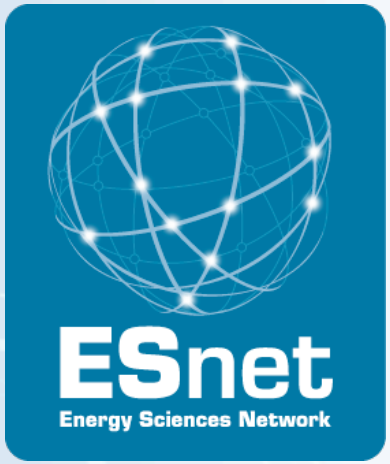


# ESnet Requirements Workshops Summary for Sites

Eli Dart, Network Engineer

ESnet Network Engineering Group

ESnet Site Coordinating Committee Meeting

Clemson, SC

February 2, 2011


# Overview

ESnet requirements workshops – what are they?

Common themes

Discussion of subset of requirements learned

- Examples of trends
- Success stories
- Upcoming needs that are currently unmet

Thoughts for discussion

# ESnet Requirements Workshops

Means by which ESnet is informed of programmatic needs by the DOE Office of Science Program Offices

Each office has a requirements workshop every three years (we do two workshops a year and there are six program offices)

- BES and BER (2007, 2010, … )
- FES and NP (2008, 2011, … )
- ASCR and HEP (2009, 2012, … )

Workshop format – round table discussions of science case studies

- Science is discussed from the point of view of the scientists (though the narrative is network-centric)
- Figure out what the scientists are actually trying to do, then figure out how the network can provide them with services
- Workshop report includes case studies, analysis, program input

# Common Themes - Science

New science processes such as remote instrument control, experiment health monitoring, etc will place new demands on networks

- Multi-site near-real-time or real-time network interaction
- Need expressed by multiple science communities (light sources, biology, HPC users, etc)
- Many of these communities are not network experts, and will need help from networking organizations in order to progress

Increasing data intensity of science across many disciplines

- Many collaborations that have historically not used the network for data transport must begin soon – sneakernet will no longer be practical
- Many collaborations that have gotten by with using SCP/rsync/etc for WAN transfers will no longer be able to do so – must change to GridFTP or something similar to increase performance

# Common Themes - Troubleshooting


Still many performance problems due to packet loss

- Networks aren't clean
- Figure out how to clean them up and keep them clean
  - perfSONAR (see also DICE Diagnostic Service talk)
  - Architectural changes (see Science DMZ talk)

Many scientists/collaborations are not network experts, and it is unreasonable to expect them to become experts

- We in the networking community must reach out to the science community
- Networks are key to the emerging modes of scientific discovery
  - Widely distributed collaboration
  - Machine-consumable / programmatic interfaces to data
  - Massive data volumes → automated data handling and analysis
- There is a lot of scientific leverage here, but only if the network is an effective scientific tool

# HEP – LHC

Large data sets (transfers of tens of terabytes are routine)

Automated data distribution over multiple continents

LHC collaborations are large and well-funded, but they have their own challenges

- Network monitoring, tuning, and troubleshooting is difficult
- Many site and regional networks still drop packets
- Lessons here for others
  - Networks must be clean
  - Hosts must be tuned (the defaults are still wrong for WAN)
  - Use proper data transfer tools for WAN

# NP – RHIC at BNL


STAR

- Widely distributed collaboration
- Significant data transfers to NERSC
- Significant international data transfers
- Over 2PB of raw data to be produced in 2011, 400TB+ of derived data sets to be distributed

PHENIX

- 2008 data rates from ~650Mbps to 2.4Gbps to Japan (118 TB)
- Near-real-time transfer need (data sets have short lifetime on cache disk, less than 24 hours)
- Data set sizes increasing (projected to transfer 1.2PB to Japan in 2011)

# BES – Light and Neutron Sources


ALS at LBL, APS at ANL, LCLS at SLAC, NSLS at BNL, SNS at ORNL, etc.

Large number of beamlines, instruments

- Hundreds to thousands of scientists per facility
- Academia, Government, Industry

Data rates have historically been small

- Hand-carry of data on physical media has been the norm for a very long time: CDs → DVDs → USB drives
- Scientists typically do not use the network for data transfer

Near future: much higher data rates/volumes

- Next round of instrument upgrades will increase data volumes by a factor of 10 to a factor of 100, e.g. from 700GB/day to 70TB/day
- Network-based data transport is going to be necessary for thousands of scientists that will be doing this for the first time in their careers

# BES – Light and Neutron Sources


New science architectures coming

- Experiment automation leads to the need for near-real-time health checks
  - Stream sample experiment output to remote location for verification of experiment setup
  - Significant efficiencies of automation are driving this
- Multi-site dependencies (e.g. need for analysis at supercomputer centers)
  - Need a general model for streaming from detectors to supercomputer centers
  - Supercomputer centers often say that allocations change from year to year, therefore significant effort to support one particular scientist may not be wise resource allocation
  - However, many light source users will need to stream data to supercomputer centers – generalized support for this use model will result in significantly increased scientific productivity

# BES – Light and Neutron Sources


Some of these data increases have already taken place

Dedicated data transfer hardware and perfSONAR have been used to fix performance problems

- Networks must be loss free
- Networks must be monitored to ensure that they stay clean

These solutions will need to be generalized

- Science DMZs and/or Data Transfer Nodes (DTNs) for light sources
- Assist users with figuring out the “other end” (e.g. suggestions for common architectures such as DTN or Science DMZ)
- Requiring that every collaboration implement their own solution (as many light sources do currently) will result in tens of one-offs over the next few years
  - Difficult to troubleshoot
  - High support load for facility, system and network support staff
  - Therefore, leadership now is in our collective best interest

# BER – Climate Science

Supercomputer centers at NERSC and ORNL

Data repositories at LLNL, ORNL, NCAR, NOAA, NCDC, BADC (UK), Germany, Japan, Australia

2PB (~1.6PB and growing) to replicate over multiple continents in 2011

They are behind schedule, lots of performance issues (they will need our help)

- ESG data nodes going up at LLNL, NERSC, NCAR, BADC, etc.
- Network performance tuning is not part of the climate community's institutional knowledge
- Proactive participation by networking folks will be needed (find and help your local climate science group)

# BER - Genomics

JGI does a lot of sequencing and analysis

- Scientists send in samples to be sequenced
- JGI does sequencing/analysis, sends back genome

Significant changes coming

- Price of sequencing equipment dropping dramatically (~$500k to ~ $50k)
- Data rates going up dramatically (1PB/year today, sequencing machine output to go up by as much as 12x over 5 years)

# BER - Genomics

Genomics is about to be stood on its head

- Many sites will be able to deploy their own sequencers as costs come down
- Many will not have the local processing/storage infrastructure or systems expertise to do their own analysis
- Instead of sending biological samples to JGI, many sites will send raw sequence data (much larger than the completed genome) to JGI → for analysis significant networking component

Sites with sequencers can expect larger data flows to/from JGI

# FES - Experiments

Data collection/transfer between EAST in China and DIII-D at GA is a good illustration of the value of dedicated DTNs

- Dedicated servers make data transfer possible
- Two GridFTP boxes built, one shipped to EAST, one deployed at GA
- Data transfers now keep up with experiments
- This is probably a good fit for future Fusion experiments also (FES workshop to be conducted later this year)

# FES - Simulation

Simulations can generate data sets of essentially arbitrary size (e.g. GTC code running at ORNL is expected to generate 500TB/week in a few years)

- Full or reduced data sets must be transferred to other sites (e.g. NERSC, Princeton) for analysis
- Sites that consume fusion data sets from supercomputer centers might want to consider DTNs for their fusion groups if such infrastructure does not currently exist

Fusion Simulation Program (FSP)

- FSP will require integration of multiple sites to run distributed simulation codes

# Discussion Topics


Many scientists are expressing the need for multi-site services

- Unlike HEP and NP, most will not be able to implement these for themselves – BES facilities in particular are going to need a lot of help here
- However, the potential scientific payoff is huge (networks will be able to legitimately claim to have enabled the next round of breakthroughs in materials, biology, energy efficiency, etc)

Significant increase in nominal performance expectations

- For a significant user population, SCP/SFTP won't cut it anymore
- Security policies, architectures, and tools deployments must evolve
- Networks *must* be clean!

End to end infrastructure focus (this means multi-site planning)

# Questions?

Thanks!

# Questions?

Thanks!