ESnet

ENERGY SCIENCES NETWORK

Unique instrument enabling big-data science

Inder Monga

Director, Energy Sciences Network

Director, Scientific Networking

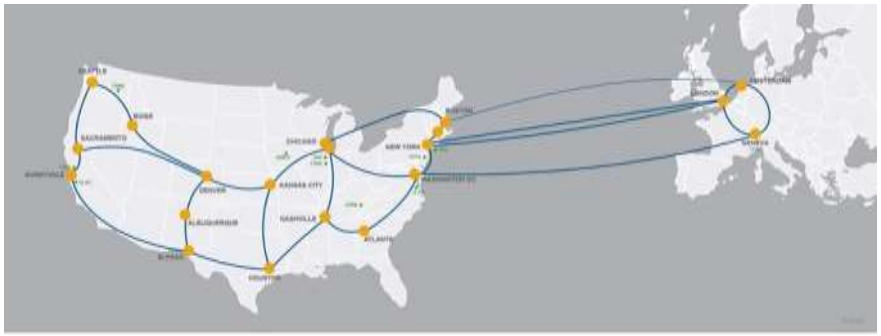
Lawrence Berkeley National Lab

BERAC

October 19th, 2018

Talk

ESnet Introduction



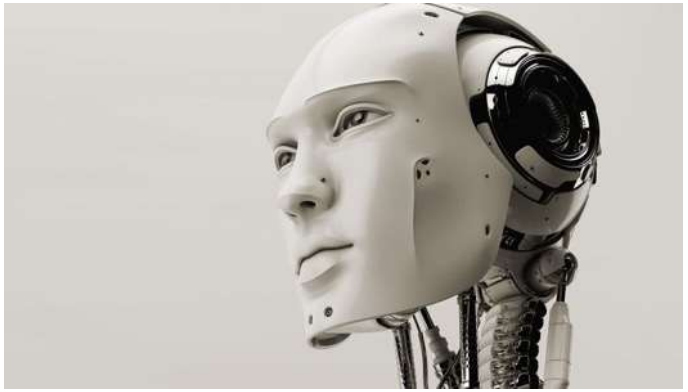
Scaling with Design Patterns

Future Directions

ESnet

Networks are central to all ‘smart’ human life

facebook



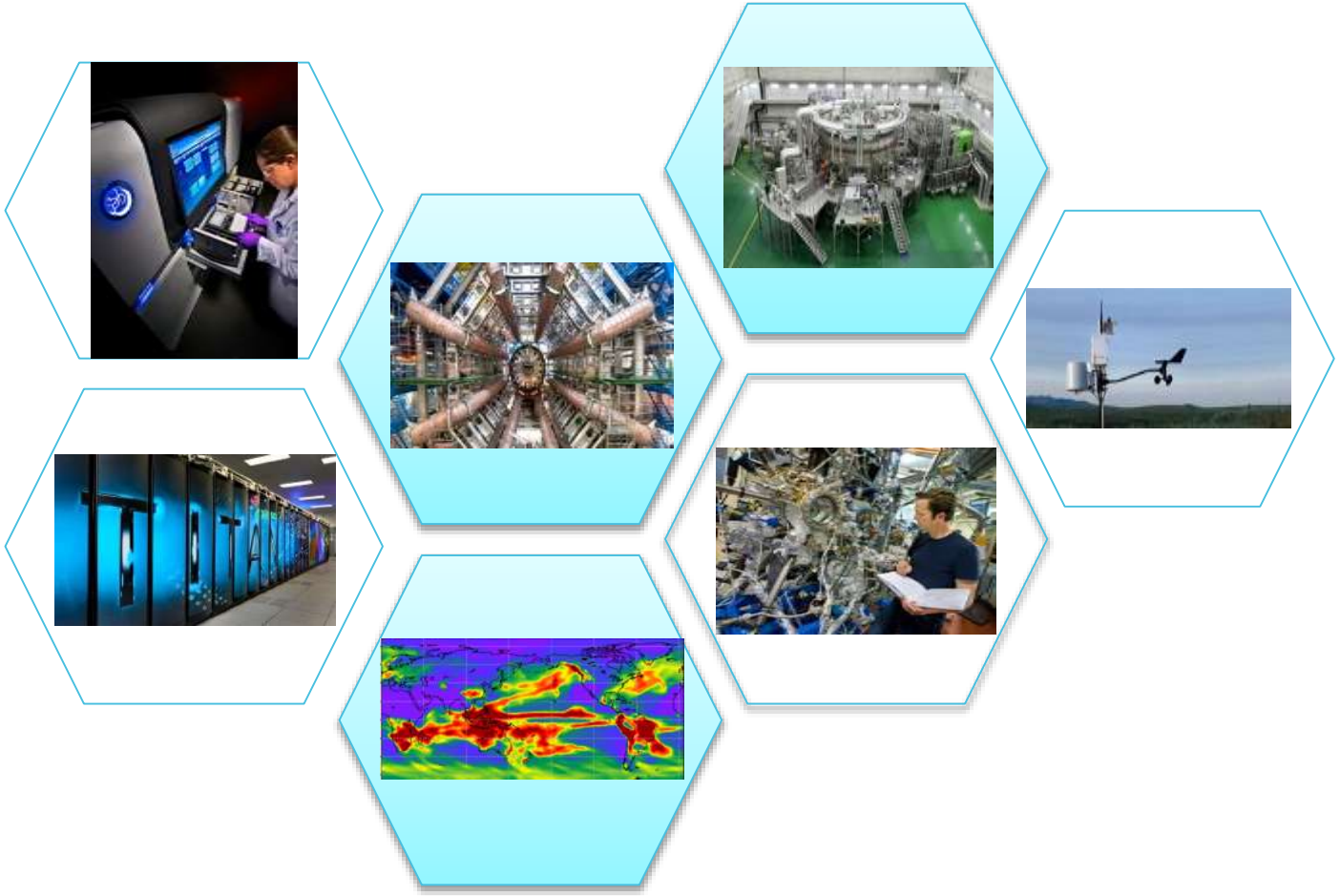
Google

amazon

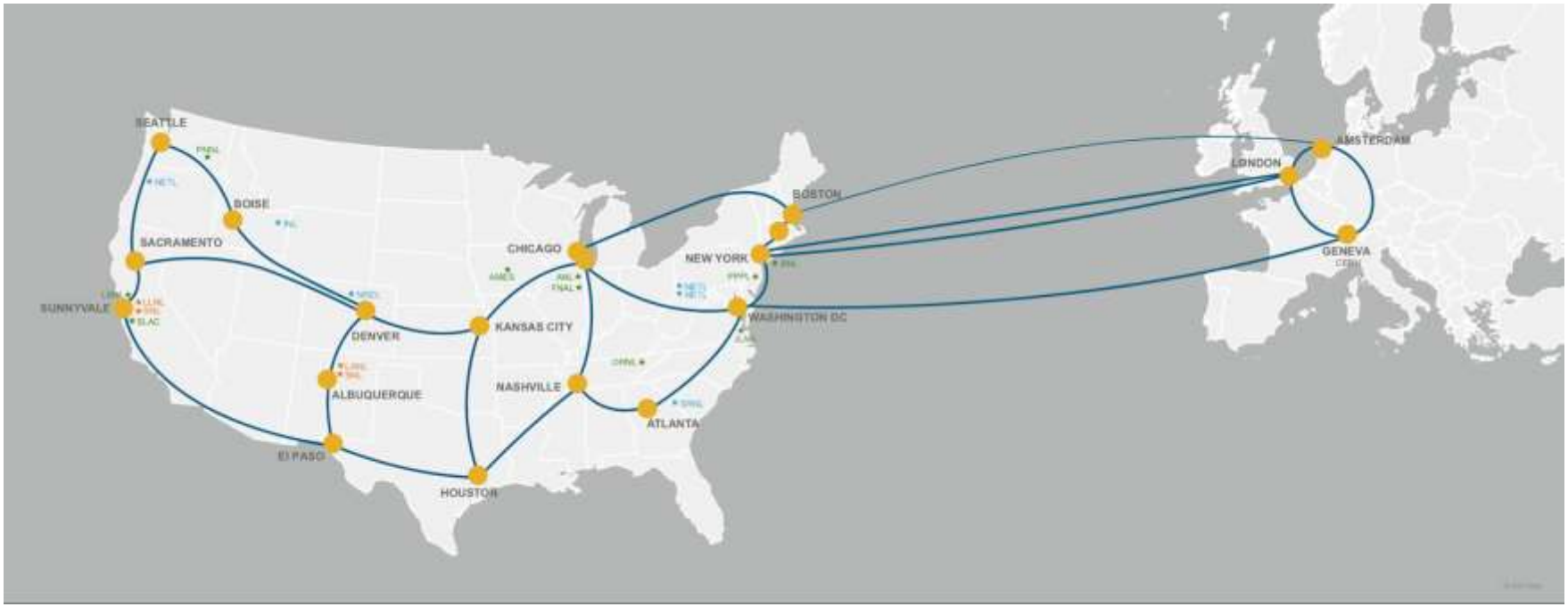
Artificial Intelligence
Machine Learning

ESnet

Additionally,
Networks are central to science collaborations



DOE's high-performance network (HPN) user facility optimized for enabling big-data science



ESnet provides connectivity to
all of the DOE labs, experiment sites, & supercomputers

Our vision:

Scientific progress will be completely unconstrained by the physical location of instruments, people, computational resources, or data.

ESnet

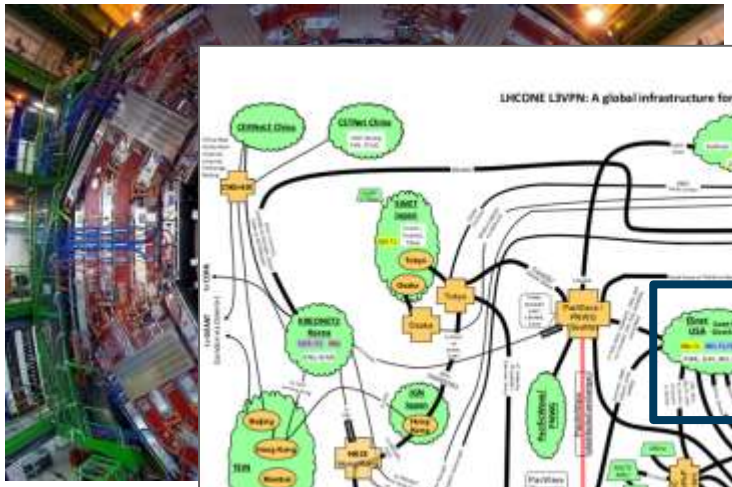
Global partnerships and network connections key to meeting mission

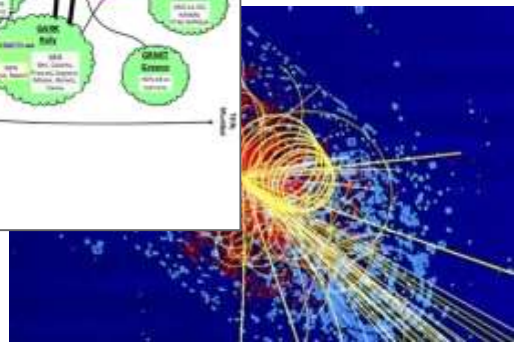
80% of carried traffic originates or terminates outside the DOE complex

Serve all interests: Commercial peers, private peering with popular cloud providers, R&E networks worldwide, regionals, universities, agencies etc.

Global science collaborations like LHC depend on high-speed networking for science discovery

Example 1: High Energy Physics / Large Hadron Collider Science



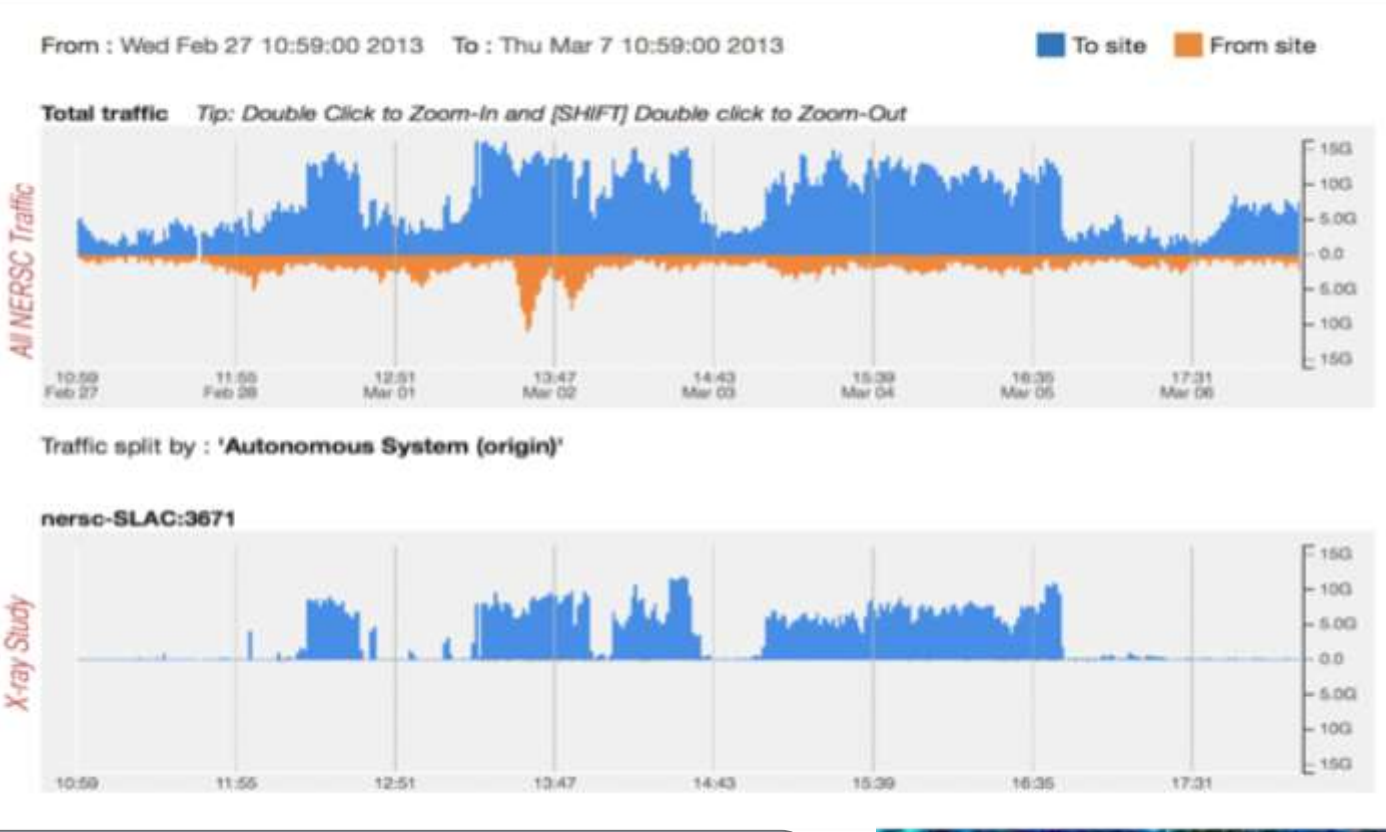


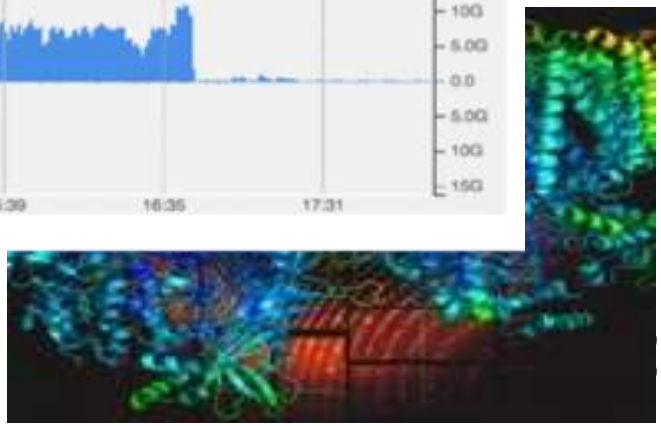
ESnet supports high-performance data movement to/from LHC in CERN, Switzerland to FNAL and BNL (Tier 1 sites) and 20 other universities

Discovery of

High-performance data movement needed to access supercomputing resources in near real-time

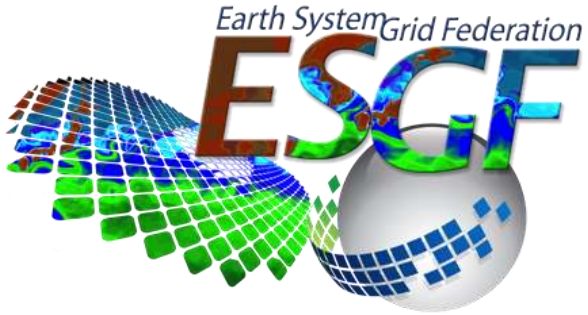
Example 2: Basic Energy Sciences / LCLS at SLAC





A single LCLS run of the Photosynthesis II experiment, representative of future LCLS II workflow, generated 3x times the usual traffic on the network

ESnet/BER Science Partnerships: ICNWG



- International Climate Network Working Group created in 2014
  - Started as part of the Enlighten Your Research Global program
  - Now an ESGF working group
- Purpose: improve data transfer performance between climate data facilities
- Current focus: data replication between Tier1 data centers
- ESnet engagement has brought data portal architecture and performance engineering expertise to ESGF

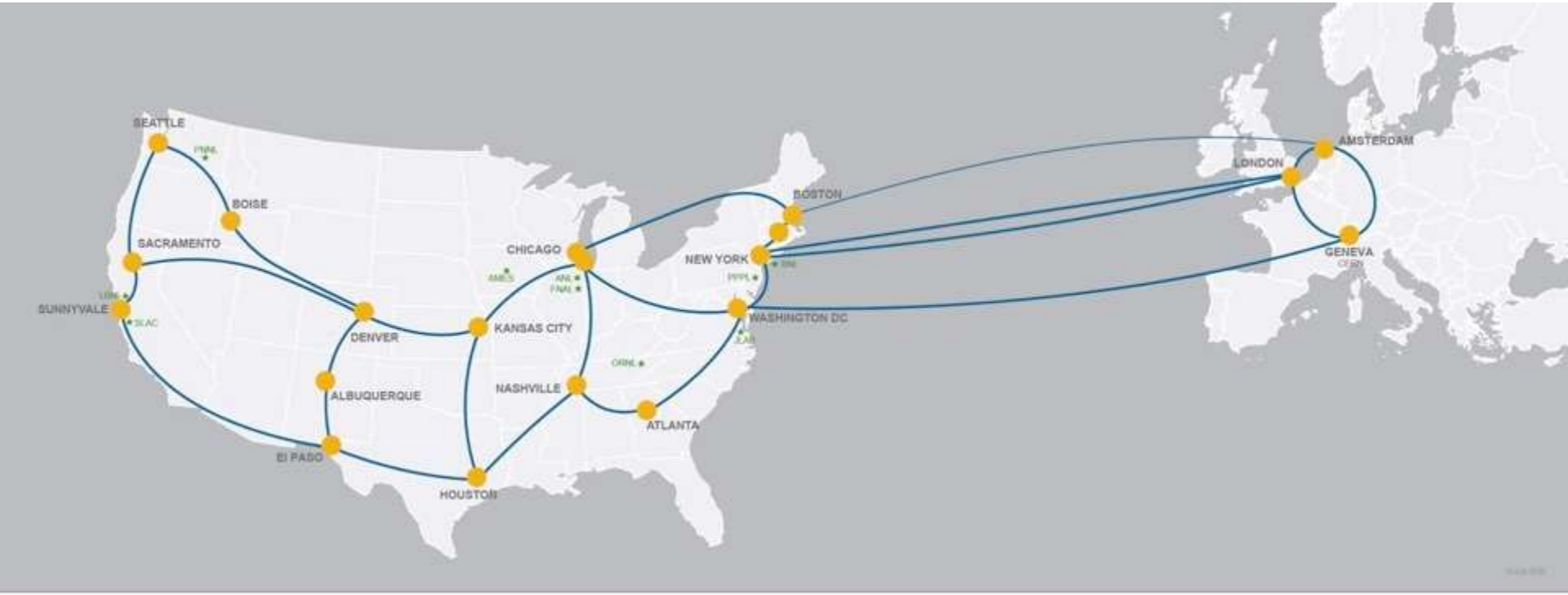
ESnet

ESnet/BER Science Partnerships: JGI

- ESnet and JGI work together both tactically and strategically
- Work with JGI staff and users on transfer performance for large data sets
- Consult and collaborate on data portal architecture and design
- Strategic engagement on topics related to new building (IGB)
- Network requirements, today and for the future

ESnet

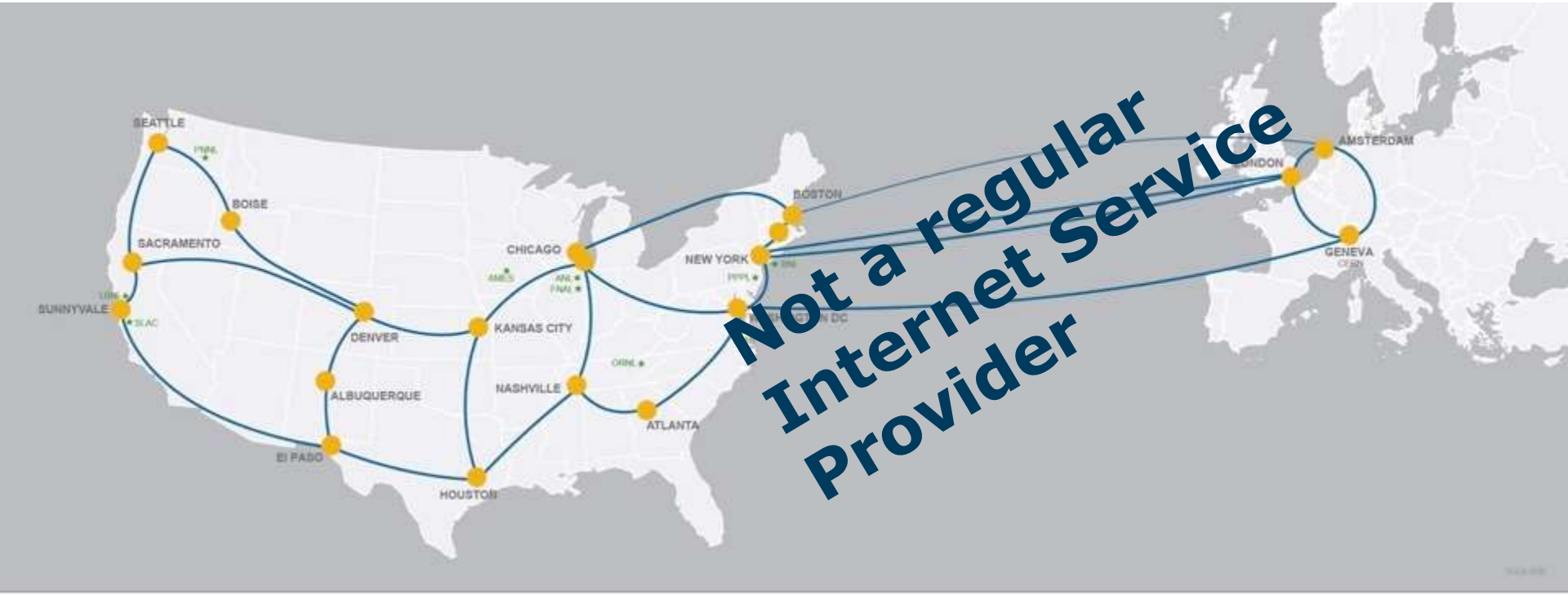
Even though ESnet builds and operates a network, it's focus is on data…



...by offering unique capabilities aka "services", and optimizing the network for

data acquisition, data placement, data sharing, data mobility

Even though ESnet builds and operates a network, it's focus is on data…



...by offering unique capabilities aka "services", and optimizing the network for

data acquisition, data placement, data sharing, data mobility

Talk

ESnet Introduction

Scaling with Design Patterns

Future Directions

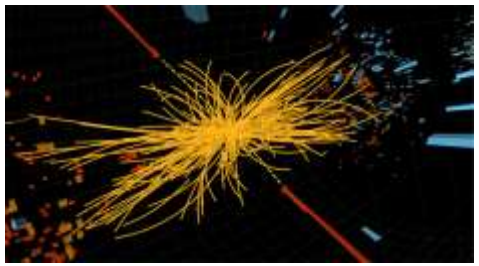
ESnet

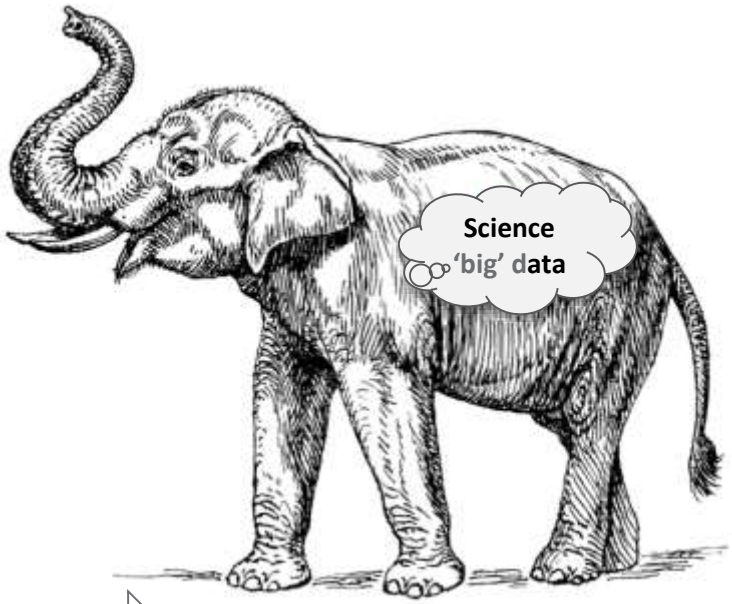
Learning from nature:
Infer and Codify the underlying design pattern

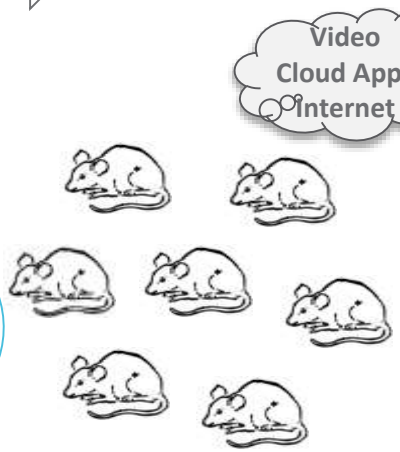
ESnet

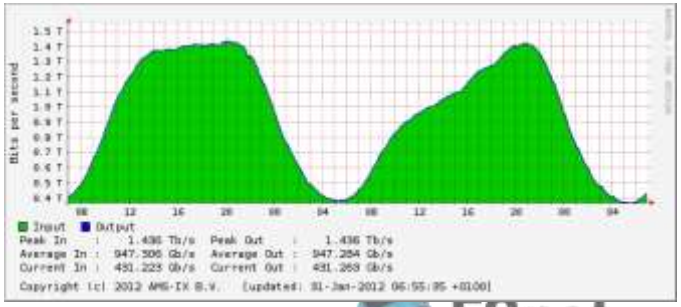
Design Pattern #1: Protect your *Elephant* Flows

HPN is built to handle science's 'big' data whose traffic patterns differ dramatically from the Internet







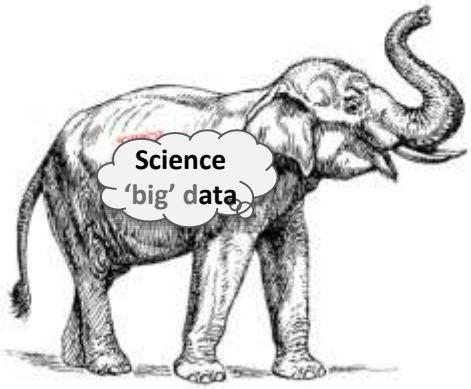


ESnet

Elephant science flow's performance suffers in case of loss in the network

Physical pipe that leaks water at rate of .0046% by volume.

Result
99.9954% of water transferred, at "line rate."



Network 'pipe' that drops packets at rate of .0046%.

Result
100% of data transferred, *slowly*, with upto 20x slowdown

essentially fixed

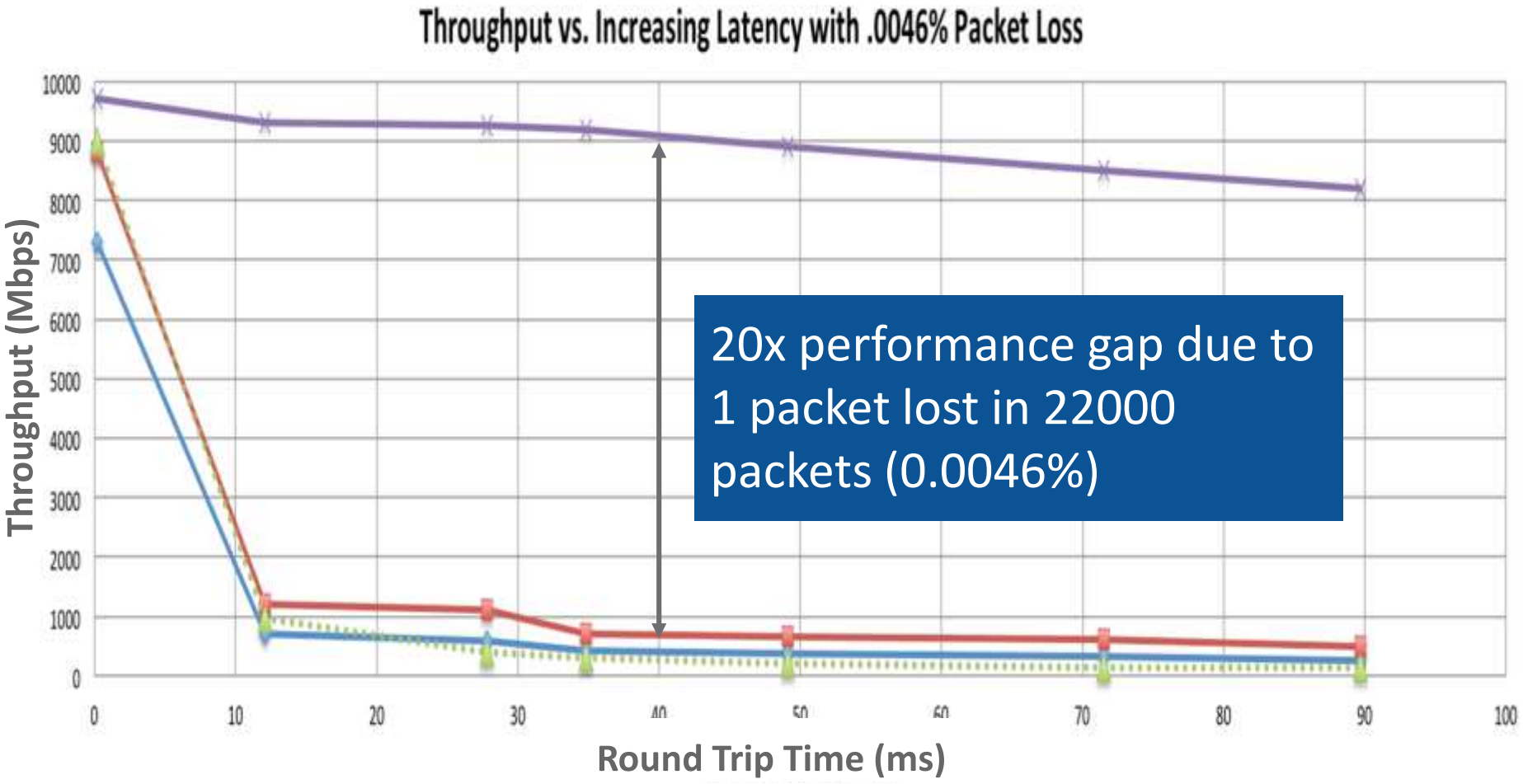
$$\frac{\text{maximum segment size}}{\text{round-trip time}} \times \frac{1}{\sqrt{\text{packet-loss rate}}}$$

determined by speed of light

Through careful engineering, we can minimize packet loss.

Assumptions: 10Gbps TCP flow, 80ms RTT.
See Eli Dart, Lauren Rotman, Brian Tierney, Mary Hester, and Jason Zurawski. The Science DMZ: A Network Design Pattern for Data-Intensive Science. In *Proceedings of the IEEE/ACM Annual SuperComputing Conference (SC13)*, Denver CO, 2013.

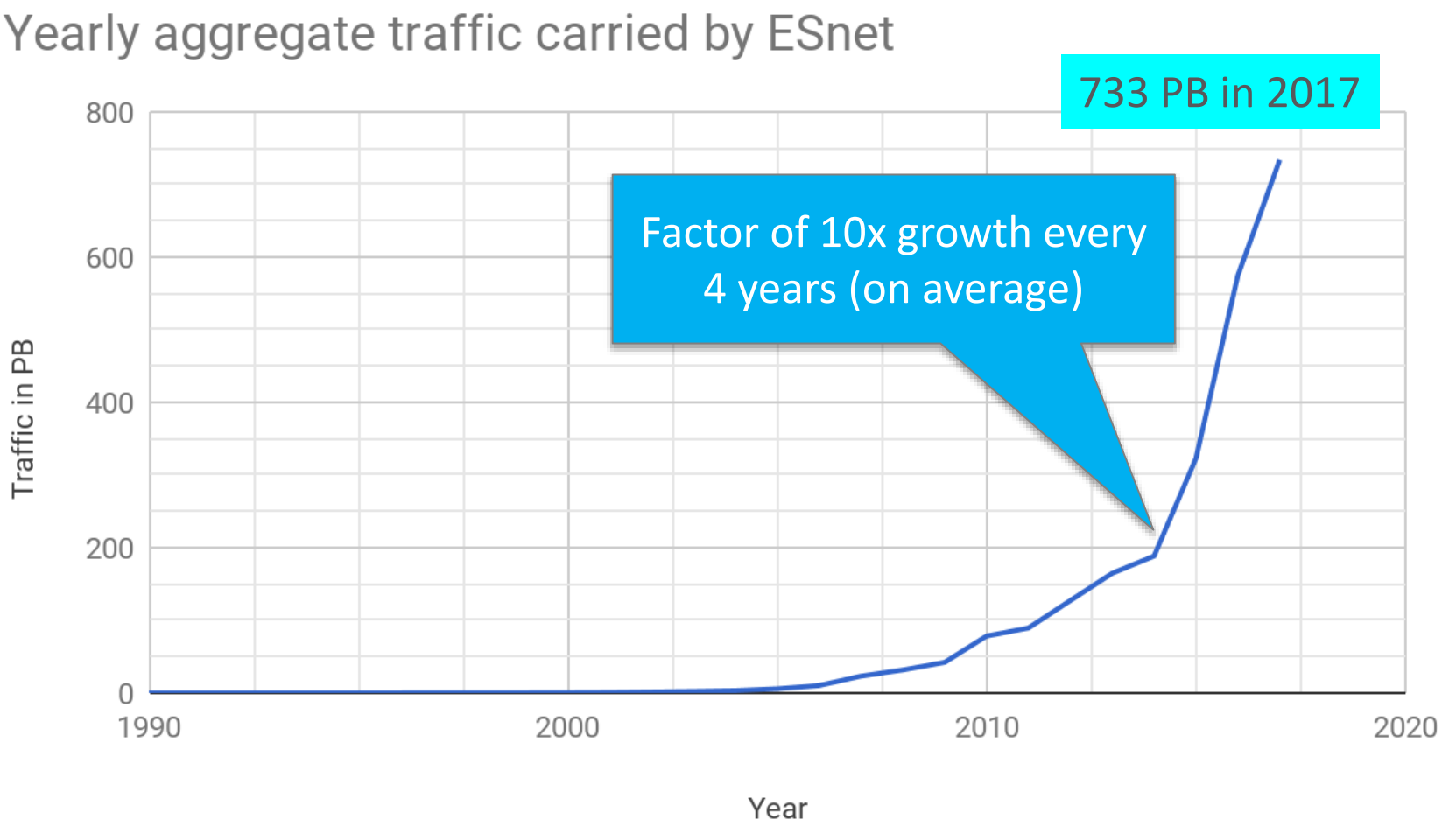
Application throughput more important than bandwidth



See Eli Dart, Lauren Rotman, Brian Tierney, Mary Hester, and Jason Zurawski. The Science DMZ: A Network Design Pattern for Data-Intensive Science. In *Proceedings of the IEEE/ACM Annual SuperComputing Conference (SC13)*, Denver CO, 2013.

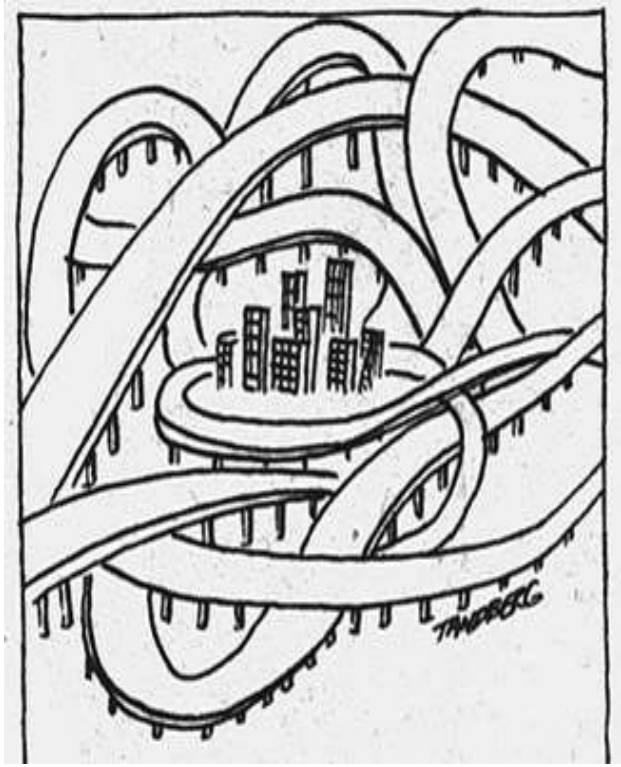
Science applications take full advantage of well engineered networks

exponential traffic growth over past 28 years



Design Pattern #2: There is no highway without the ramps

Problem and Solution explained illustratively

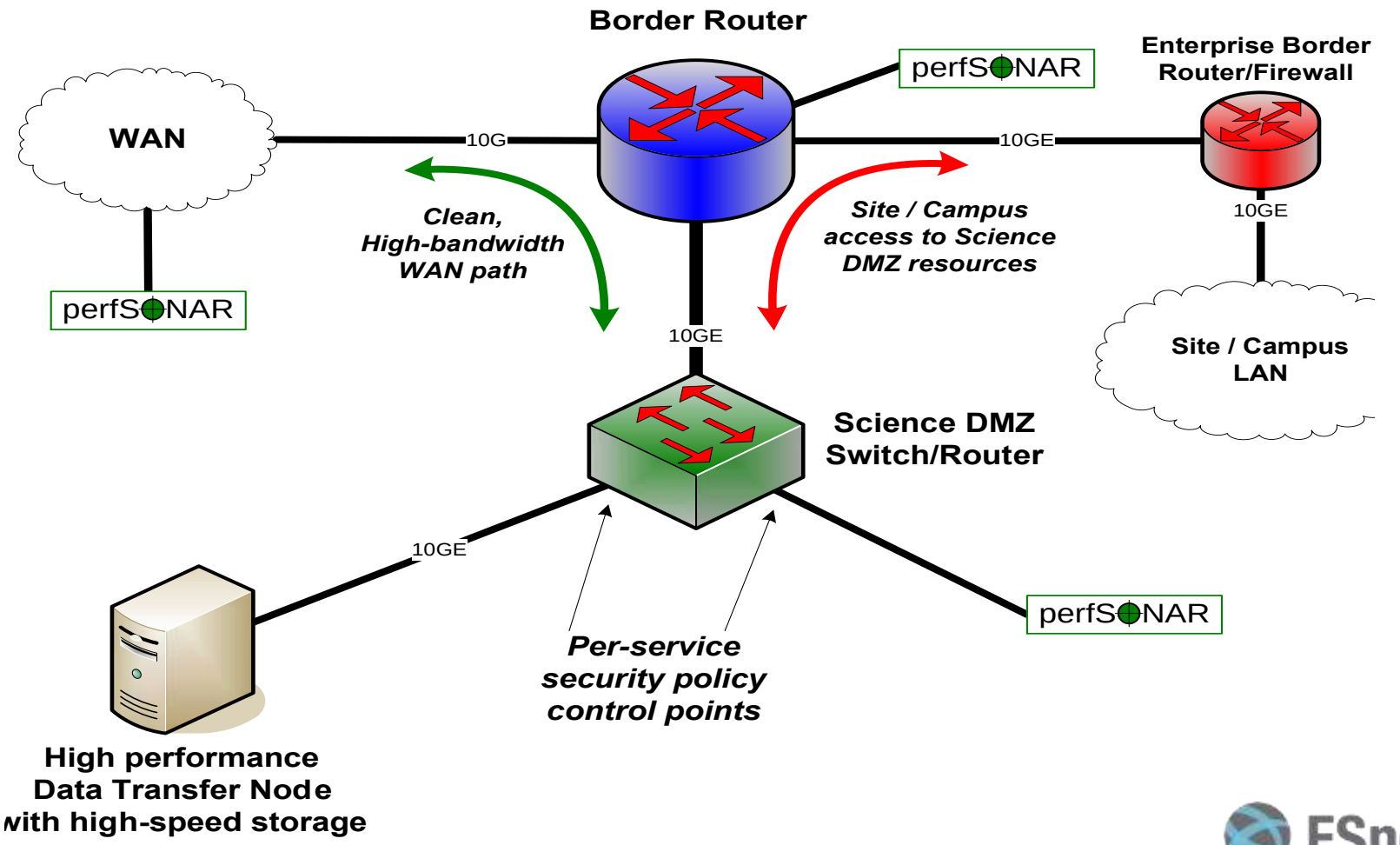


Big-Data assets **not optimized** for **high-bandwidth access** because of **convoluted campus network and security design**

Science DMZ is a **deliberate, well-designed architecture** to simplify and **effectively on-ramp 'data-intensive' science** to a capable WAN

ESnet

Science DMZ Design Pattern (Abstract)



ESnet

Emerging global consensus around Science DMZ architecture.

>120 universities in the US have deployed this ESnet architecture.

NSF has invested >>$120M to accelerate adoption.

Australian, Canadian, NZ, and other global universities following suit.

http://fasterdata.es.net/science-dmz/

ESnet

Design Pattern #3: Prepare your data cannons

ESnet

Dedicated Systems – Data Transfer Node

- Set up ***specifically*** for high-performance data movement
  - System internals (BIOS, firmware, interrupts, etc.)
  - Network stack
  - Storage (global filesystem, Fibrechannel, local RAID, etc.)
  - High performance tools
  - No extraneous software
- **Limitation of scope and function is powerful**
  - No conflicts with configuration for other tasks
  - Small application set makes cybersecurity easier

I LIKE HOW WE'VE HAD THE INTERNET FOR DECADES, YET "SENDING FILES" IS SOMETHING EARLY ADOPTERS ARE STILL FIGURING OUT HOW TO DO.

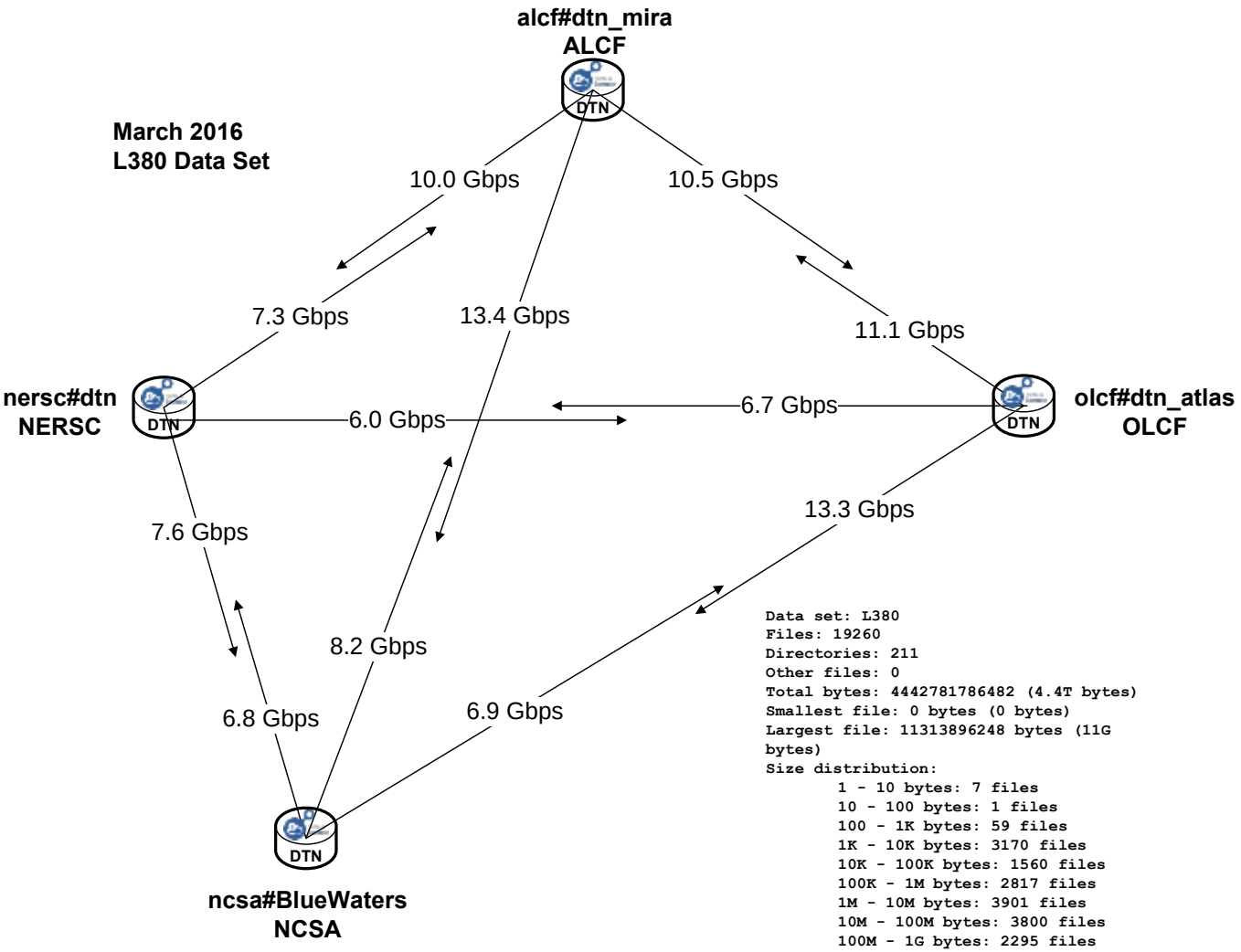
ESnet

Data And HPC: The Petascale DTN Project

- Effort to improve data transfer performance between the DOE ASCR HPC facilities at ANL, LBNL, and ORNL, and also NCSA.
  - Multiple current and future science projects need to transfer data between HPC facilities
  - Performance was slow, configurations inconsistent
  - Performance goal of 15 gigabits per second (equivalent to 1PB/week)
  - Realize performance goal for routine Globus transfers without special tuning
- Reference data set is 4.4TB of cosmology simulation data
- Benefit for all users, including climate and biology (BER)

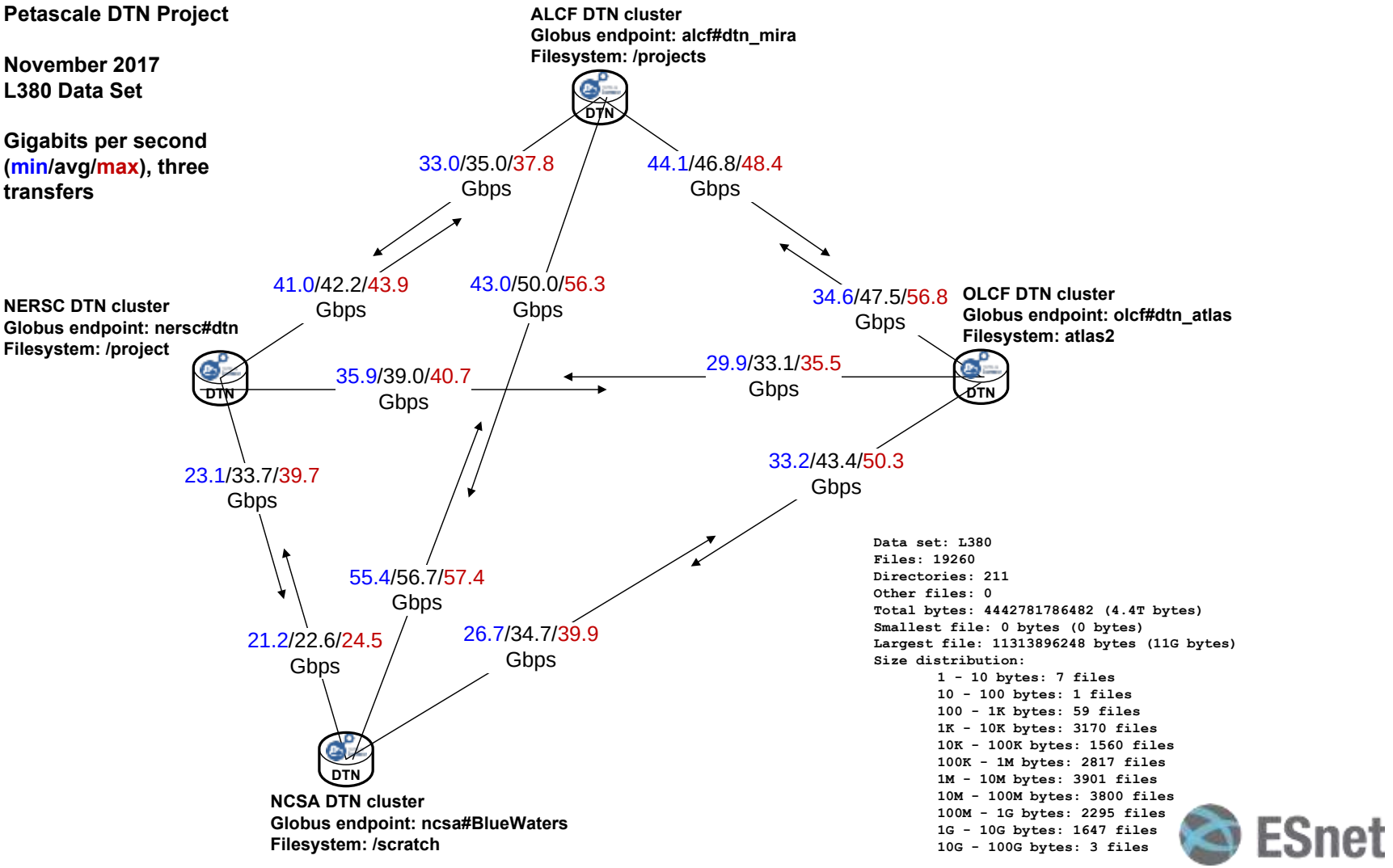
ESnet

Non-optimized DTNs – HPC Facilities (2016)




```
Data set: L380
Files: 19260
Directories: 211
Other files: 0
Total bytes: 4442781786482 (4.4T bytes)
Smallest file: 0 bytes (0 bytes)
Largest file: 11313896248 bytes (11G
bytes)
Size distribution:
        1 - 10 bytes: 7 files
        10 - 100 bytes: 1 files
        100 - 1K bytes: 59 files
        1K - 10K bytes: 3170 files
        10K - 100K bytes: 1560 files
        100K - 1M bytes: 2817 files
        1M - 10M bytes: 3901 files
        10M - 100M bytes: 3800 files
        100M - 1G bytes: 2295 files
        1G - 10G bytes: 1647 files
        10G - 100G bytes: 3 files
```

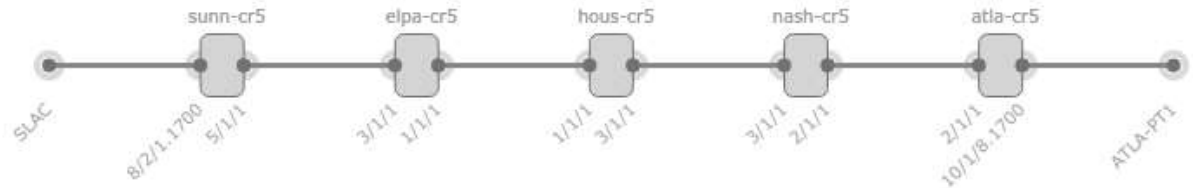
DTN Cluster Performance – HPC Facilities (2017)



From 1 PB/week to 1 PB/day (approx.)

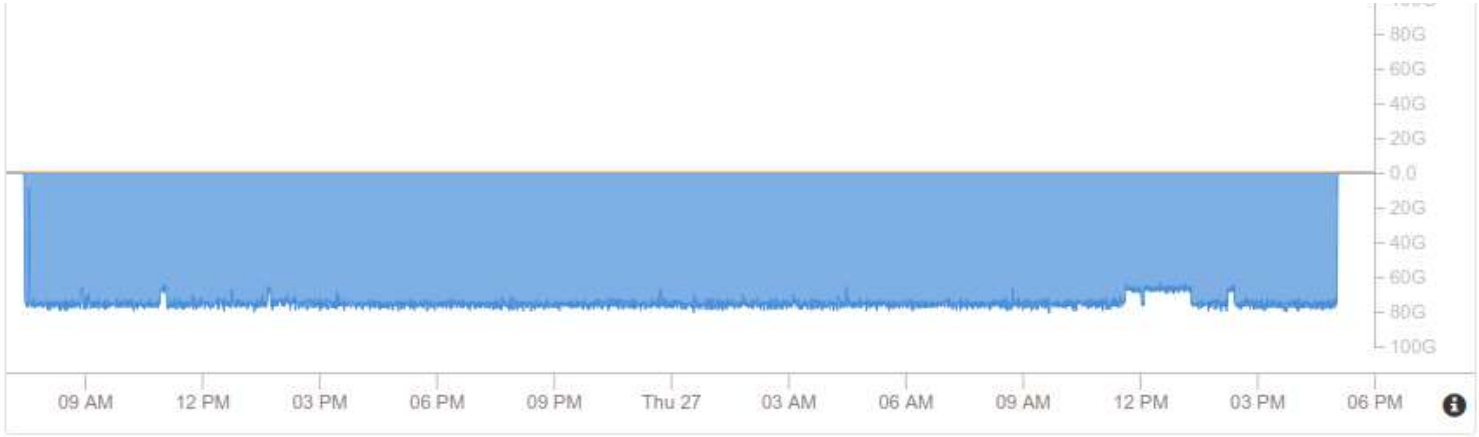
HOME › OSCARS »

SLAC latency loop - 1 of 2 - OVERRIDE - VLAN 1700



ESnet's Network, Software Help SLAC Researchers in Record-Setting Transfer of 1 Petabyte of Data

Using a 5,000-mile network loop operated by ESnet, researchers at the SLAC National Accelerator Laboratory (SLAC) and Zettar Inc. (Zettar) recently transferred 1 petabyte in 29 hours, with encryption and checksumming, beating last year's record by 5 hours, almost a 15 percent improvement.



ESnet

Talk

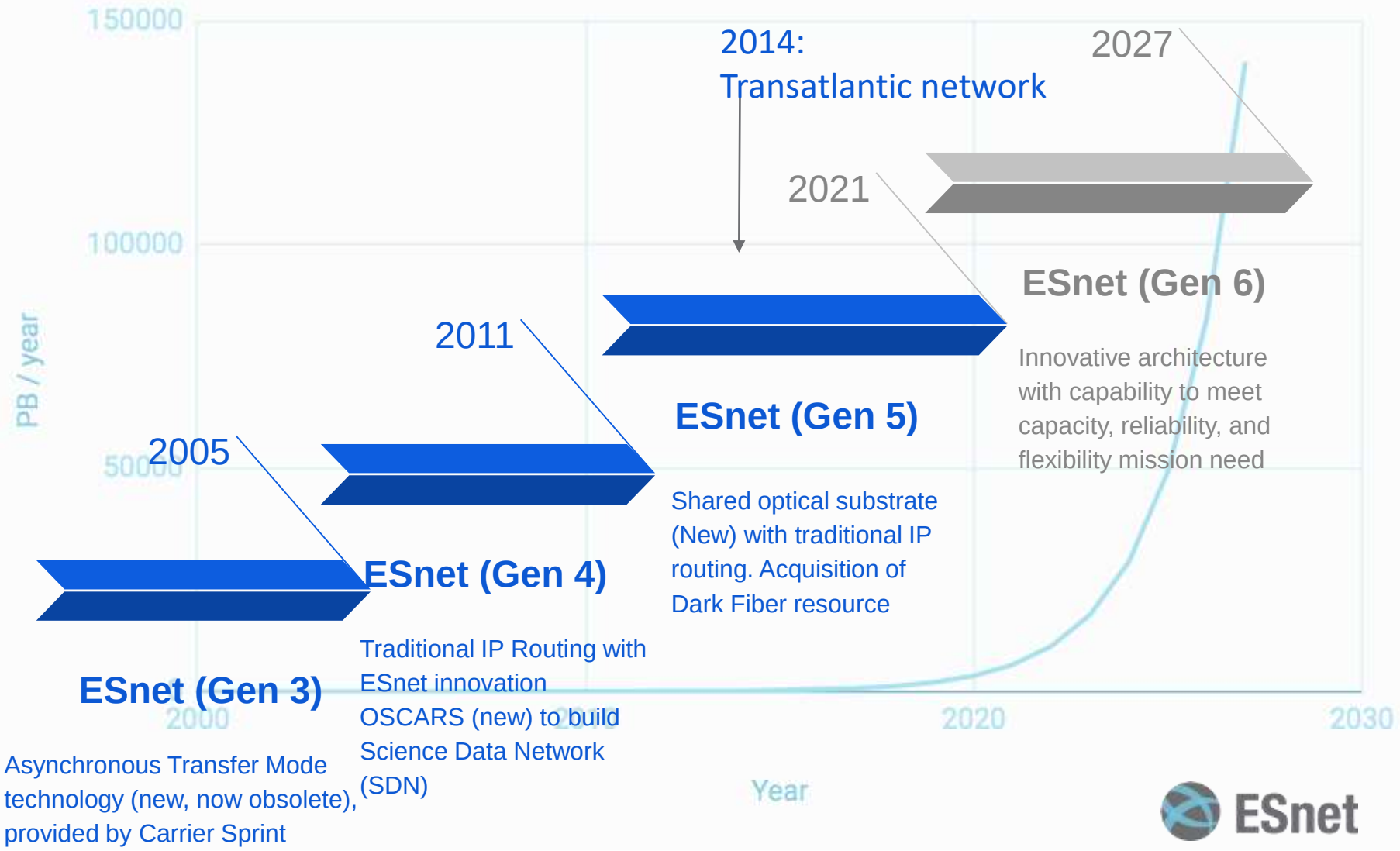
ESnet Introduction

Scaling with Design Patterns

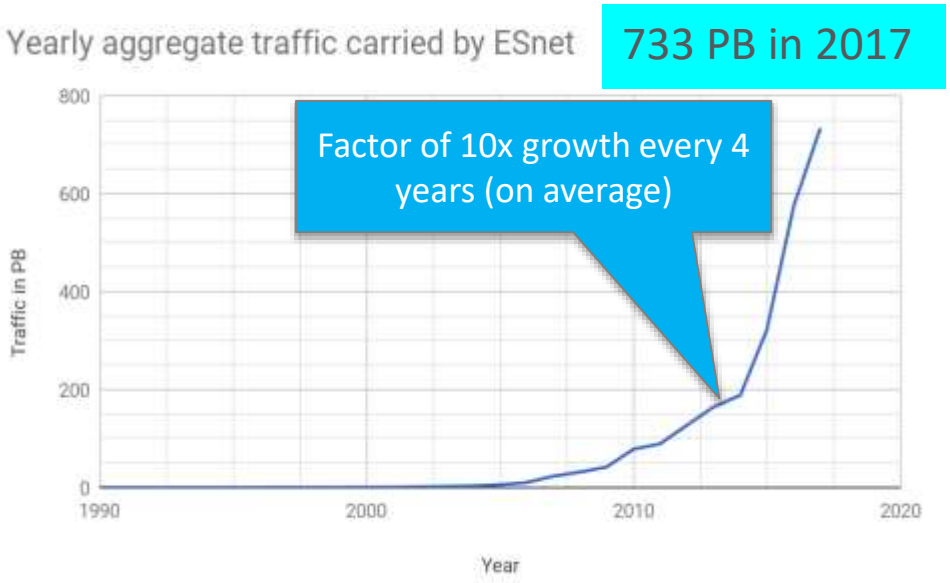
Future Directions

ESnet

Each major upgrade transforms the facility with innovative, cutting edge technologies



ESnet Upgrade: ESnet6 Mission Need



1. Capacity to handle exponential increase in science data

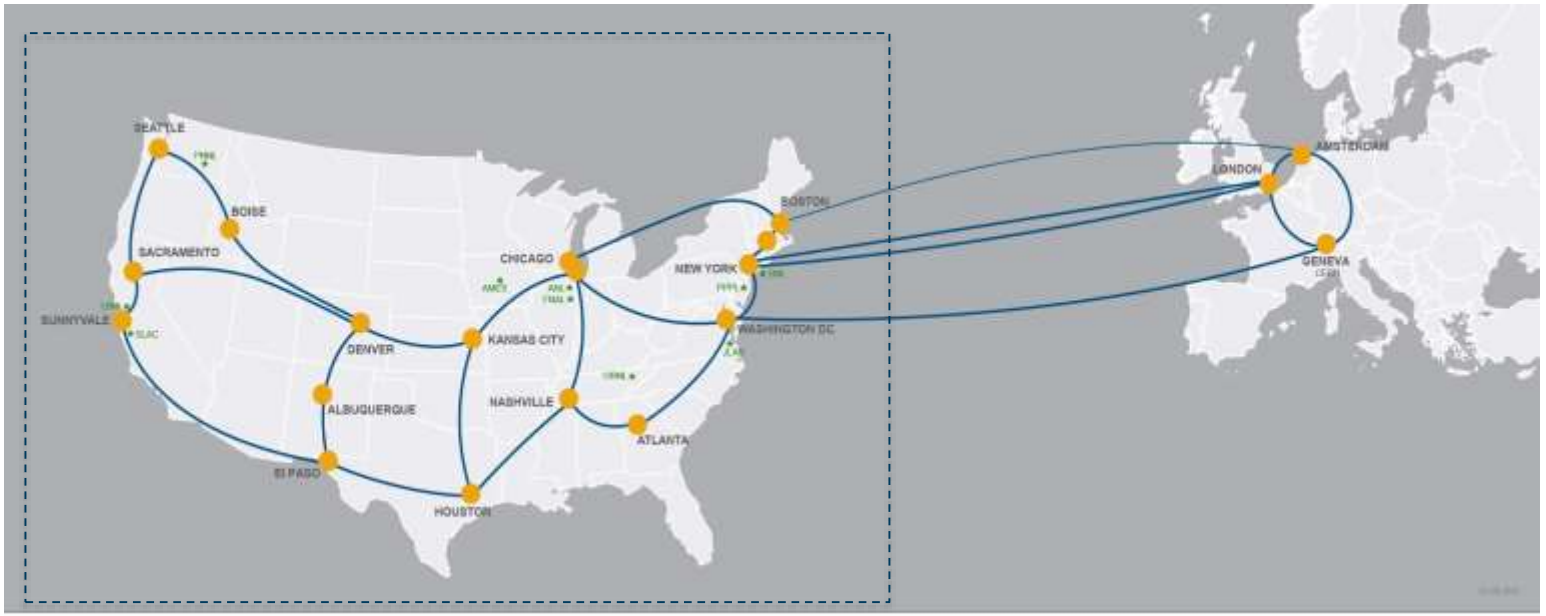
2. Replace end-of-life equipment with an architecture that inherently provides reliability and cyber-resiliency.

3. Flexibility to create network services to meet new scientific opportunities.

CD 1/3A in August 2018, CD 2 planned mid-late next year

ESnet

Novel programmable network architecture on nationwide unlit fiber**



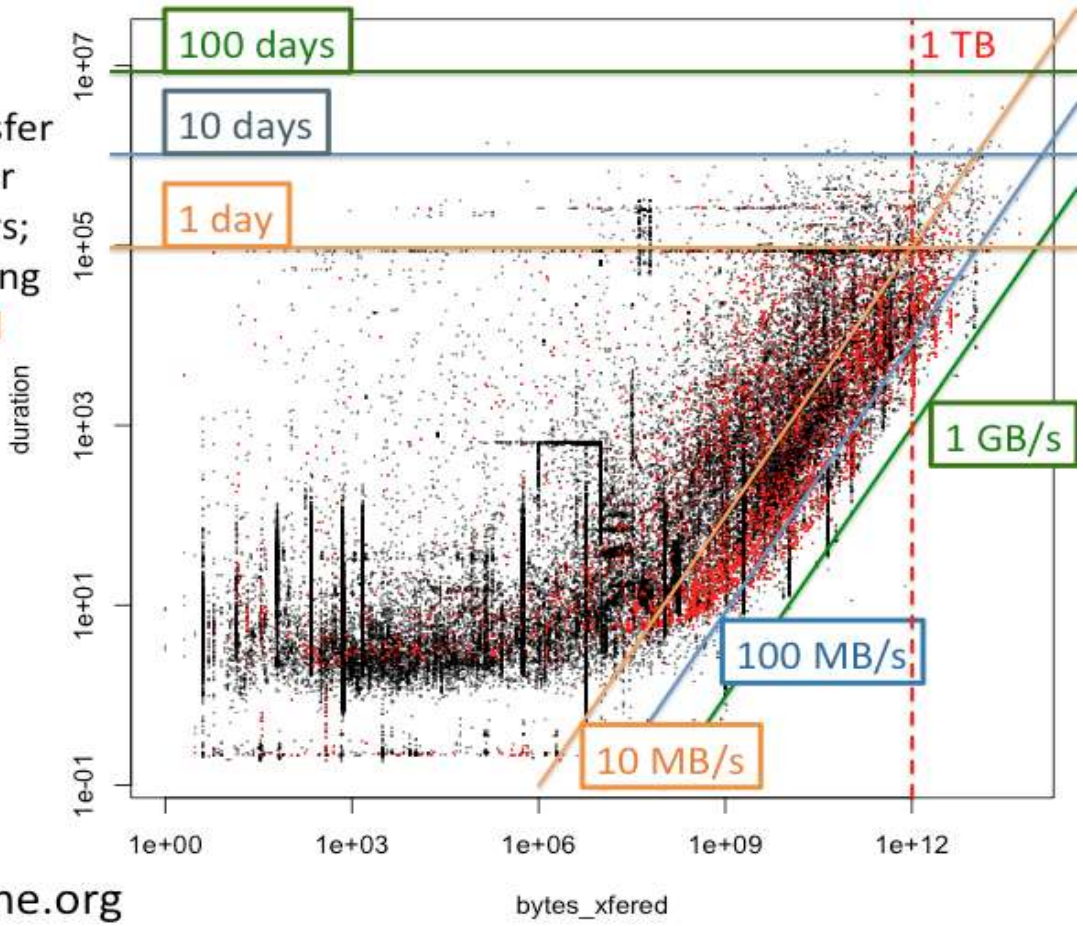
- Architecture is based on a scalable ‘switching core’ coupled with a flexible and dynamic ‘intelligent services edge’
- Integration of compute, storage and network
  - **Aligned with BER’s Data Grand Challenge**
- Automation and programmability of network services planned as key features
- Early finish planned for Q1 FY2023

** transatlantic links were recently renewed

Research Challenge: Predictable Network Transfers at scale

- Transfers over a shared network are not predictable
- Best-effort delivery can also mean worst-effort delivery

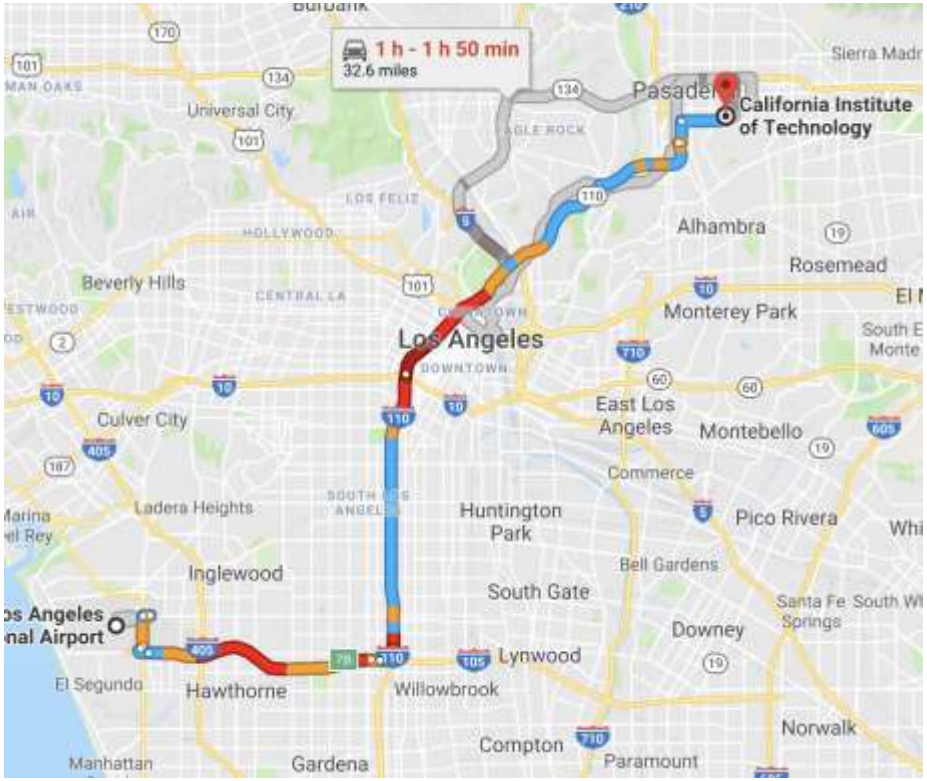
Globus Transfer requests over last two years; those involving NERSC in **red**

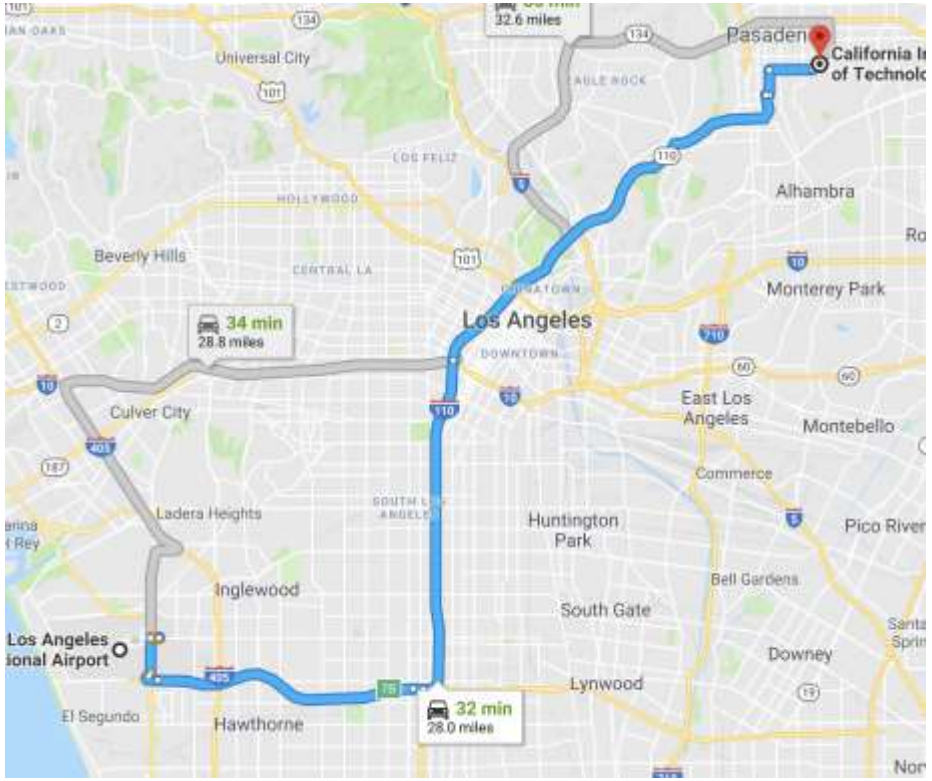


globusonline.org

ESnet

When will I get home?





LAX– Caltech, 6 pm:
1 hr – 1hr 50 min

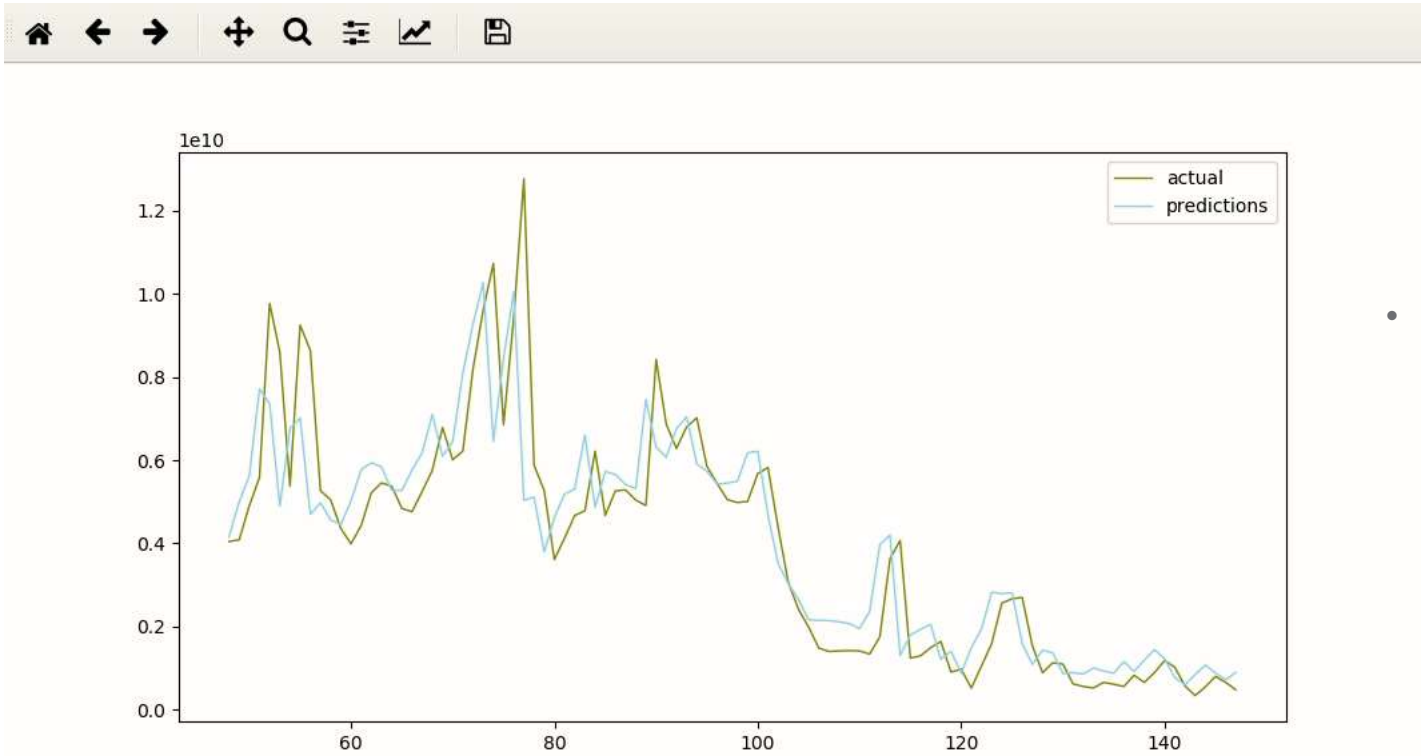
LAX– Caltech, 11 pm:
32 min

ESnet

Machine Learning applied to network telemetry data – learn, understand and optimize

Predicting traffic per link/site
Method: Deep Machine Learning (Recurrent Neural Network) for time-series data

- Predict anomalies (or peaks) accurately 15 minutes in future



- Graph showing 1 minutes prediction in future

ESnet

Research Challenge:
Applications cannot ‘dialogue’ with the network

Application Workflow

Service is not working, please check status

available 10Gbps is ok.

Endpoint Listing

Heavy traffic in your area

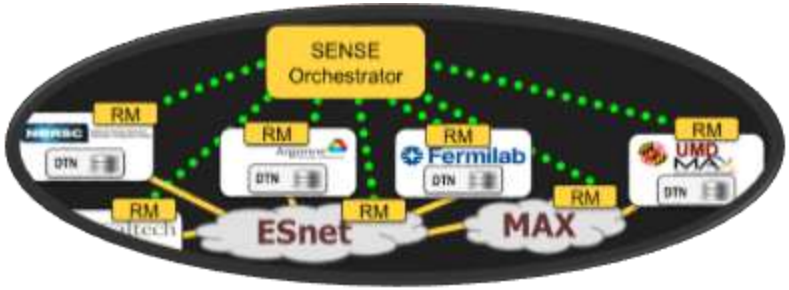
4:37 PM

Slower than usual with delays up to 8 min

service between Caltech and Fermilab. If 20 Gbps not available 10Gbps is ok.

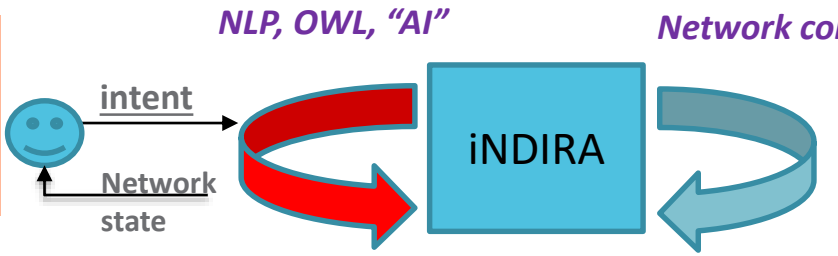
15 Gbps P2P service between Caltech and Fermilab Instantiated

Failure on a network element, problem fixed



Machine Learning applied to dialogue with applications and workflows – understand the intent

" I want to send data to my SuperComputer at NERSC by 5:00pm today"



" Ok ill reconfigure the network to make this possible!"

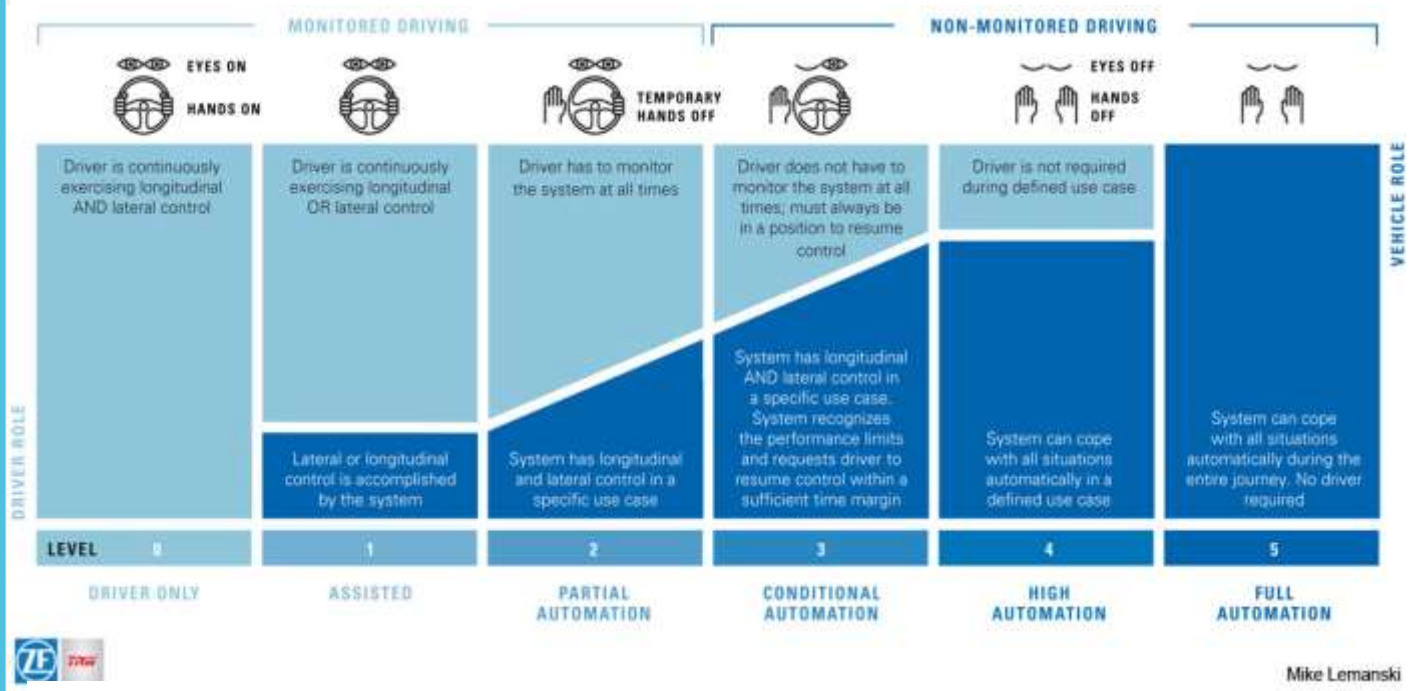
Language processing to take intent input

Renderer translates intent

- Understand English (e.g. transfer, connect)
- Check conditions, conflicts and permissions
- ML in Natural Language Processing for intelligent negotiation with user

- Automate rendering into network commands like bandwidth, time schedule, topology
- Optimize the network
- Return success or failure to user

Vision (or a challenge): A cognitive network



Network can:

Assimilates internal and external information (e.g. usage, maintenance schedules, component MTF, driver's schedule, etc.)

Anticipates trips based on routines and disruptions (e.g. scheduled maintenance)

Adapts route and departure time due to road, (current and expected) traffic, and weather conditions

ESnet

Networks are **the** circulatory system for digital data

1. ESnet facility is **engineered and optimized** to meet the diverse needs of DOE Science
2. We aim to create a world in which **discovery is unconstrained by geography.**
3. An effective dialogue between the **network and application** is extremely important to accomplish the end-to-end vision

ESnet

Thank You and Questions?

imonga@es.net

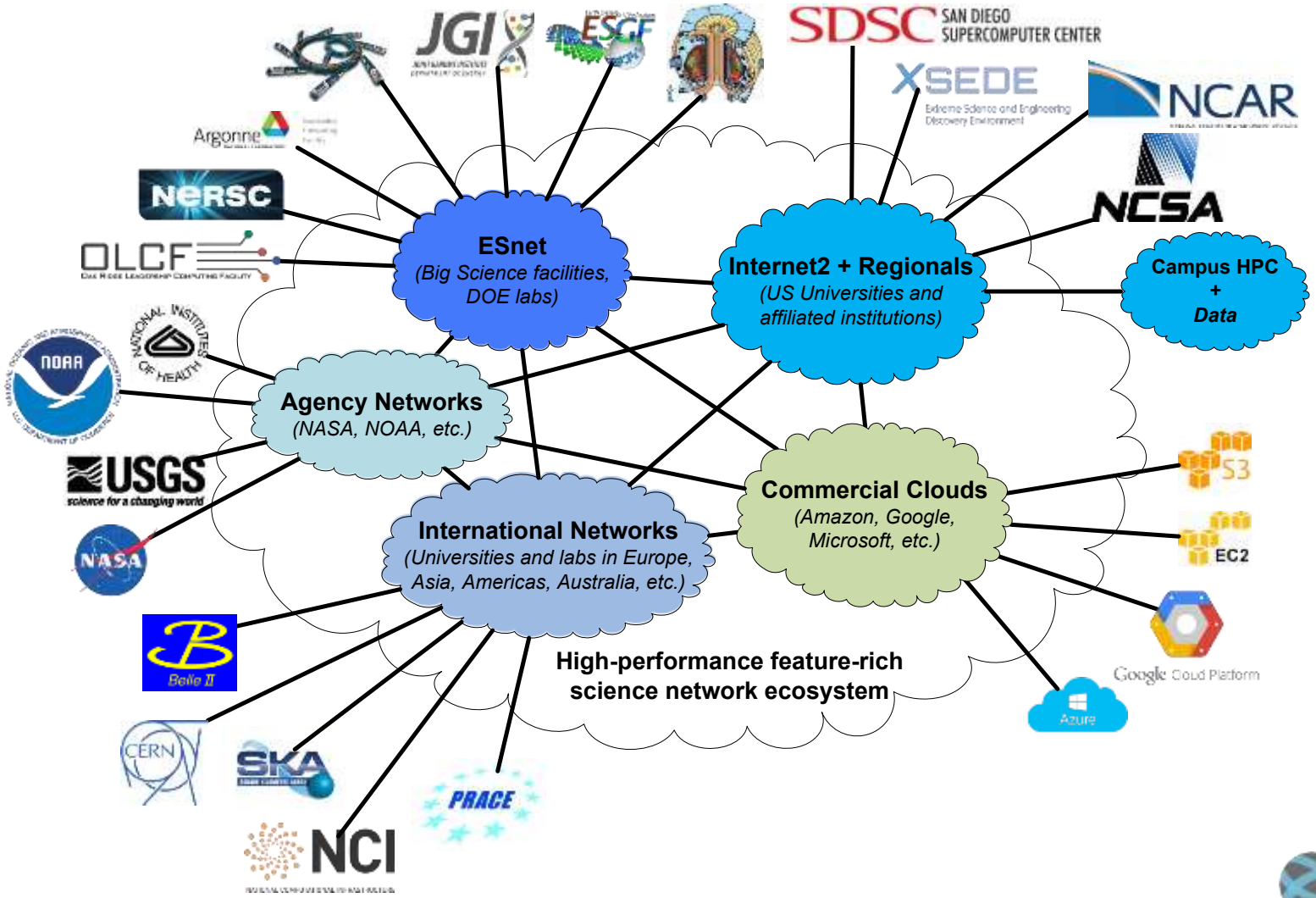
ESnet

In conclusion – ESnet's vision:

Scientific progress will be **completely unconstrained** by the physical location of instruments, people, computational resources, or data.

ESnet

Long-Term Vision For Facilities



ESnet

A reputation for innovation and excellence.

CENIC

InformationWeek
Government

OPEN NETWORKING
FOUNDATION

"The entire staff conscientiously and continually lead their field."

[report from recent operational review]

ESnet

ESnet is a 31-Year Old Mission Organization

Mission of DOE Office of Science:
Deliver knowledge and tools for transforming our understanding of the universe.

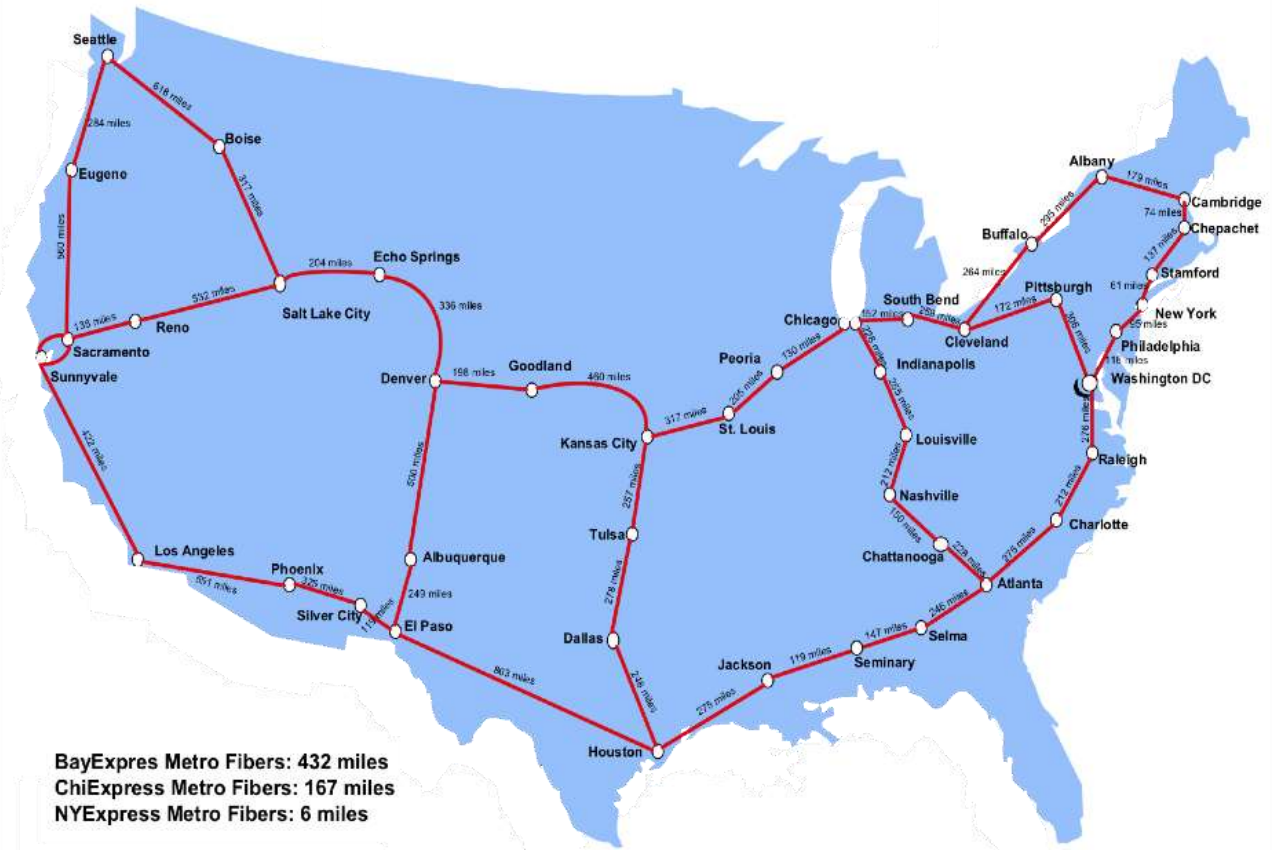
Mission of Energy Sciences Network:
Accelerate this research and discovery.

$5B/year for the US National Lab Complex, which includes:

- world's largest collection of scientific user facilities
- supercomputers, accelerators, xray / neutron sources, electron microscopes, sequencers, fusion facilities, **Energy Sciences Network**
- >100 Nobel Prizes

ESnet

Leverage key asset – 13,000 miles of Dark Fiber IRU



ESnet

In conclusion – ESnet's vision:

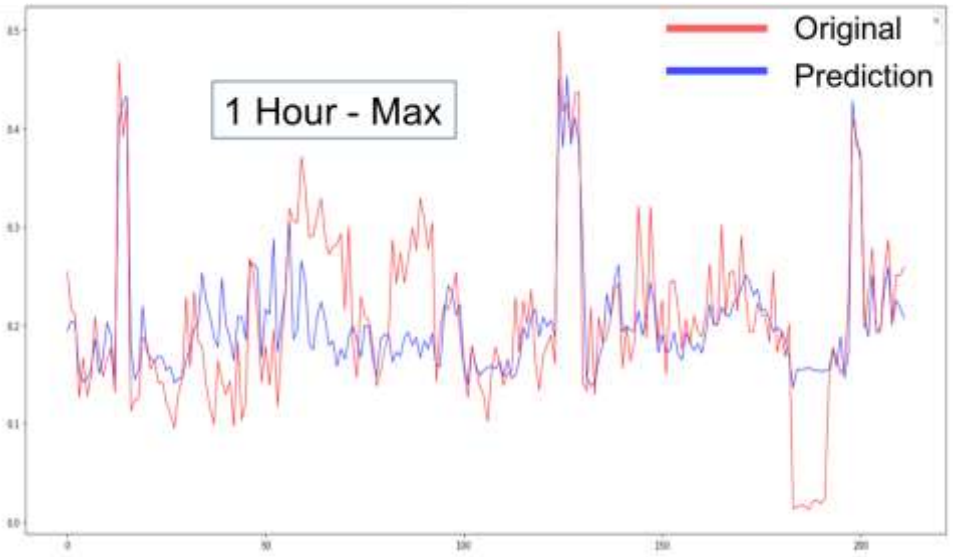
Scientific progress will be **completely unconstrained** by the physical location of instruments, people, computational resources, or data.

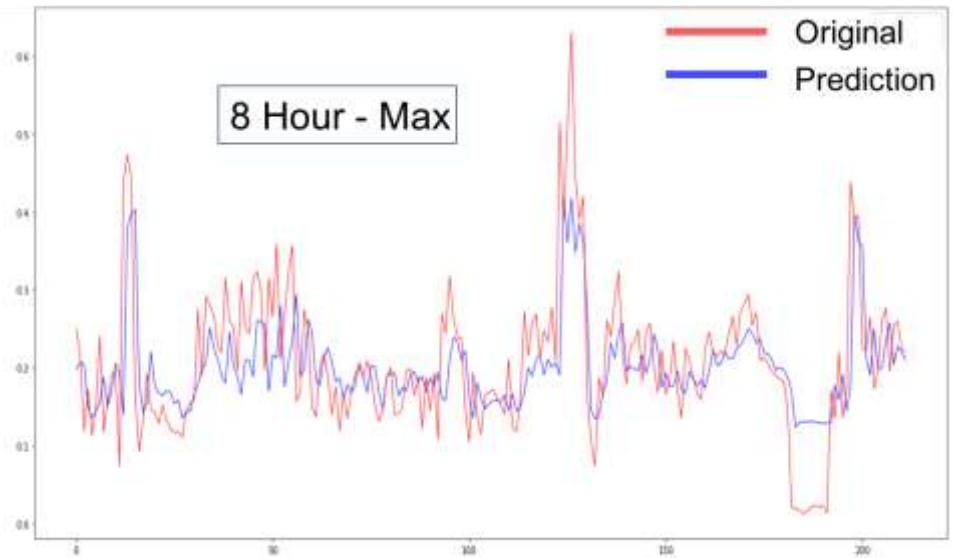
ESnet

Predicts 8 hours!

- Using current 8 hours on pretrained model
- Follows trend accurately
- Predicts magnitude fairly well
- Predicts high anomalies
- Mean Square Error (MSE) of our method performs better the traditional approaches:

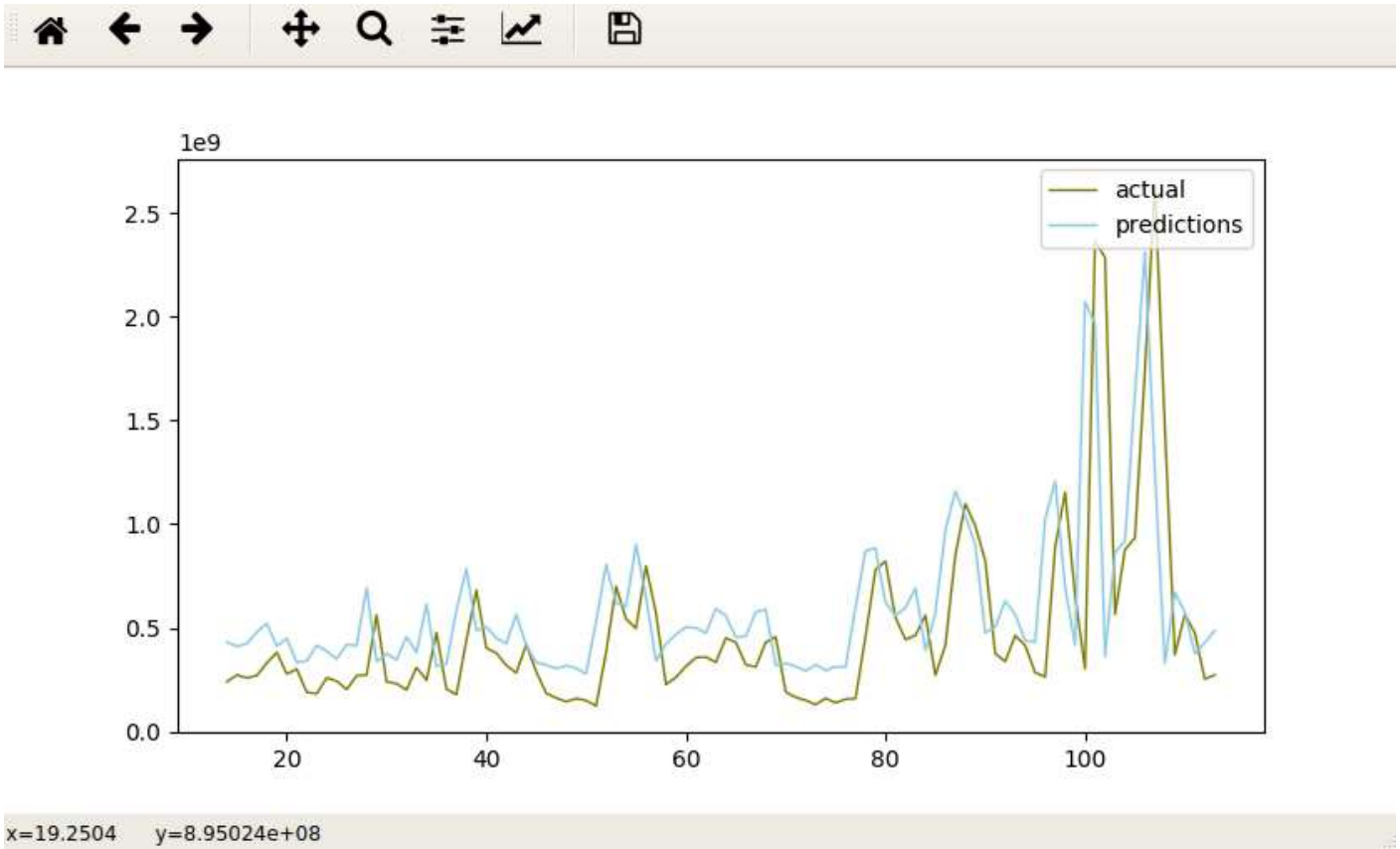
Link	Our Model	ARIMA	Holt Winters
WASH-CR5	**0.00413**	0.01198	0.02267
ESNET-LSW1	**0.00377**	0.05601	0.06923





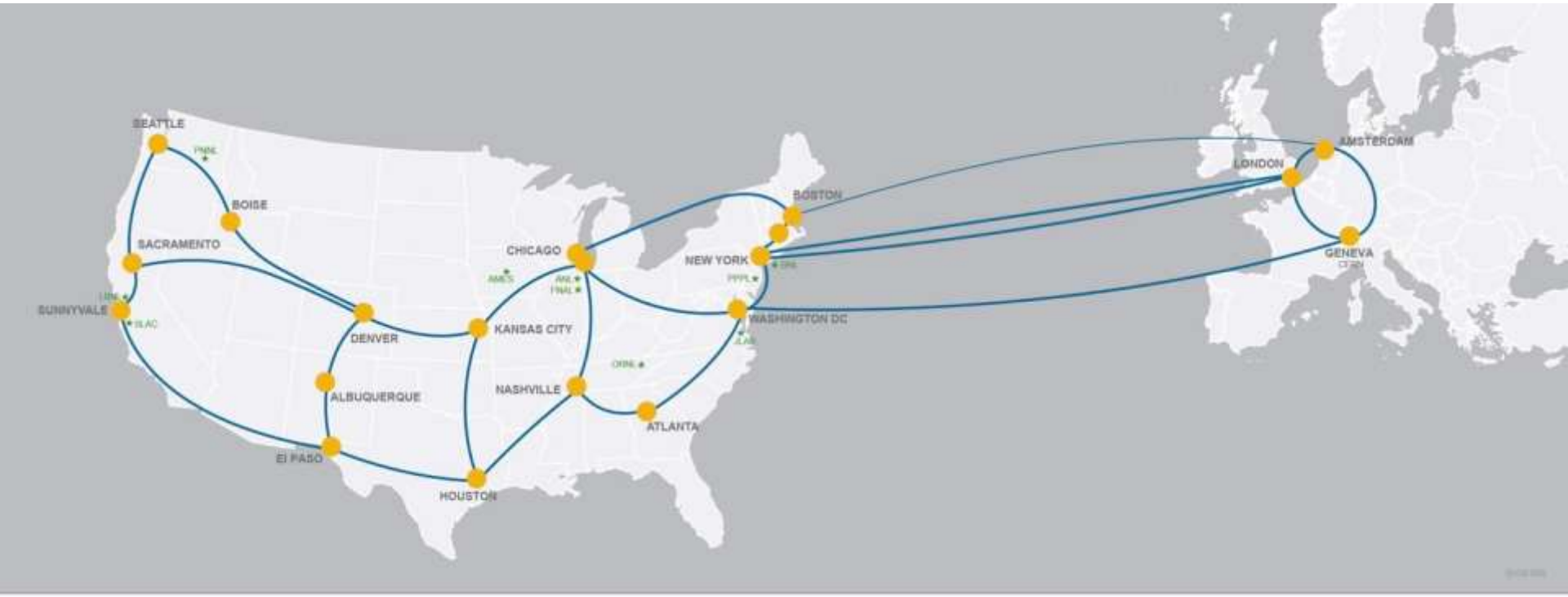
ESnet

Real-time plotting (showing just one step ahead)



ESnet

ESnet: DOE's international science network user facility – an instrument to accelerate discovery



Office of Science Facility connecting all of the DOE labs, experiment sites, & supercomputers

Interconnects to 100's of other science networks around the world and to the Internet