ESnet

ENERGY SCIENCES NETWORK

Data Mobility for Fusion Science

Energy Sciences Network (ESnet)

Lawrence Berkeley National Laboratory

Jason Zurawski, Science Engagement Engineer

December 16, 2021

U.S. DEPARTMENT OF ENERGY

Office of Science

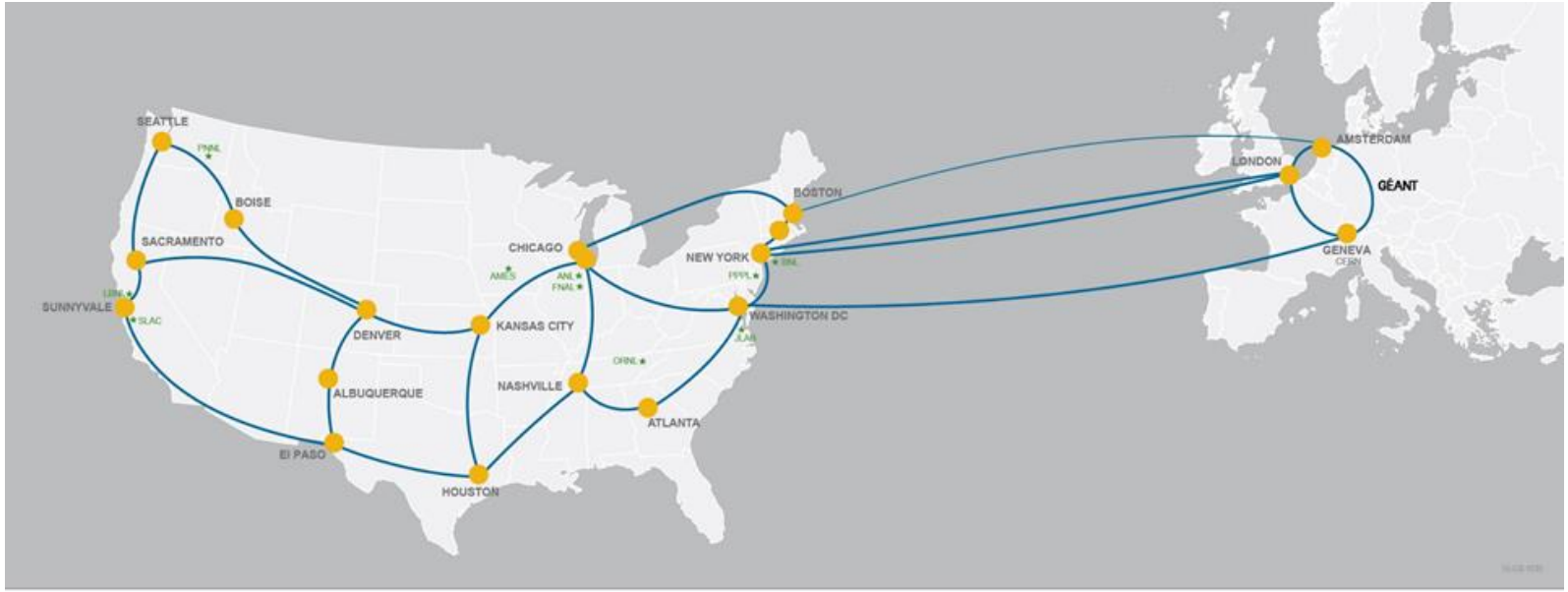
BERKELEY LAB

Topics

- ***ESnet: Introduction & Motivation***
- Understanding Scientific Requirements
- Digging Deeper: Data Mobility
- How We Can Help
- Conclusions

ESnet

ESnet is DOE's high performance network user facility



ESnet
ENERGY SCIENCES NETWORK

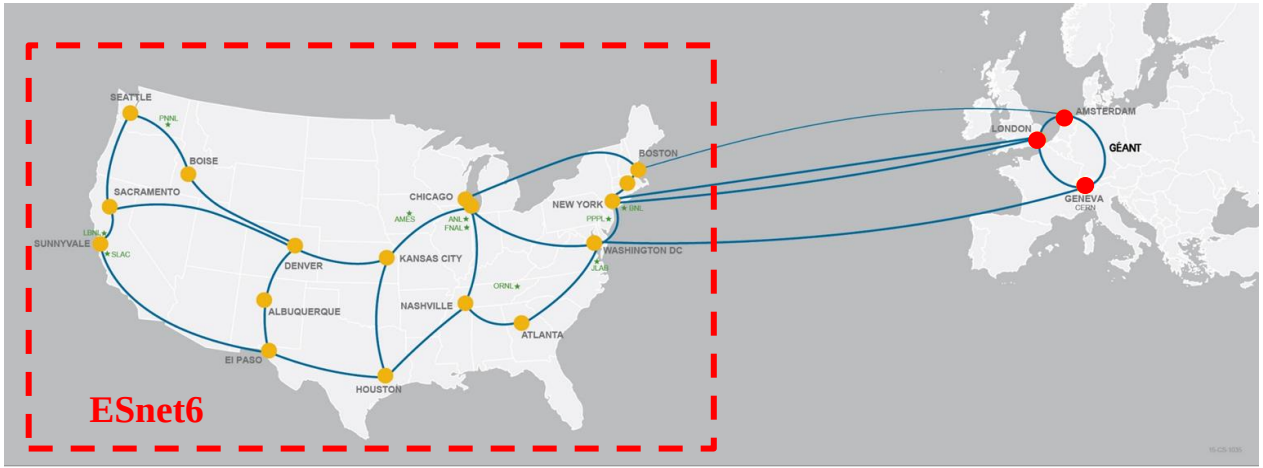
★ Department of Energy Office of Science National Labs

Ames	Ames Laboratory (Ames, IA)	LBNL	Lawrence Berkeley National Laboratory (Berkeley, CA)
ANL	Argonne National Laboratory (Argonne, IL)	ORNL	Oak Ridge National Laboratory (Oak Ridge, TN)
BNL	Brookhaven National Laboratory (Upton, NY)	PNNL	Pacific Northwest National Laboratory (Richland, WA)
FNAL	Fermi National Accelerator Laboratory (Batavia, IL)	PPPL	Princeton Plasma Physics Laboratory (Princeton, NJ)
JLAB	Thomas Jefferson National Accelerator Facility (Newport News, VA)	SLAC	SLAC National Accelerator Laboratory (Menlo Park, CA)

The ESnet Team is ~100 People as of December 2021

ESnet

ESnet6 – Next Generation Network



To satisfy Mission Need, the ESnet6 project scope resolves capability gaps:

1. Increasing basic network capability and capacity while replacing end of life equipment
2. Automating network operations and improving security services
3. Providing programmable network flexibility to support evolving computation and data models

ESnet

Topics

- ESnet: Introduction & Motivation
- ***Understanding Scientific Requirements***
- Digging Deeper: Data Mobility
- How We Can Help
- Conclusions

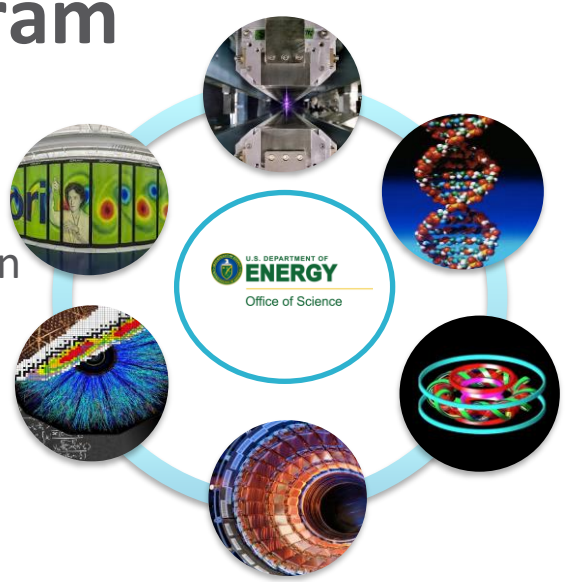
ESnet Mission

- ESnet's core mission is to serve the DOE/SC science programs.
  - Large-scale data movement
  - Network services to enable science
- To accomplish this, ESnet attempts to understand the needs of the science communities ESnet serves (including FES).
- Network implications arise from the conduct of science
  - Science instruments and facilities
  - Process of science – and how it integrates to technology (e.g. Computing, Storage, Sharing)
  - How will these change over time?
- ESnet and the DOE Science Complex must have a common strategic vision of the network needs of the science.

ESnet Requirements Review Program

ESnet's core partnership program created to comprehensively evaluate:

- ***Major science experiments and facilities***, both in operation and planned.
- ***The process of science*** used for knowledge discovery, and including scientists' interactions with the instruments and facilities.
- ***The volume of data*** produced now, and anticipated in the future, with an emphasis on geographical location of where the data must be shared, computed and/or stored.
- ***The current technology capabilities*** (network, computation, storage, and software stack) used by each science collaboration/facility as well as any planned upgrades, additions or improvements.



ESnet

Rough Findings – Built from Years of Study

- Networks are an essential part of data-intensive science
  - Connect data sources to data analysis
  - Connect collaborators to each other
- Performance is critical, but often overlooked
  - Exponential data growth
  - Constant human factors
  - Data movement and data analysis must keep up
- Effective use of wide area (long-haul) networks by scientists has historically been difficult

ESnet

Topics

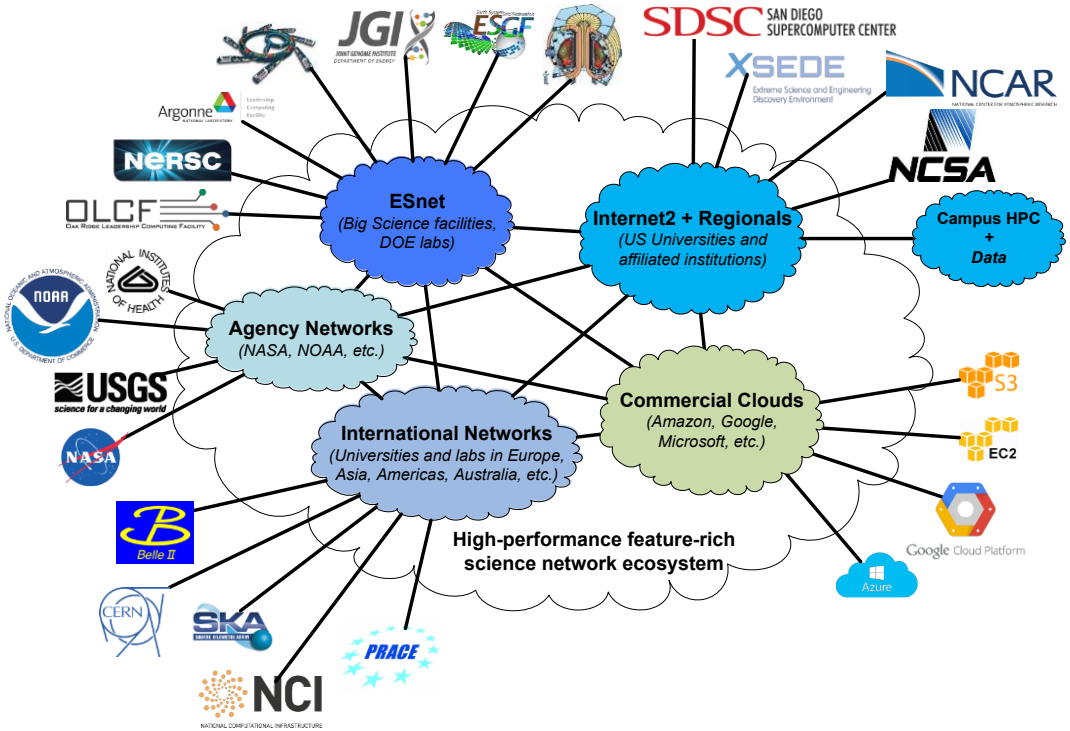
- ESnet: Introduction & Motivation
- Understanding Scientific Requirements
- ***Digging Deeper: Data Mobility***
- How We Can Help
- Conclusions

'Data Mobility' & Computation are Related

- FES experimental locations, and DOE HPC facilities, exist in a larger scientific ecosystem
  - Collaborators
  - Data repositories, data portals, science gateways
  - Site-local resources (e.g. mid-scale clusters)
  - Commercial clouds
- In this context, "Data Mobility" means putting science data where it needs to be in order for the data to be useful
  - Available to a running job; staging for a future job
  - Move data to compute or move compute to data
- In order to be scalable, we must do this using networks
  - Hand-carried hard drives don't scale
    - (Google, Amazon, et. al. do not ship their stuff around on hard drives!!)
  - Data transfer performance must scale with the science need

ESnet

Science Networks Interconnect Facilities



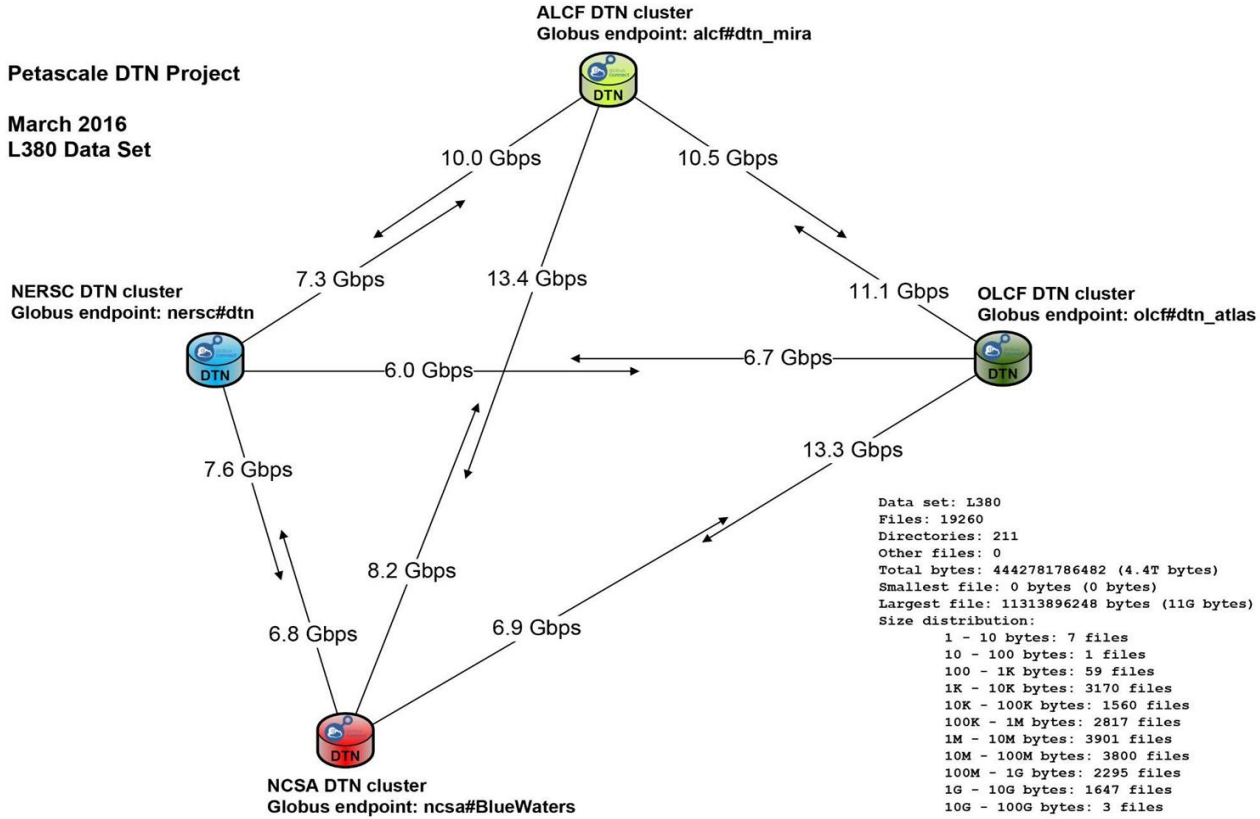
ESnet

Moving Data Between Facilities

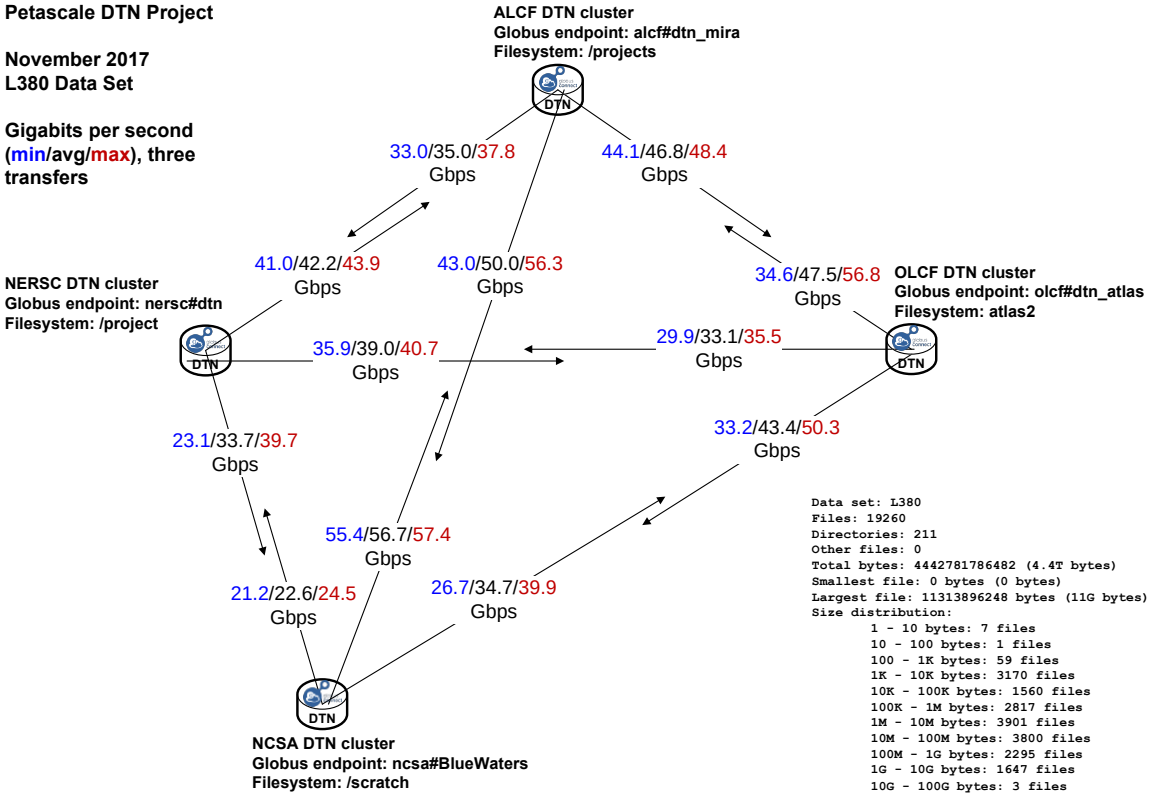
- In order to move data to/from/between facilities, the facilities must interface with the wide area network
  - Historically, this has been difficult
  - Accepted best practice exists
- Efforts like the “Science DMZ” architectural paradigm provides an framework
  - Flexible way of using best practice
  - Accommodates a wide range of data transfer and data access workflows
  - Uses simple ‘off the shelf’ approaches – known software and hardware configurations

ESnet

Data Movement Capabilities (2016)



Data Movement Improvements (2017)



Topics

- ESnet: Introduction & Motivation
- Understanding Scientific Requirements
- Digging Deeper: Data Mobility
- ***How We Can Help***
- Conclusions

ESnet

ESnet has Dedicated Staff to Assist with Data Mobility

- Evaluating the Needs
  - Understanding data volumes, where it needs to go/come from
  - Changes in the short/middle/long term
- Evaluating the Current Capabilities
  - How is data movement to collaborators/facilities done?
    - Mailing Disks?
    - Cloud Tools (Dropbox)?
    - Software transfer (rsync, ftp)?
  - How does the current solution work?
    - Can/should it be improved?

ESnet

ESnet has Dedicated Staff to Assist with Data Mobility (cont.)

- Architectural Suggestions
  - Network improvements (e.g. security, design, usage patterns, peering)
  - Server/Software improvements (e.g. data transfer tools for science, hardware that works well, workflow changes)
- Testing and Improving
  - Test the network, test the tools
  - Evaluate and change
  - Improvement is ***always*** possible

Topics

- ESnet: Introduction & Motivation
- Understanding Scientific Requirements
- Digging Deeper: Data Mobility
- How We Can Help
- ***Conclusions***

Summary

- ESnet is the DOE's High Performance Science Network
  - Available to anyone/everyone that is doing some form of DOE Science (you are probably using it and don't know it)
  - Links major facilities to the outside world
- ESnet can assist with data mobility needs for FES
  - Help in moving data to/from DOE facilities
  - Help with Software tools, Hardware Configs, testing, debugging
- Shoot us a note – we are happy to talk (even if we may not be able to completely solve a problem)

engage@es.net

ESnet

ESnet

ENERGY SCIENCES NETWORK

Data Mobility for Fusion Science

Energy Sciences Network (ESnet)

Lawrence Berkeley National Laboratory

Jason Zurawski, Science Engagement Engineer

December 16, 2021

U.S. DEPARTMENT OF ENERGY

Office of Science

BERKELEY LAB