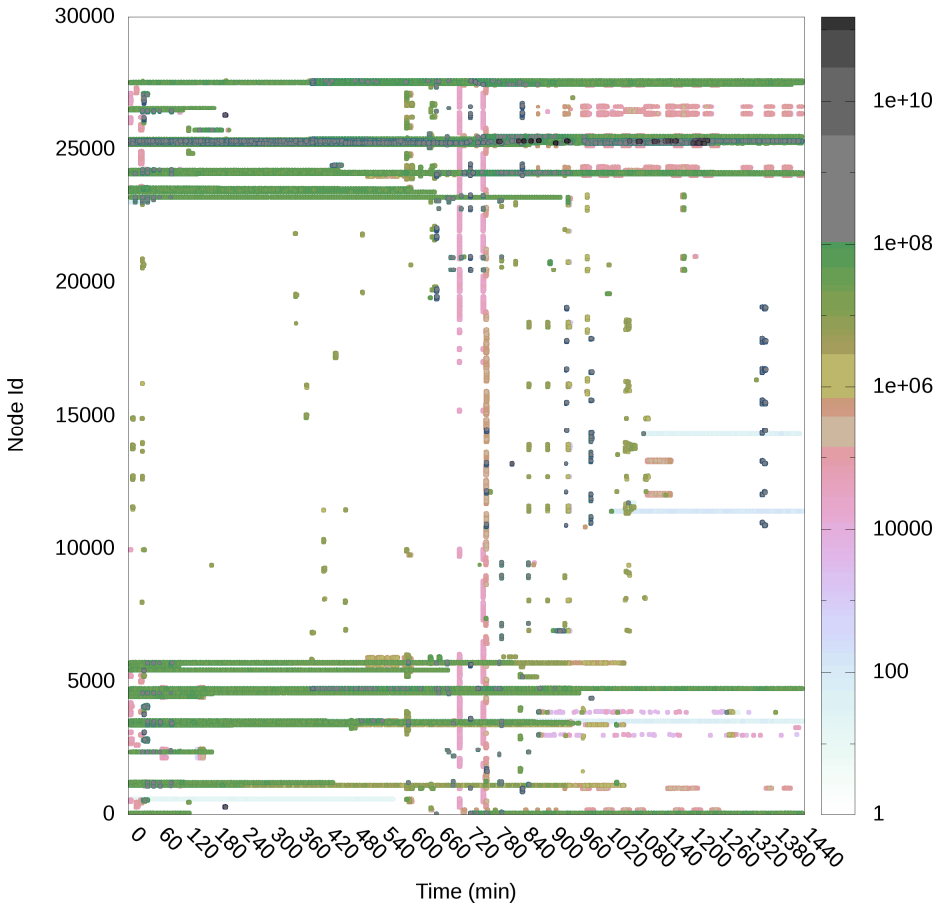


snx11001: Closes over 1 min interval

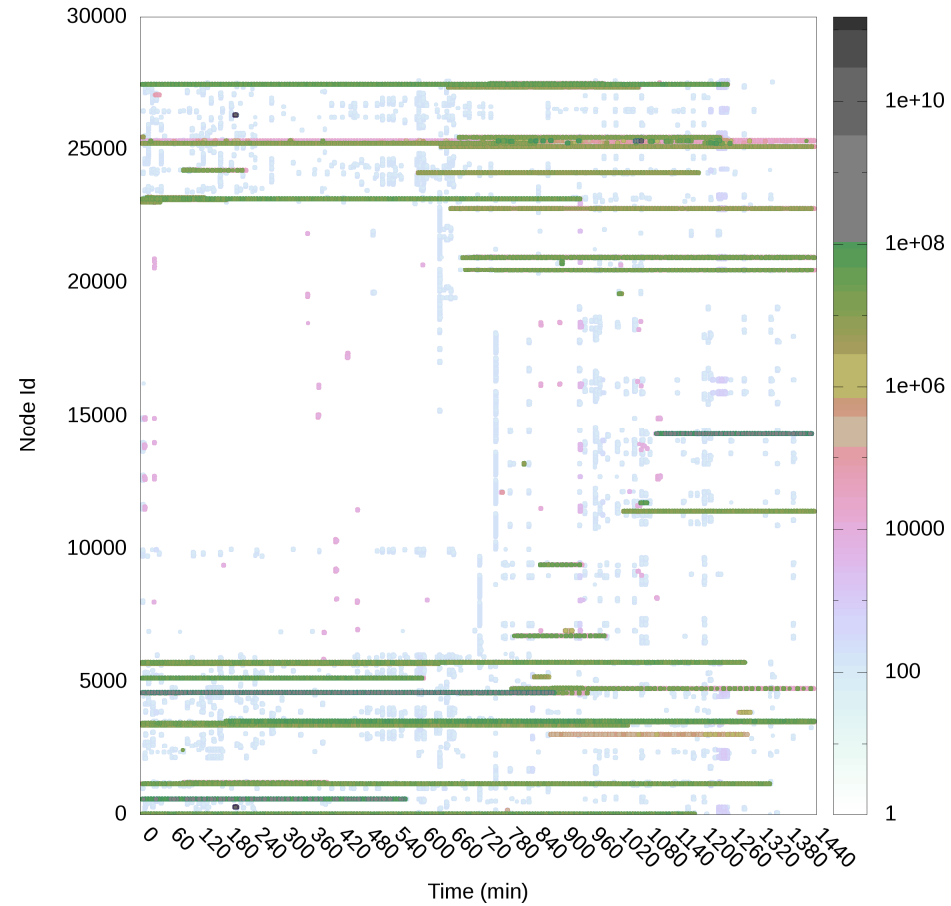


Lustre Reads/Writes

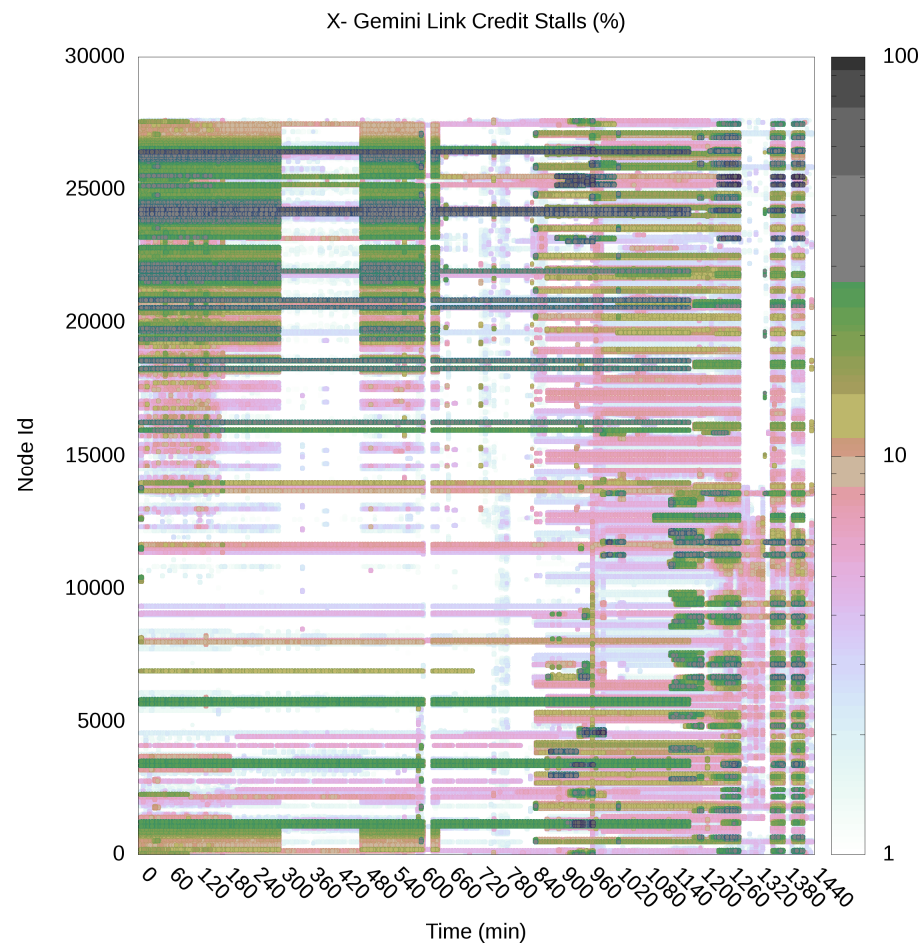
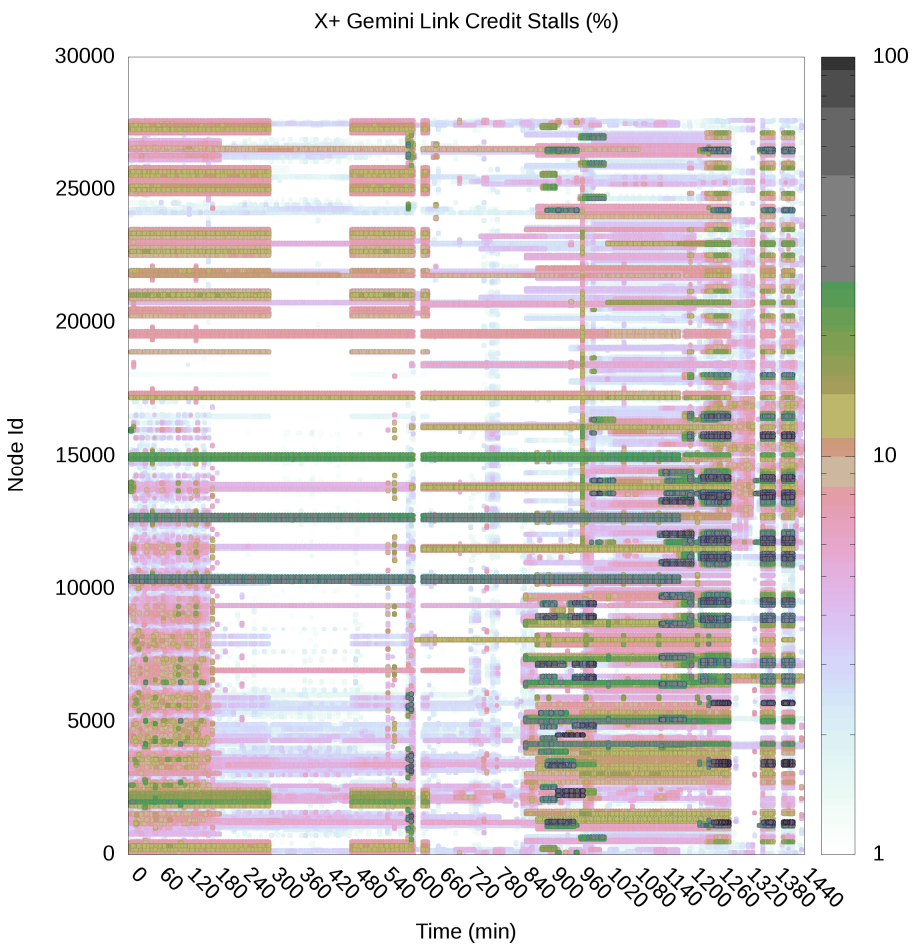
snx11001: Bytes read over 1 min interval



snx11001: Bytes written over 1 min interval

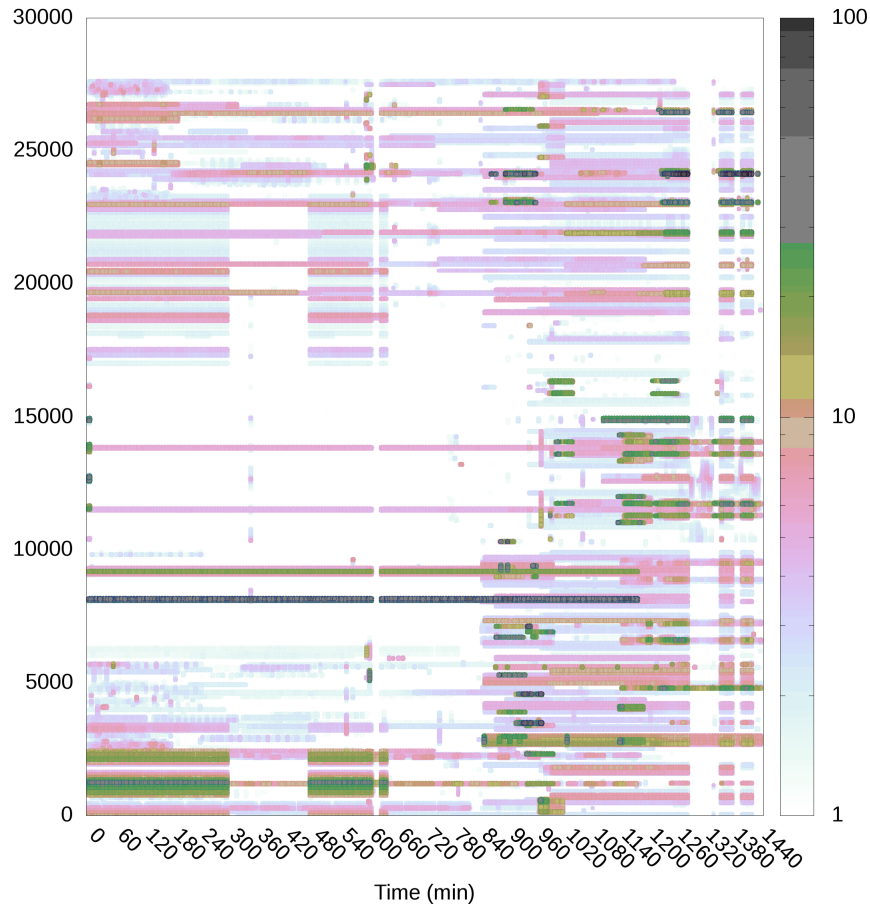


HSN Output Stalls (X)

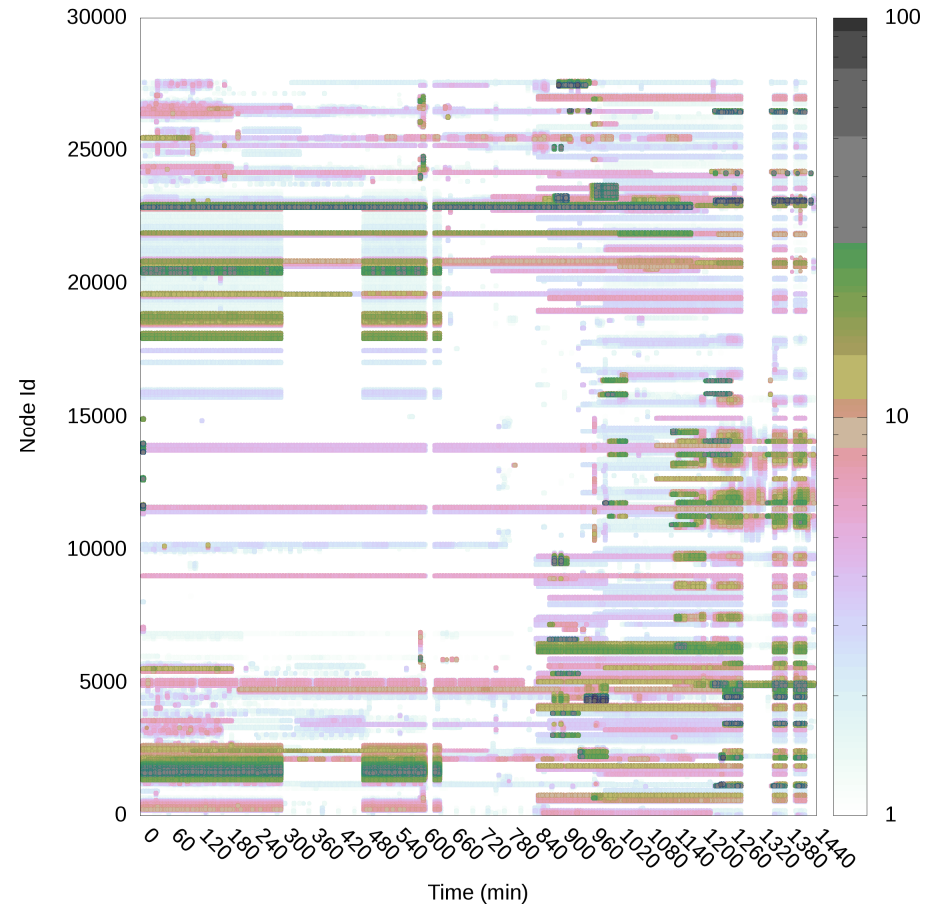


HSN Output Stalls (Y)

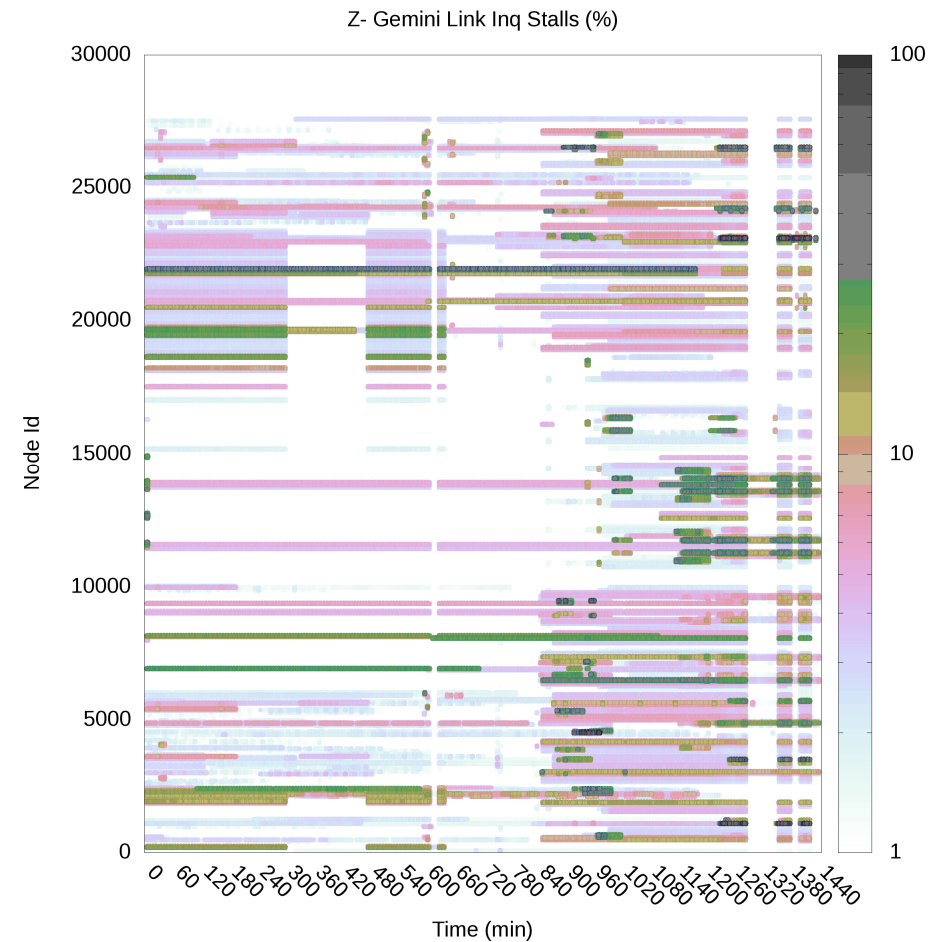
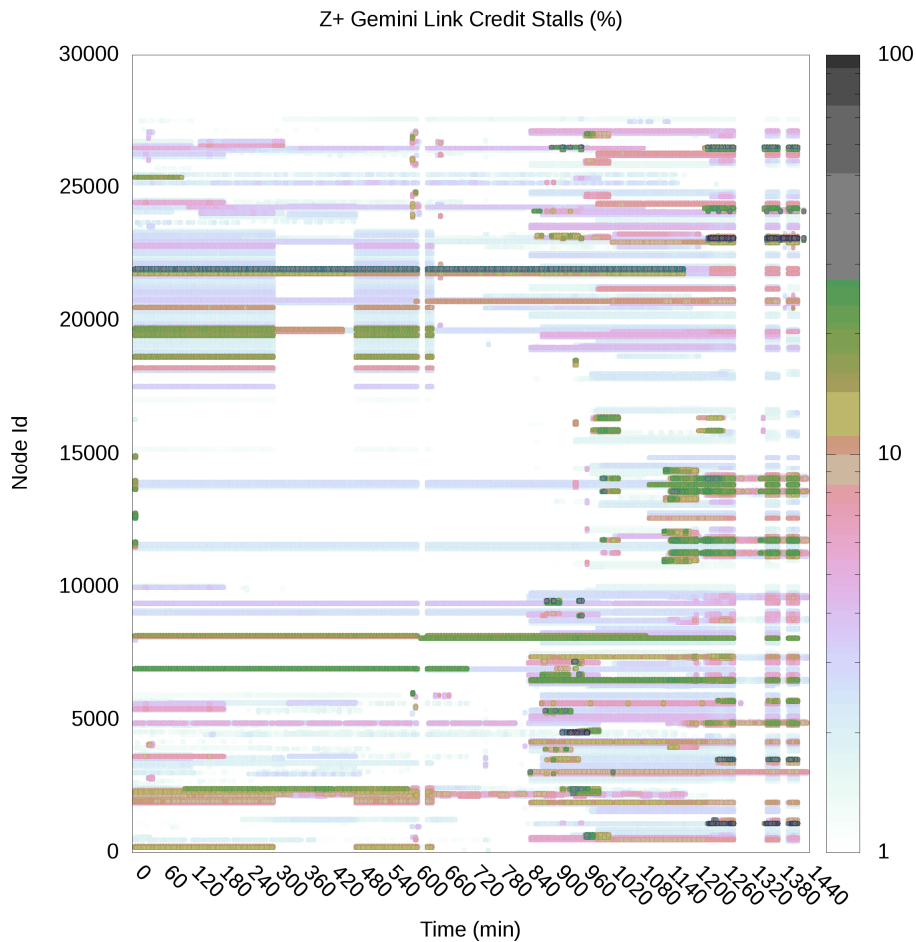
Y+ Gemini Link Credit Stalls (%)



Y- Gemini Link Credit Stalls (%)

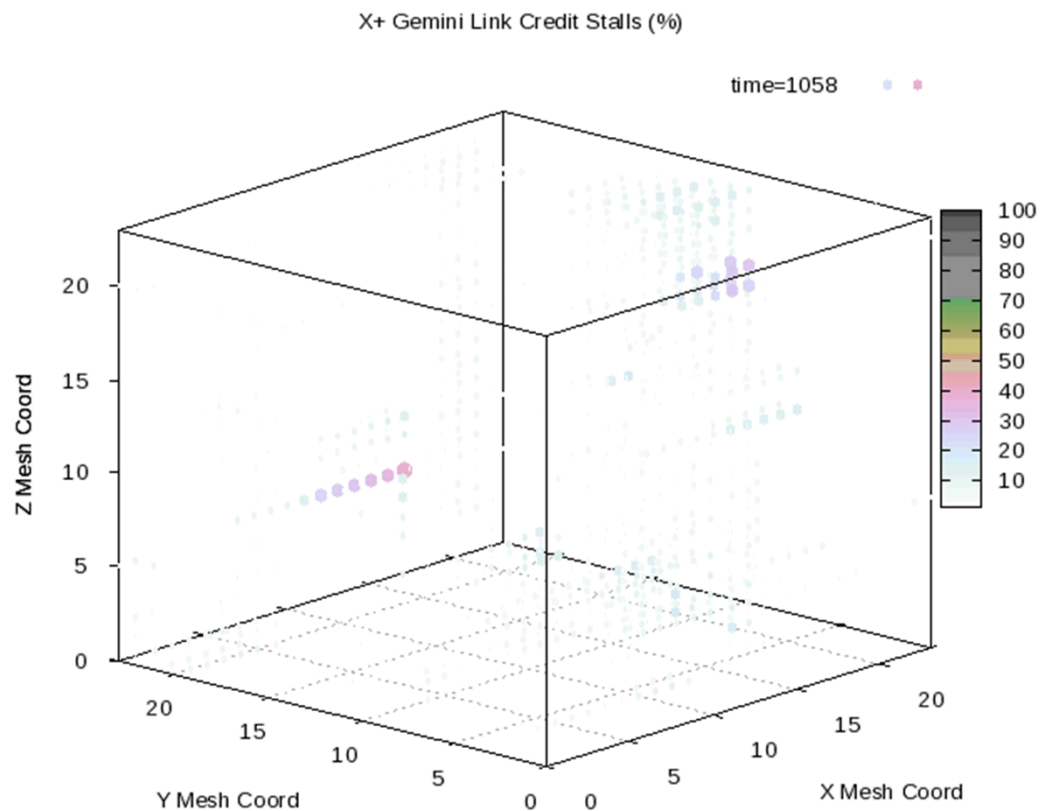
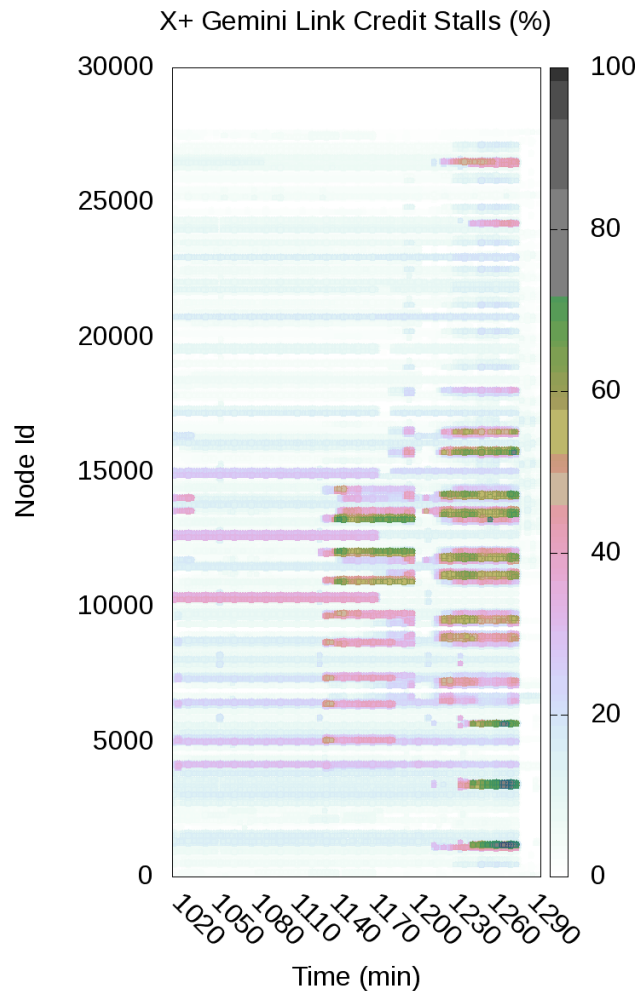


HSN Output Stalls (Z)



Mesh Topology Representation

Animation: 4 hrs @ 1059



Conclusions

- The OVIS data collection, transport, and storage infrastructure provides scalable whole system data access with no statistically significant adverse impact to applications
- Whole system snapshots of shared system resource utilization can provide valuable insights to system and application performance
- We need to develop new analysis and visualization tools to fully utilize the new wealth of data we are collecting

Future Work

- More Tools – both run-time and post processing
 - Analysis
 - Visualization
- Log collection without store for diagnostics
- “Derived Data” plugin
- Separate “connect” thread pool

Questions?