ESnet

ENERGY SCIENCES NETWORK

# Building race-tracks for big-data science

Inder Monga

Executive Director, Energy Sciences Network

Division Director, Scientific Networking

Lawrence Berkeley National Lab

MAX Participants Meeting

April 11th, 2019

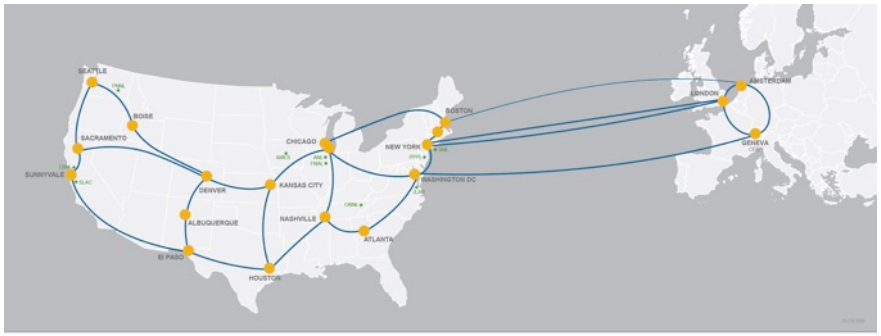
U.S. DEPARTMENT OF ENERGY

Office of Science

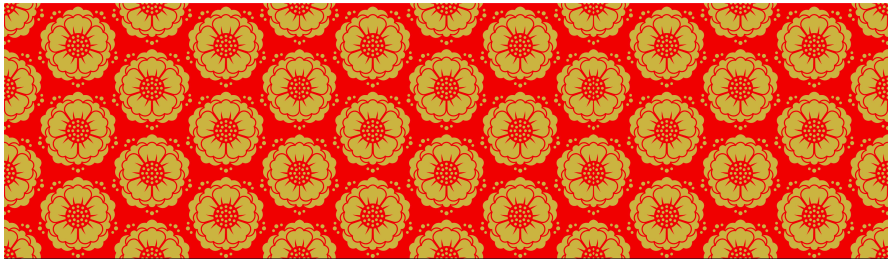
BERKELEY LAB

# Talk

ESnet Introduction




Established Design Patterns



Emerging Design Patterns

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


Provides connectivity to
all of the DOE labs, experiment sites, & user facilities (> 34417 users)

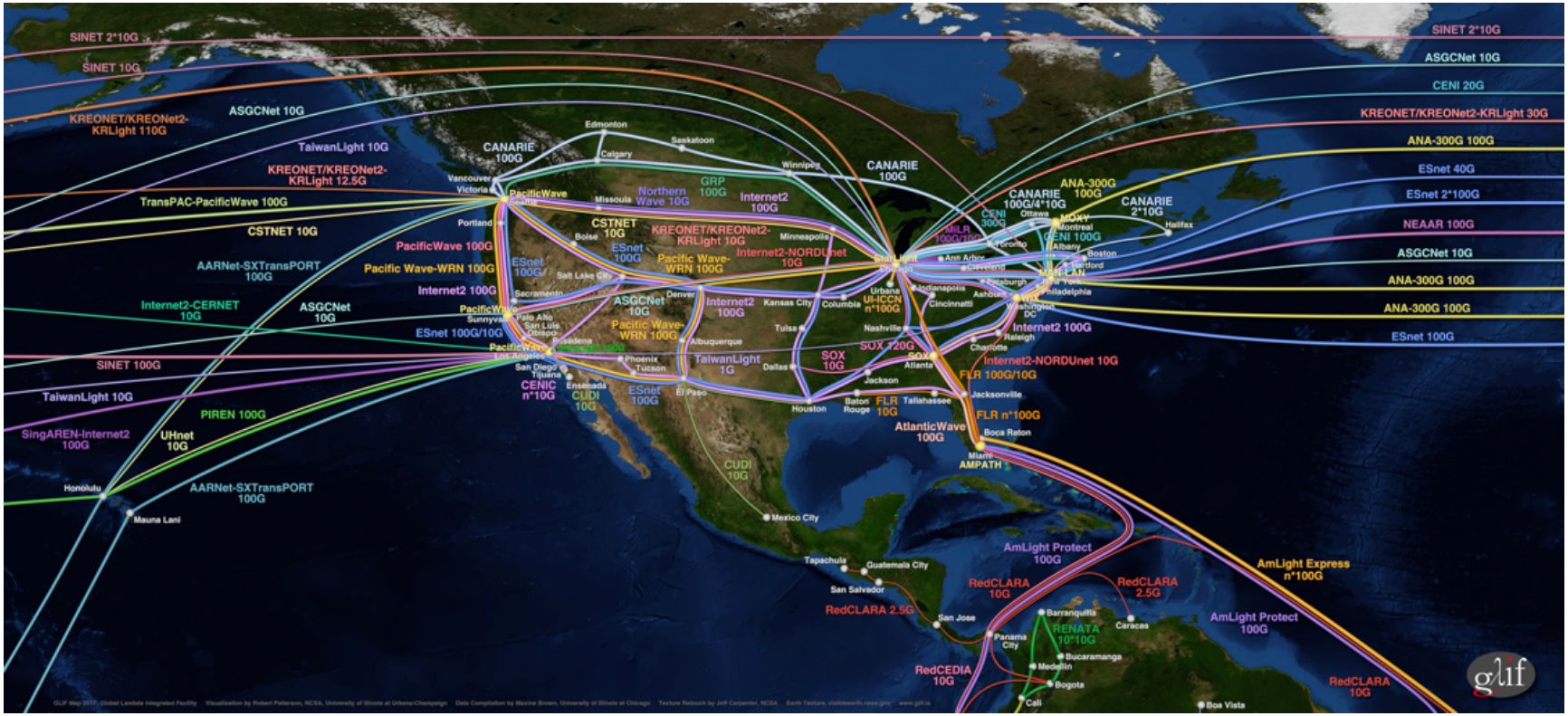
ESnet

# 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

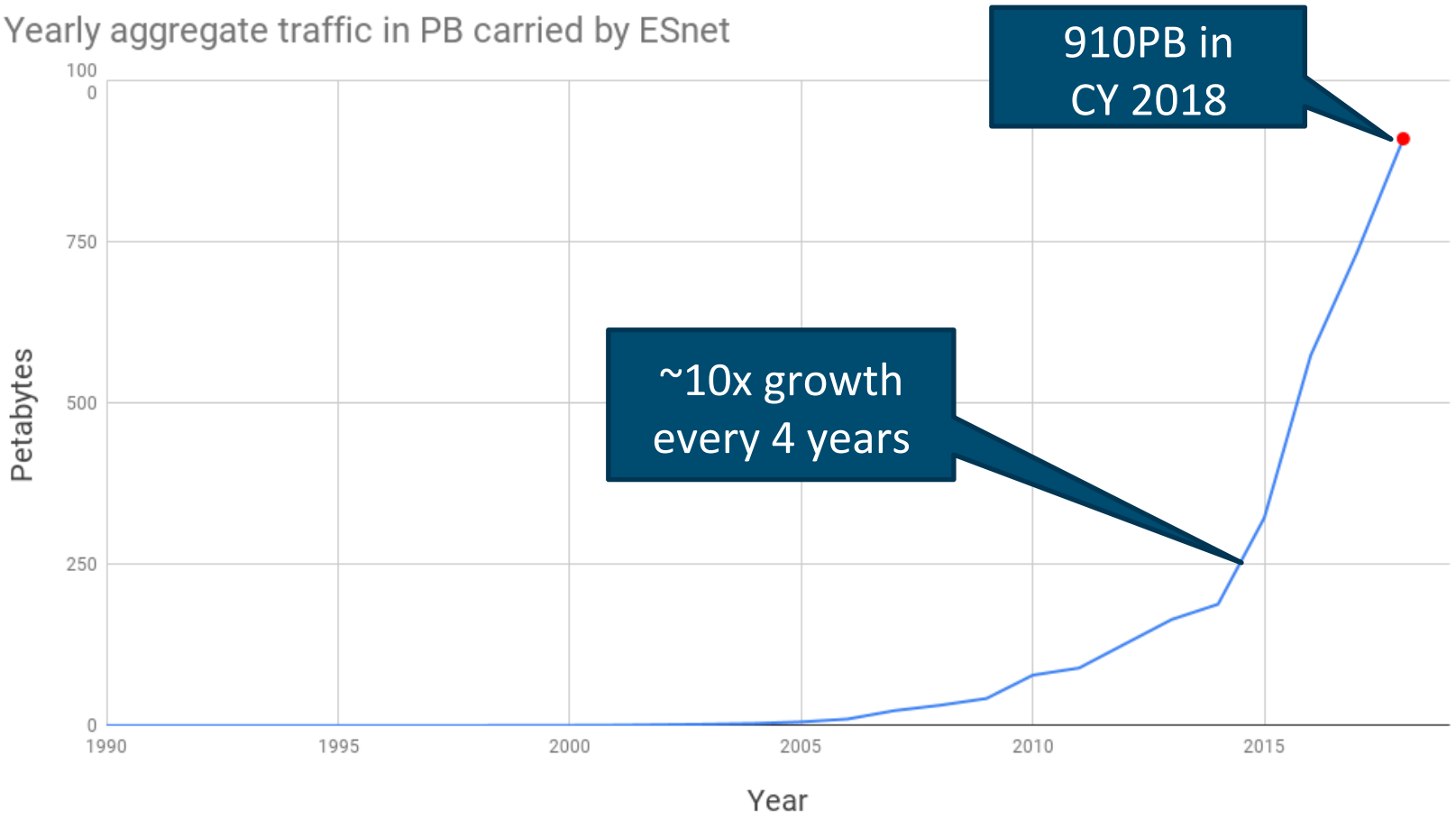



**80%** of carried traffic originates or terminates outside the DOE complex

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

ESnet

# An ~exabyte network today



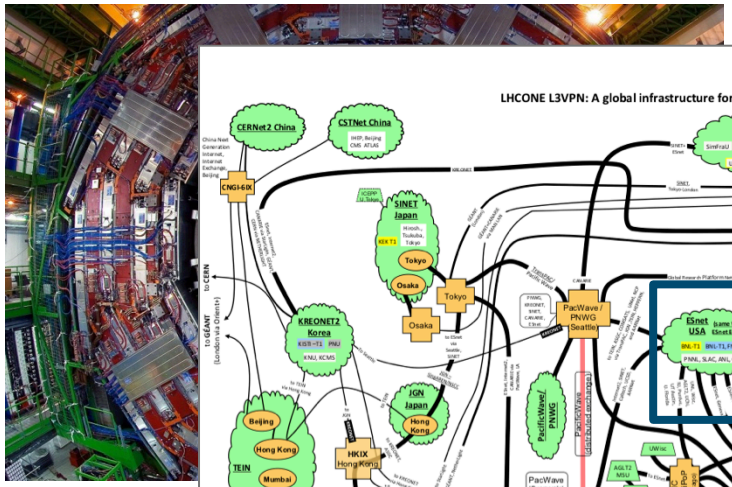

exponential traffic growth over past 28 years

measures ingress or egress only, not traffic per link

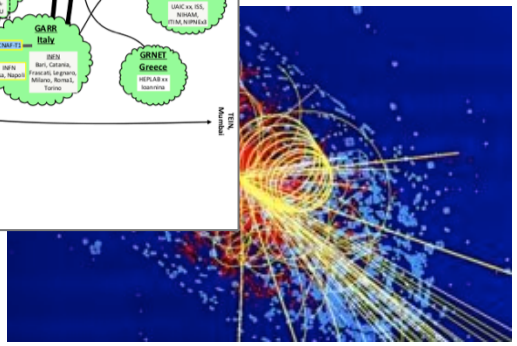
ESnet

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

***Example 1: High Energy Physics / Large Hadron Collider Science***



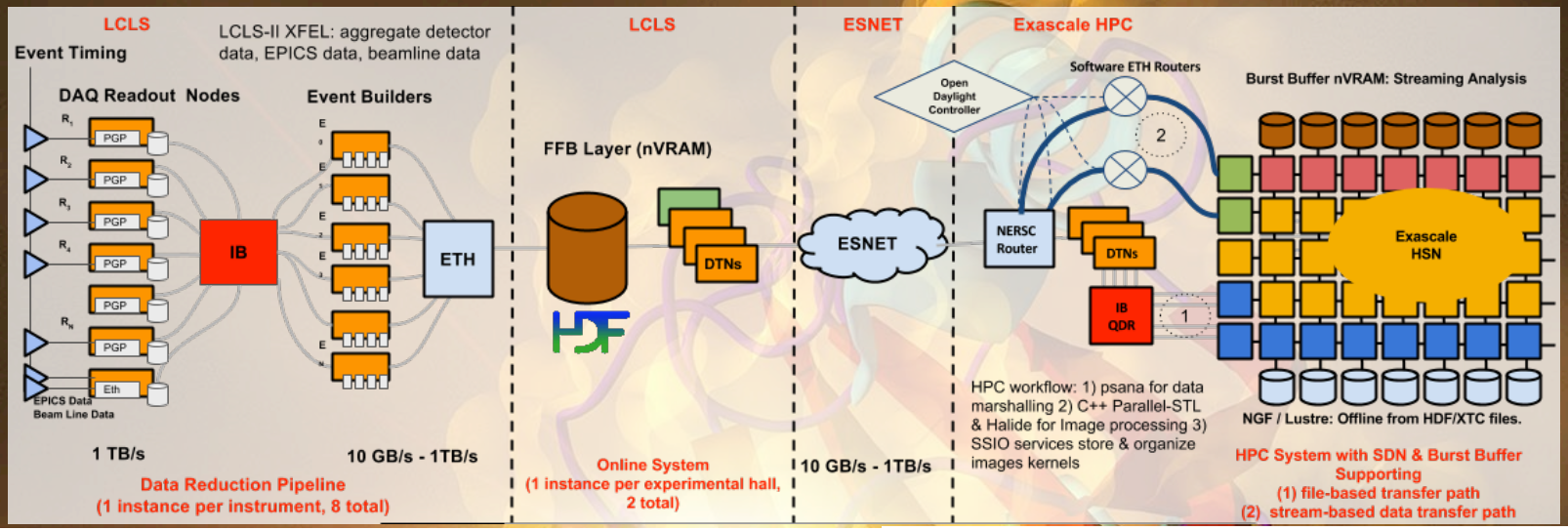


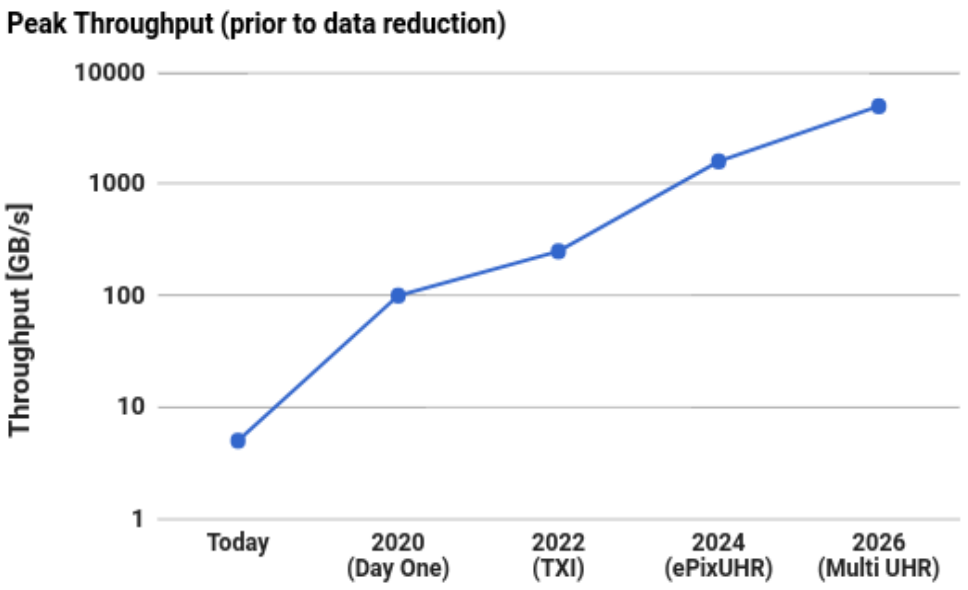



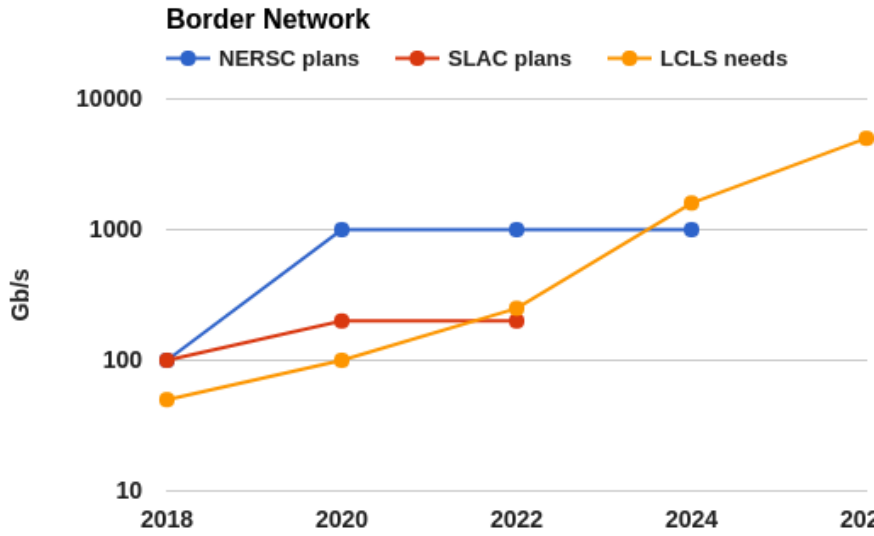
ESnet supports high-performance data movement to/from LHC in CERN, Switzerland to FNAL and BNL (Tier 1 sites) and 20 other universities

*Discovery of*

# LCLS Science Data (2020 – 2026+)

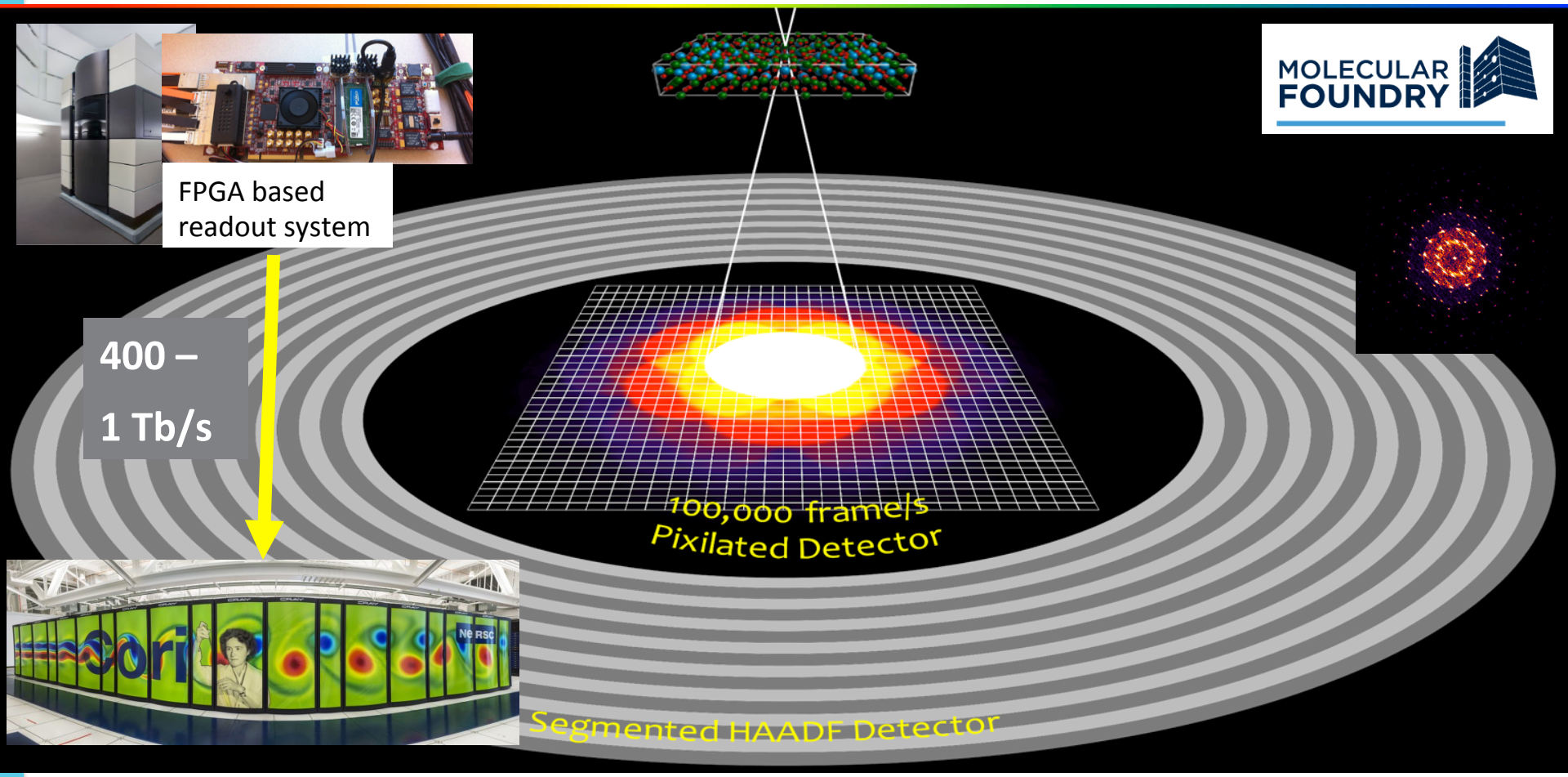


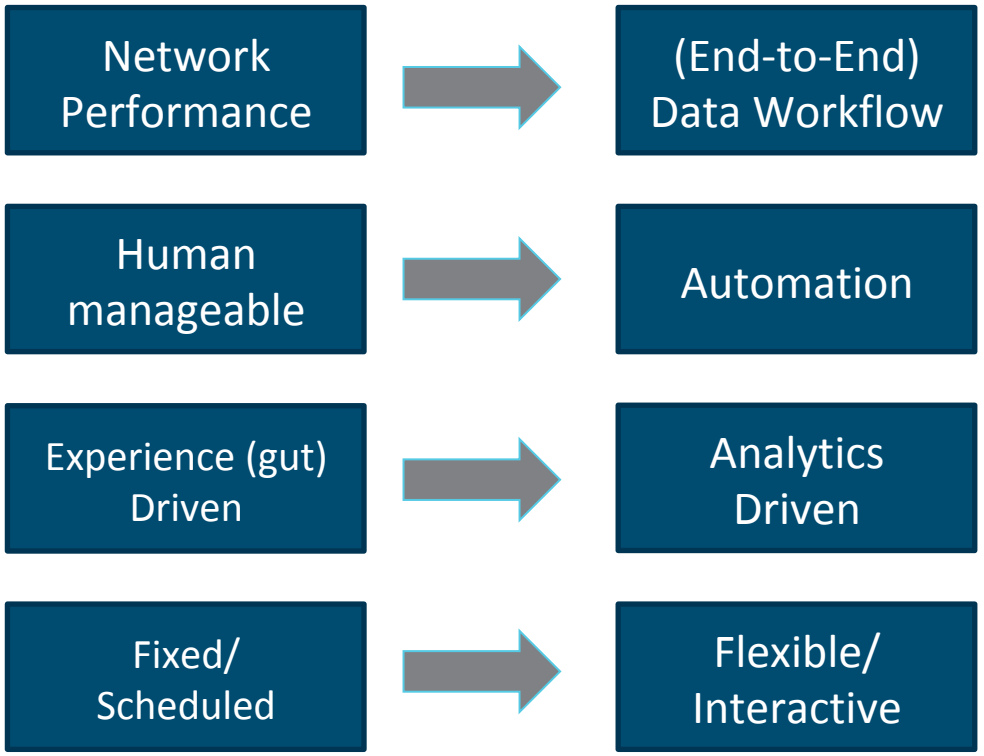

This assumes 10x data reduction is achieved

# New instruments, more data: NCEM 4D-Stem




ESnet

# Science Data 'Tsunami' driving network transformation




## Successful initiatives

Science DMZ

perfSONAR

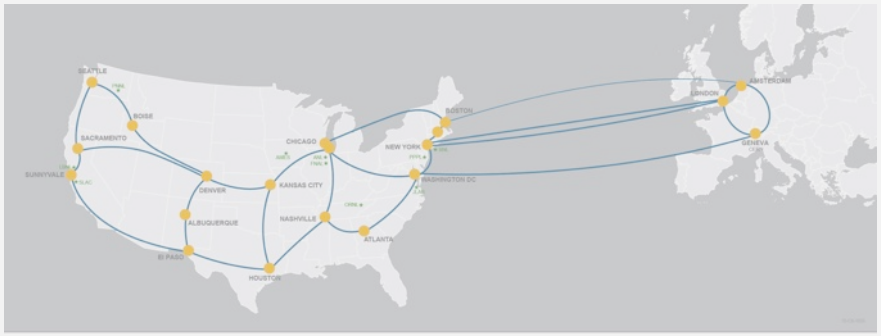
OSCARS

my.es.net Portal

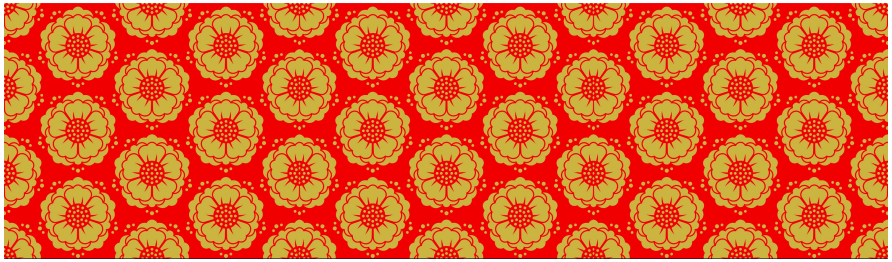
ESnet

# Talk

ESnet Introduction




**Established Design Patterns**



Emerging Design Patterns

ESnet

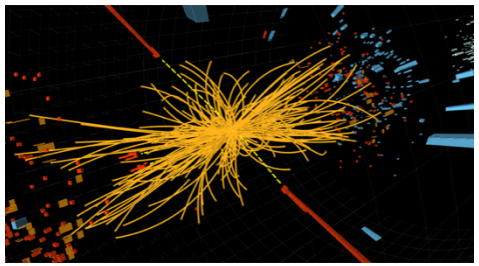
# Learning from nature:
# Infer and Codify the underlying design pattern

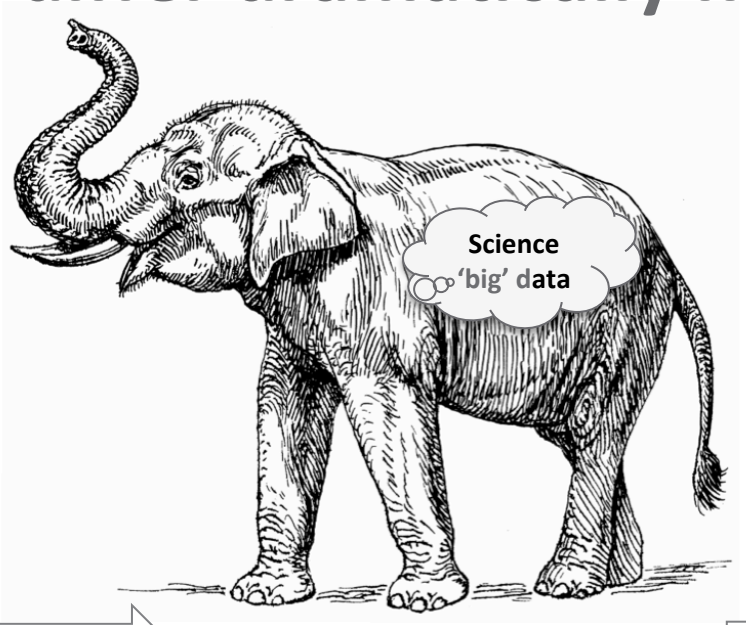
ESnet

# Design Pattern #1: Protect your *Elephant* Flows

ESnet

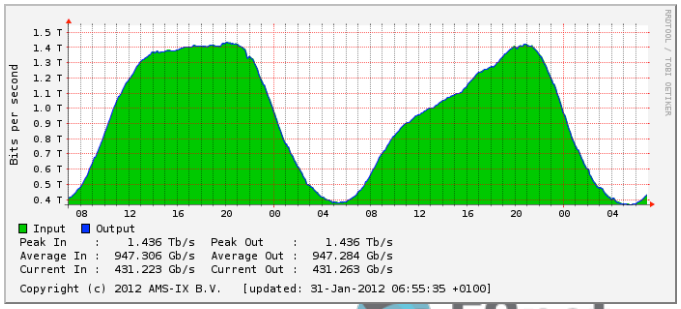
# ESnet is built to handle science's 'big' data whose traffic patterns differ dramatically from the Internet







Video
Cloud Apps
Internet

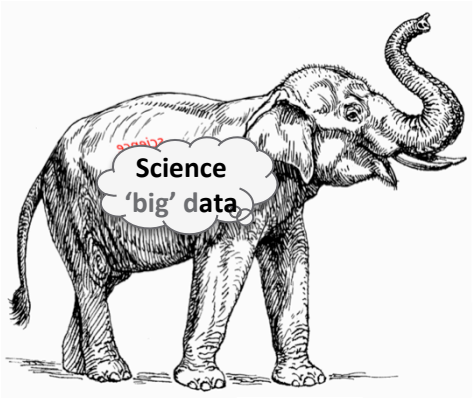




ESnet

# 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

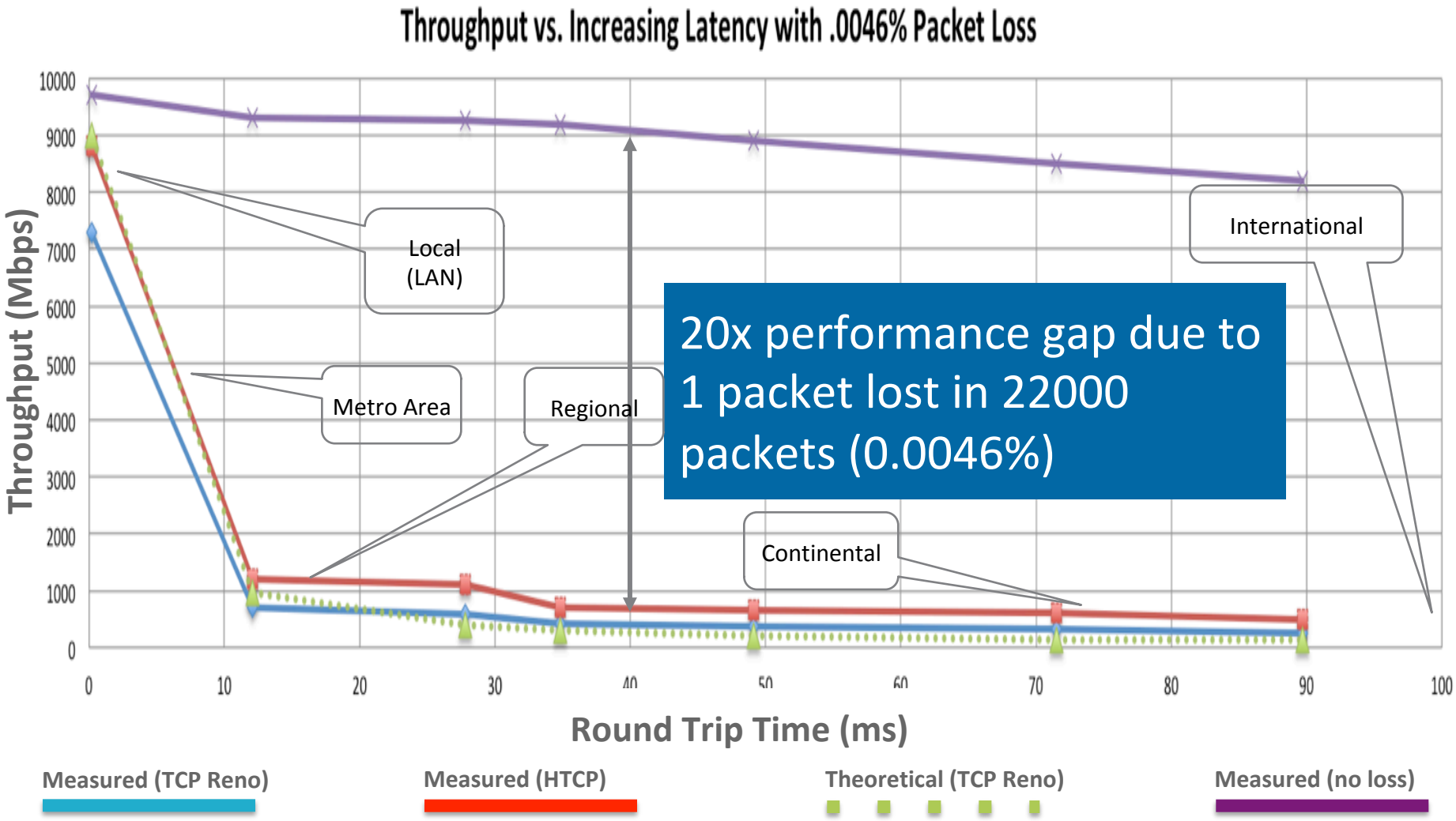
$$\frac{\text{maximum segment size}}{\text{round-trip time}} \times \frac{1}{\sqrt{\text{packet-loss rate}}}$$

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.

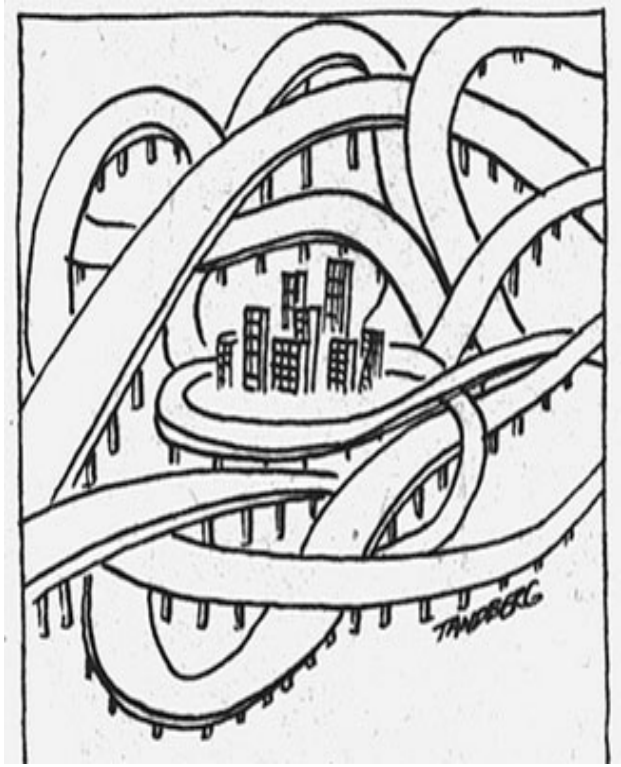
# Experimental results support the requirement to have a *lossless* network for high-performance




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.

# Design Pattern #2: Unclog your data taps

# Problem and Solution explained illustratively




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

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

ESnet

# Set right expectations with applications

| Data set size | 1 Minute | 5 Minutes | 20 Minutes | 1 Hour |
|---|---|---|---|---|
| 10PB | 1,333.33 Tbps | 266.67 Tbps | 66.67 Tbps | 22.22 Tbps |
| 1PB | 133.33 Tbps | 26.67 Tbps | 6.67 Tbps | 2.22 Tbps |
| 100TB | 13.33 Tbps | 2.67 Tbps | 666.67 Gbps | 222.22 Gbps |
| 10TB > 100Gbps | 1.33 Tbps | 266.67 Gbps | 66.67 Gbps | 22.22 Gbps |
| 1TB | 133.33 Gbps | 26.67 Gbps | 6.67 Gbps | 2.22 Gbps |
| 100GB 100Gbps | 13.33 Gbps | 2.67 Gbps | 666.67 Mbps | 222.22 Mbps |
| 10GB < 10Gbps | 1.33 Gbps | 266.67 Mbps | 66.67 Mbps | 22.22 Mbps |
| 1GB | 133.33 Mbps | 26.67 Mbps | 6.67 Mbps | 2.22 Mbps |
| 100MB < 100Mbps | 13.33 Mbps | 2.67 Mbps | 0.67 Mbps | 0.22 Mbps |
| Time to transfer | | | | |

This table available at:

http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/

ESnet

# *Emerging* global consensus around this architecture.

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

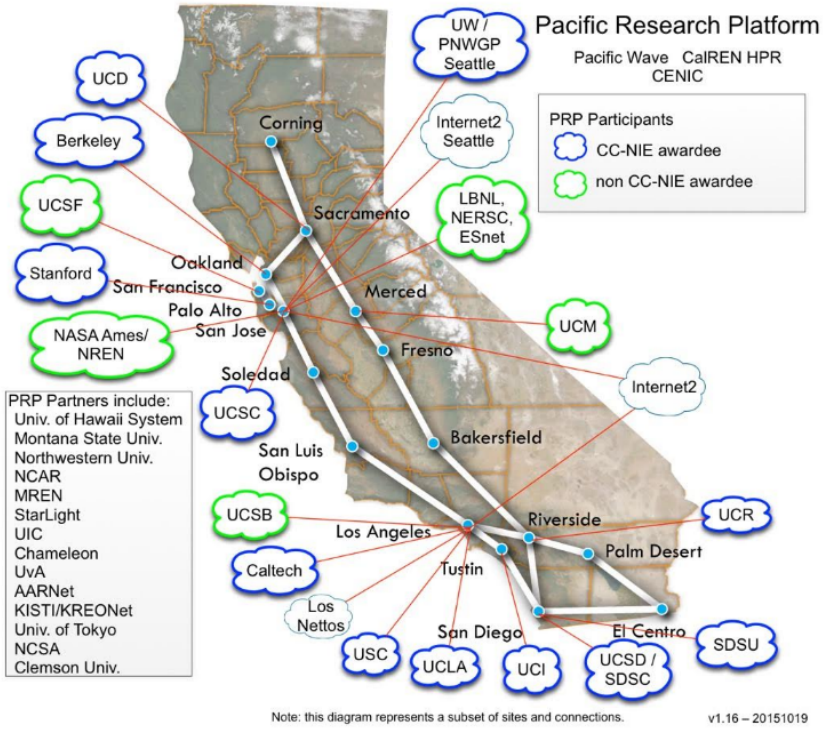
NSF has invested >>$120M to accelerate adoption.

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

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

ESnet

# Integrate and automate ScienceDMZ's for collaborative science (PRP → NRP)




# Design Pattern #3: Prepare your data cannons


ESnet

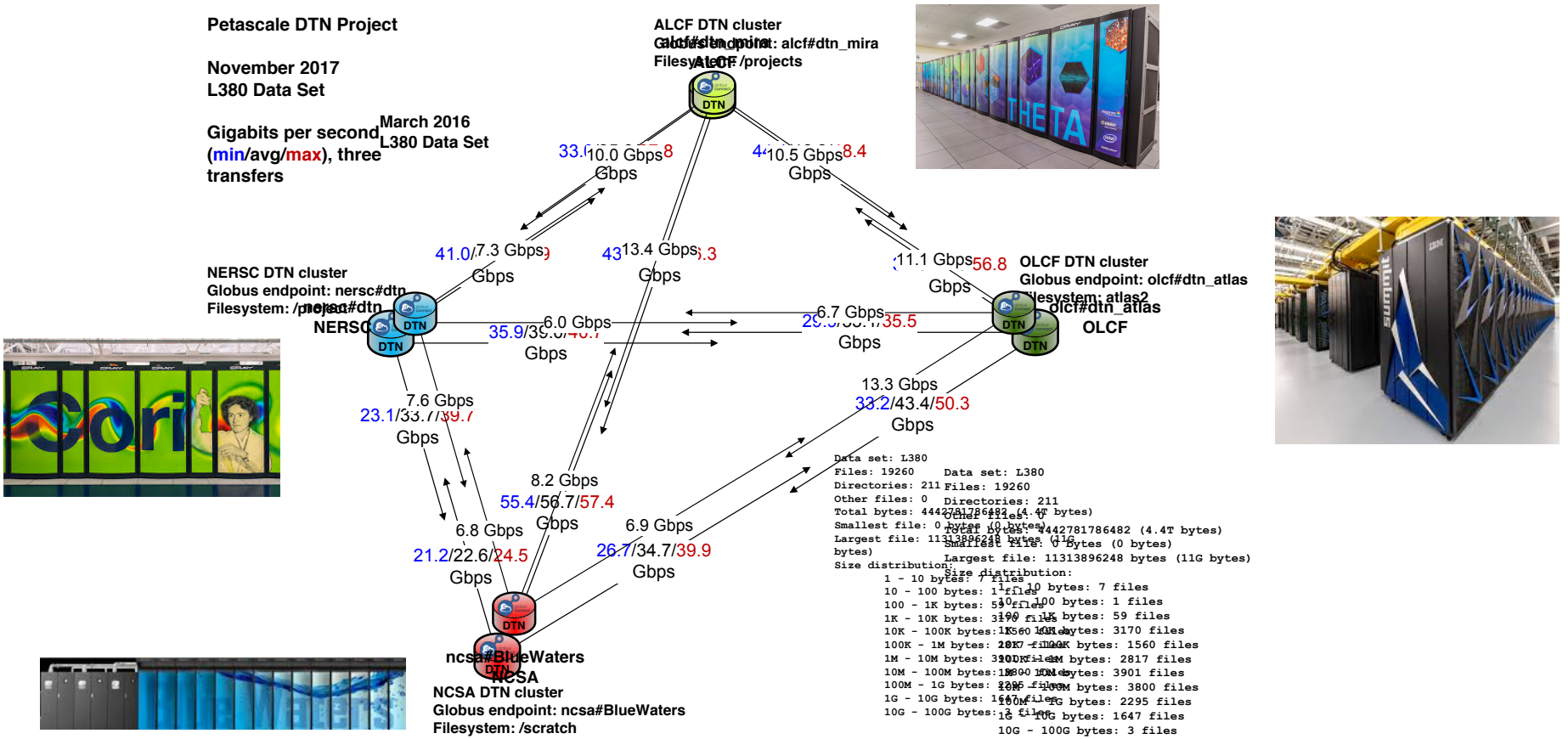
# Dedicated Systems – Data Transfer Node

- 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

ESnet

# Well-tuned Data Transfer Nodes















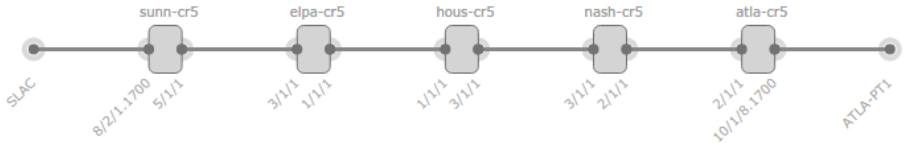


ESnet

# Data movement software keeps on improving: from 1 PB/week to 1 PB/day (approx.)

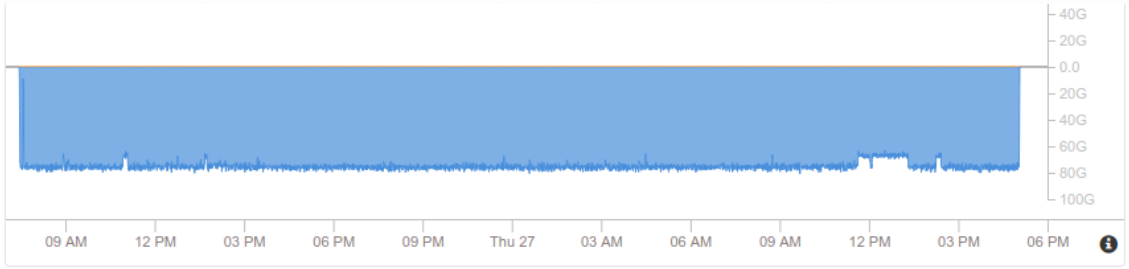
HOME › OSCARS »

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.

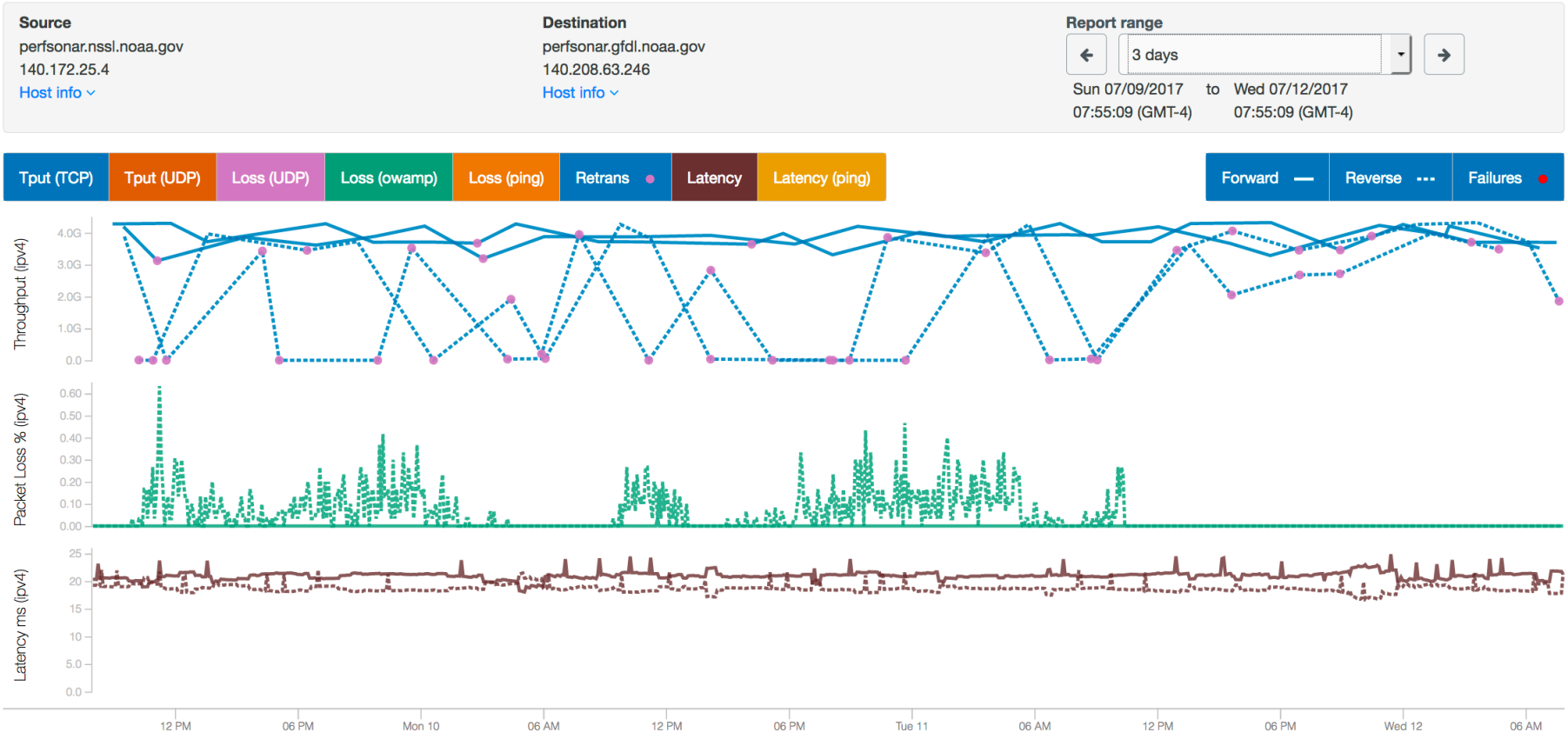



ESnet

# Design Pattern #4: Keep *flossing* the network

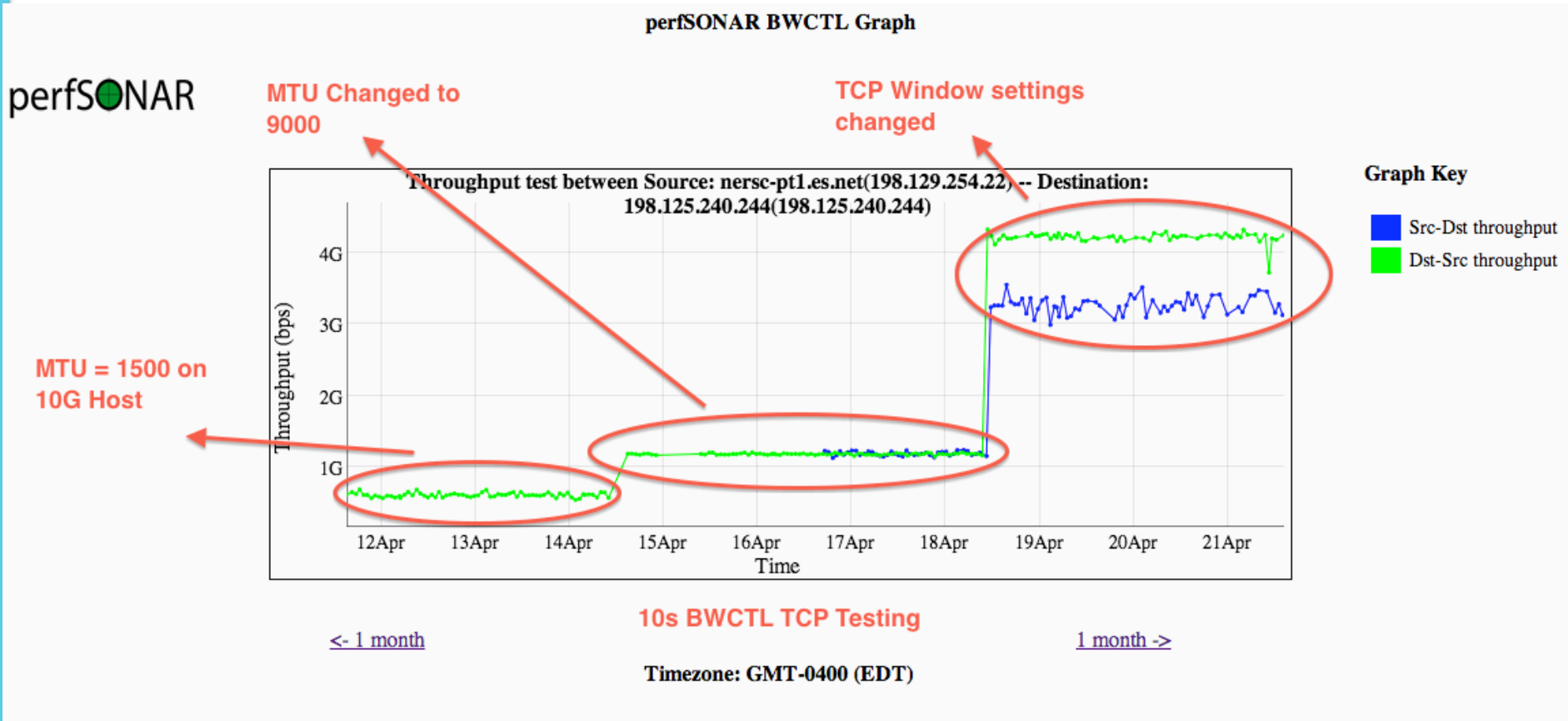
ESnet

# perfSONAR:
# continuous active monitoring of network




**Improving things, when you don't know what you are doing, is a random walk.**

# Another example, perfSONAR monitoring (contd.)

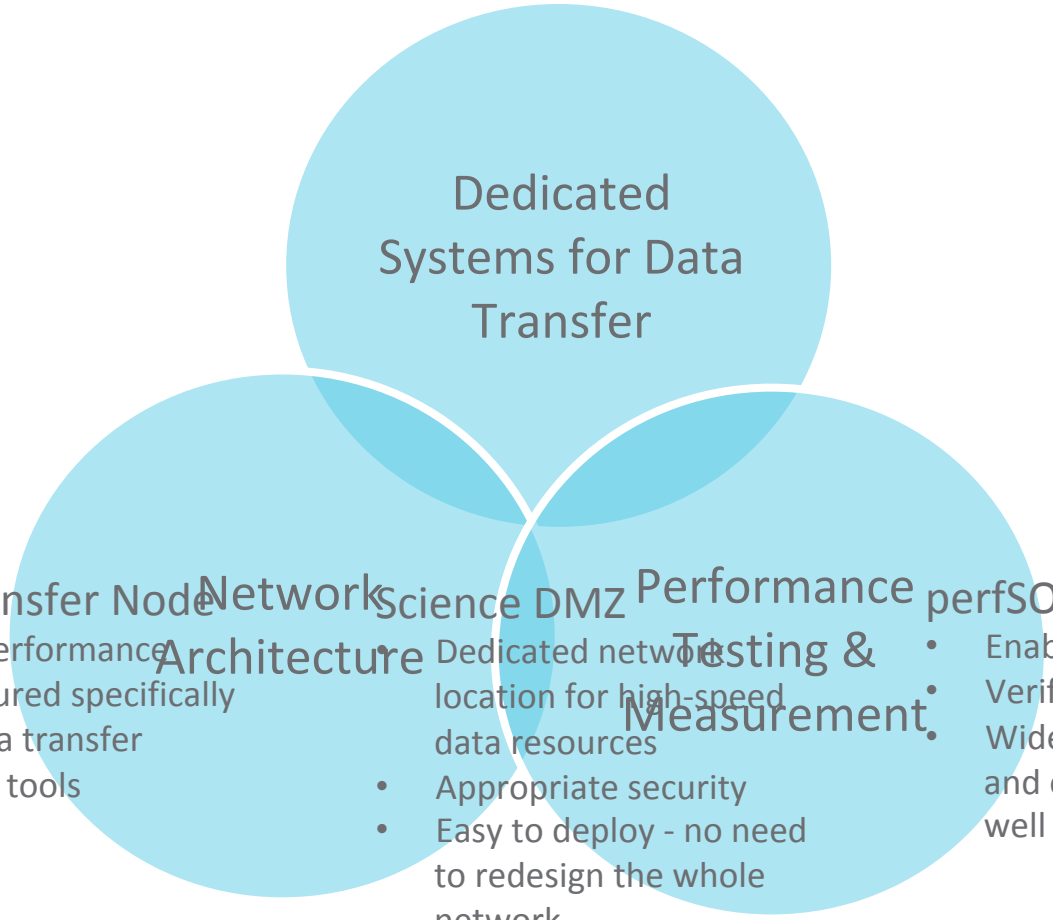


# Worldwide adoption (2000+ servers visible)


http://stats.es.net/ServicesDirectory/

ESnet

# The “Science DMZ” Design Pattern




**Data Transfer Node**
- High performance
- Configured specifically for data transfer
- Proper tools

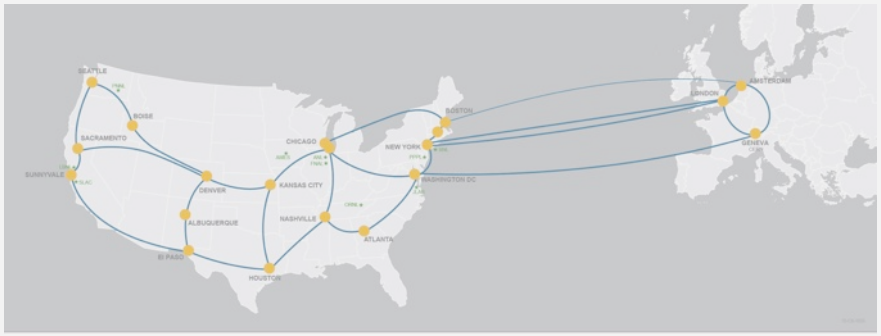
**Science DMZ**
- Dedicated network location for high-speed data resources
- Appropriate security
- Easy to deploy - no need to redesign the whole network

**perfSONAR**
- Enables fault isolation
- Verify correct operation
- Widely deployed in ESnet and other networks, as well as sites and facilities

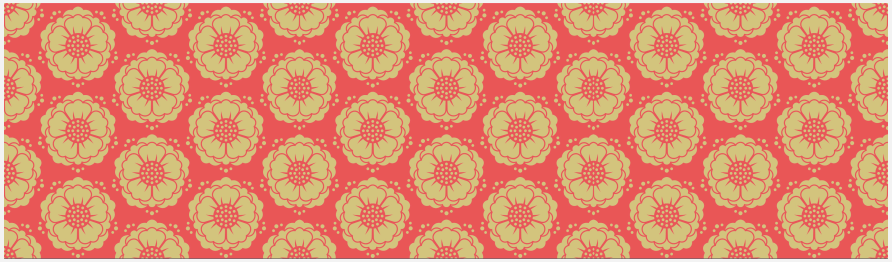
ESnet

# Talk

ESnet Introduction




Established Design Patterns



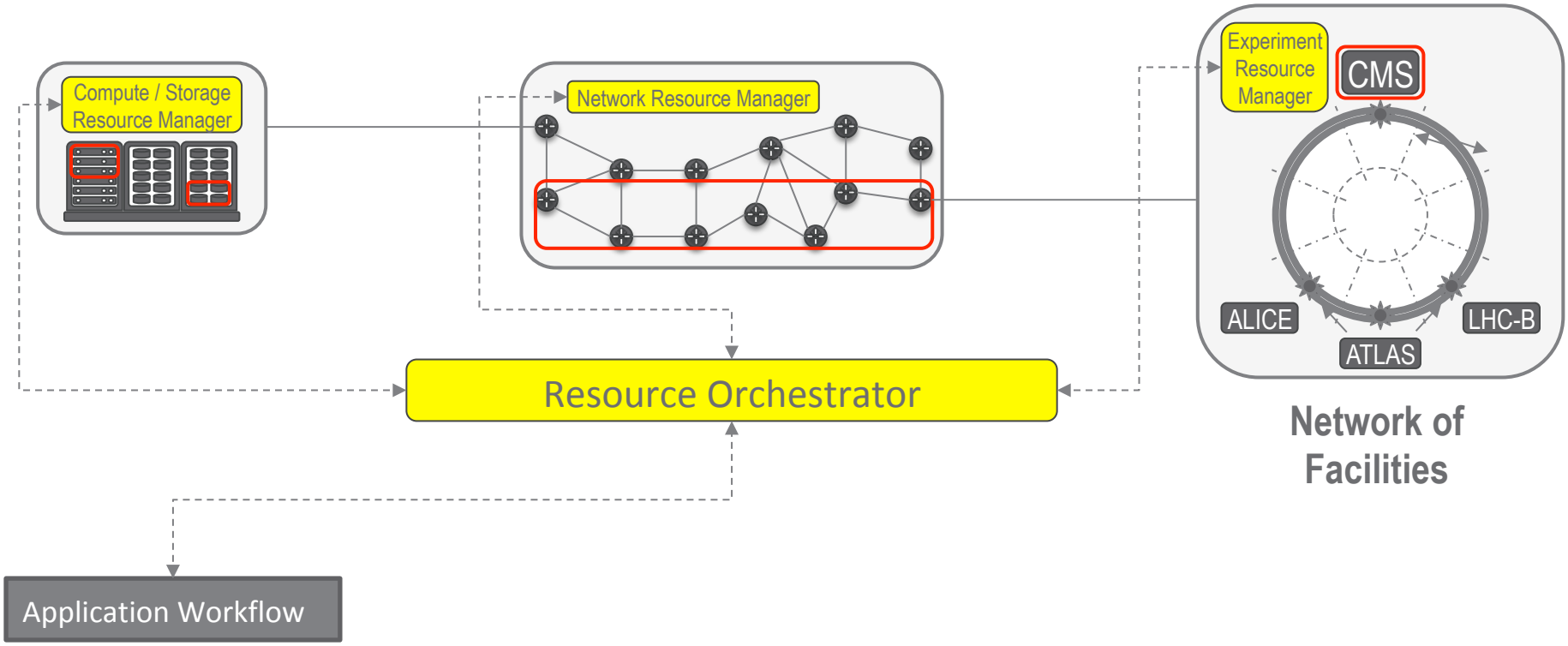
Emerging Design Patterns

ESnet

# *Emerging* Design Pattern #5: End-to-end, multi-domain science infrastructure

ESnet

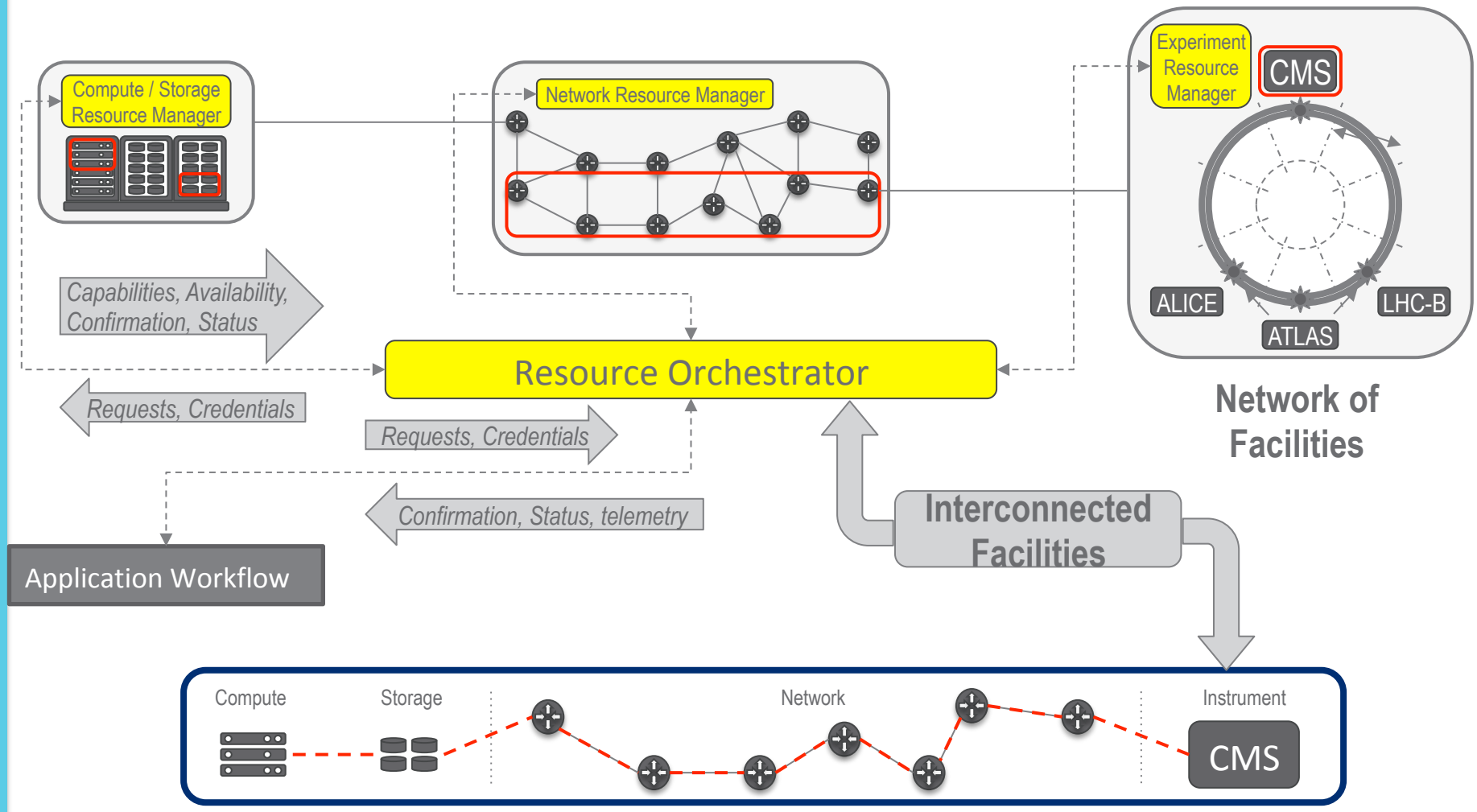
# End-to-End means more than the network





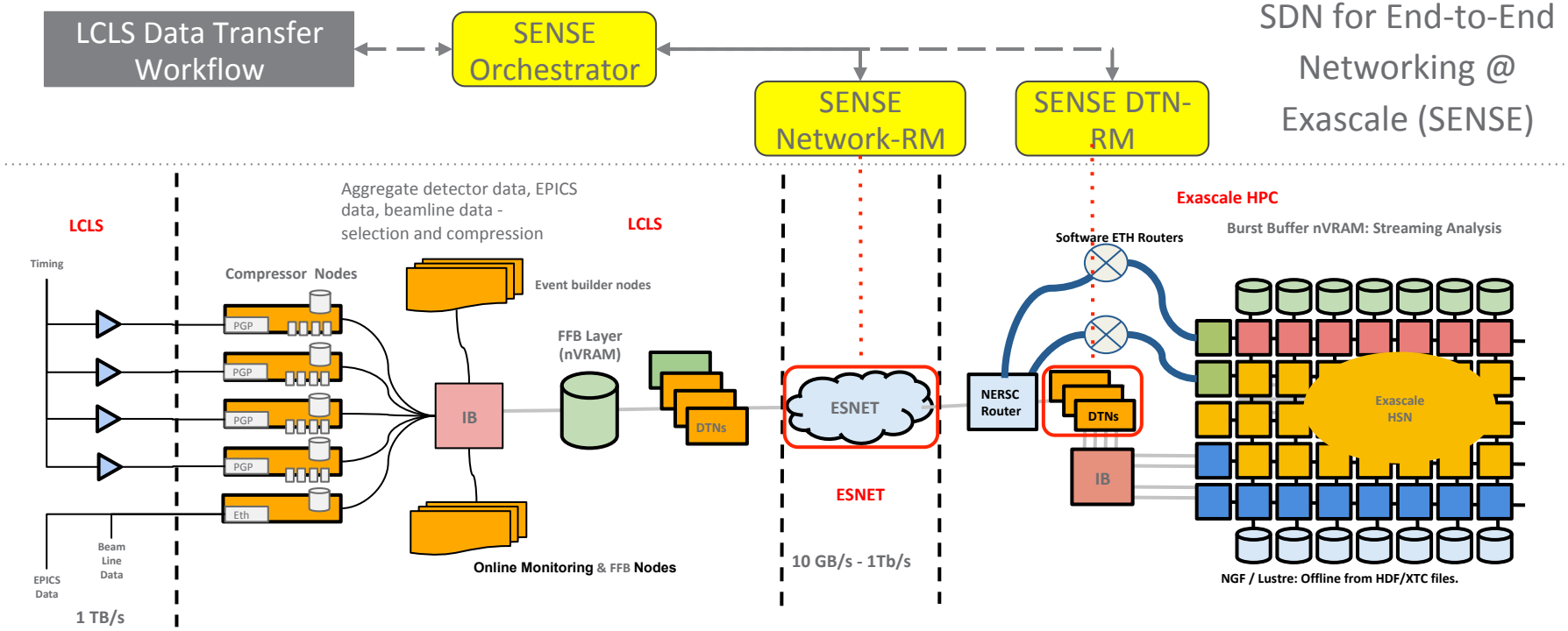

ESnet

# End-to-End means more than the network




ESnet

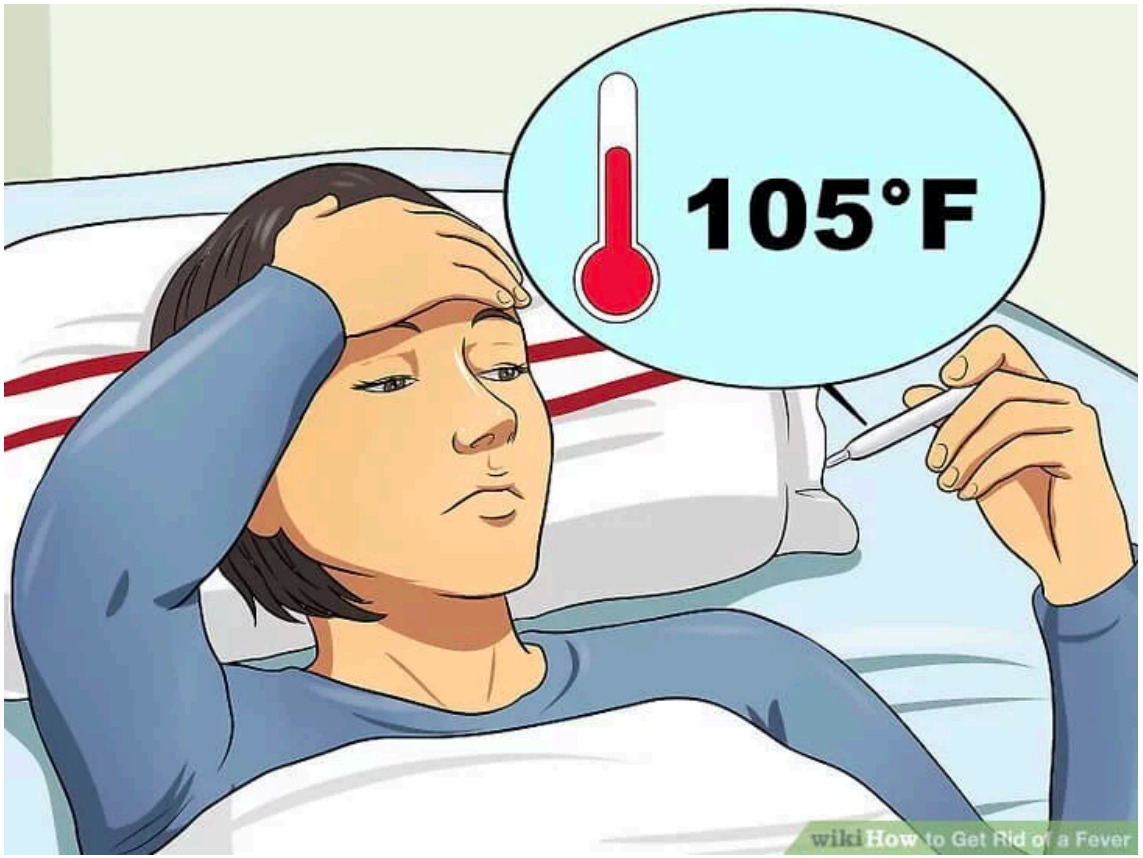
# ExaFEL: A science example of the Superfacility model




ExaFEL Data Flow

# *Emerging* Design Pattern #6: Data Driven Analytics and Learning

# Usually alert when something is broken…




**Why is this so?**

## …and then get expert help

ESnet

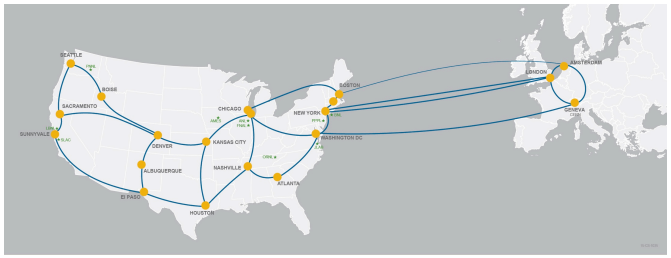
# Network Analytics

- Data being generated by the network every few seconds but not analysed or available for real-time analysis
  - The ability to ask questions of historical network data, and get answers
  - The answers updated with new data in near real-time
  - SNMP data, Flow data, Topology data, etc..

- Smart Cities, IoT, Smart Grid – have common problems

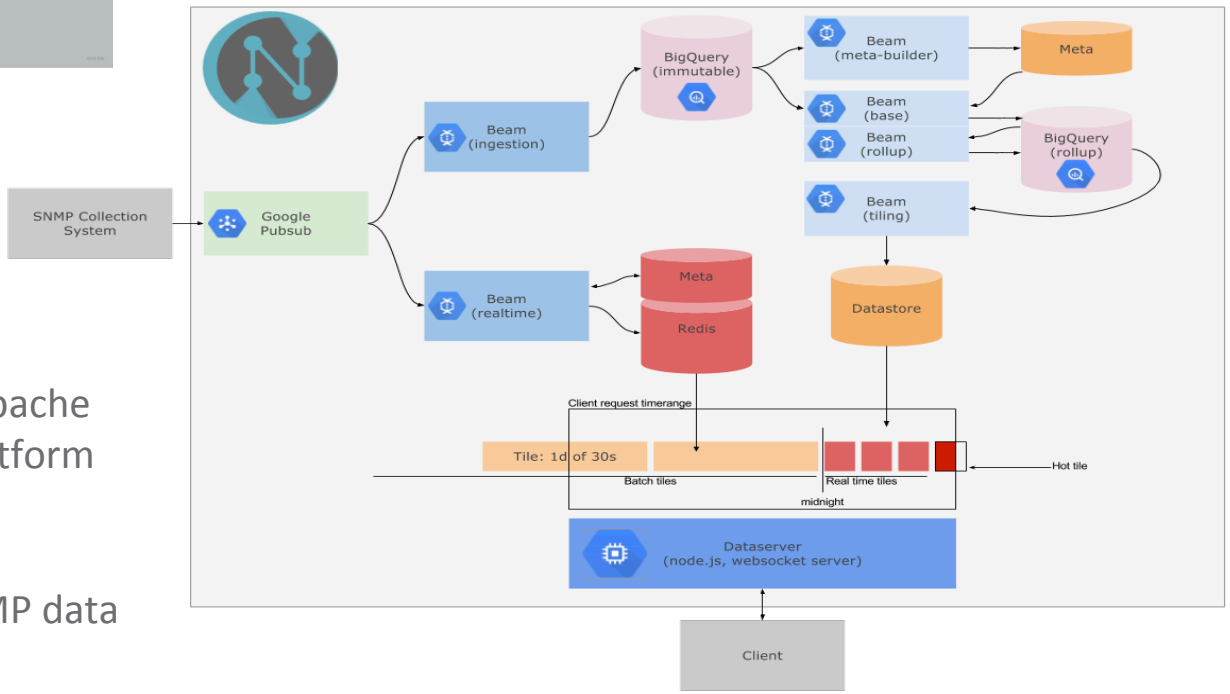
ESnet

# Deeper visibility and data-driven decisions: its about telemetry and analytics



Streams telemetry

On-demand analytics infrastructure



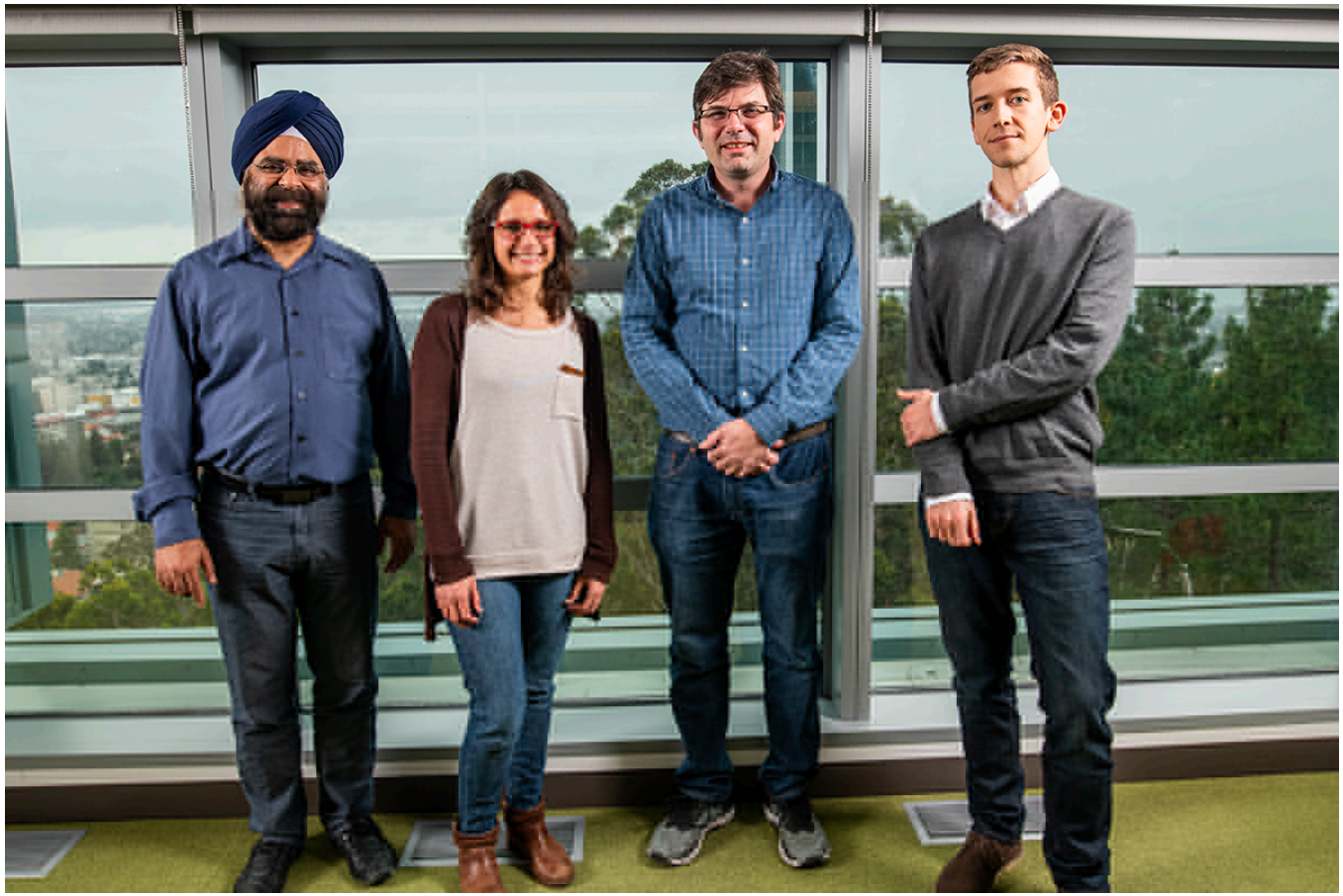

- Real-time telemetry from the network
- **netbeam** platform: Using Apache BEAM and Google Cloud Platform
- Both Batch and Stream processing in parallel
- In production for ESnet SNMP data

ESnet

# *Emerging* Design Pattern #7: Collaborations produce unexpected result

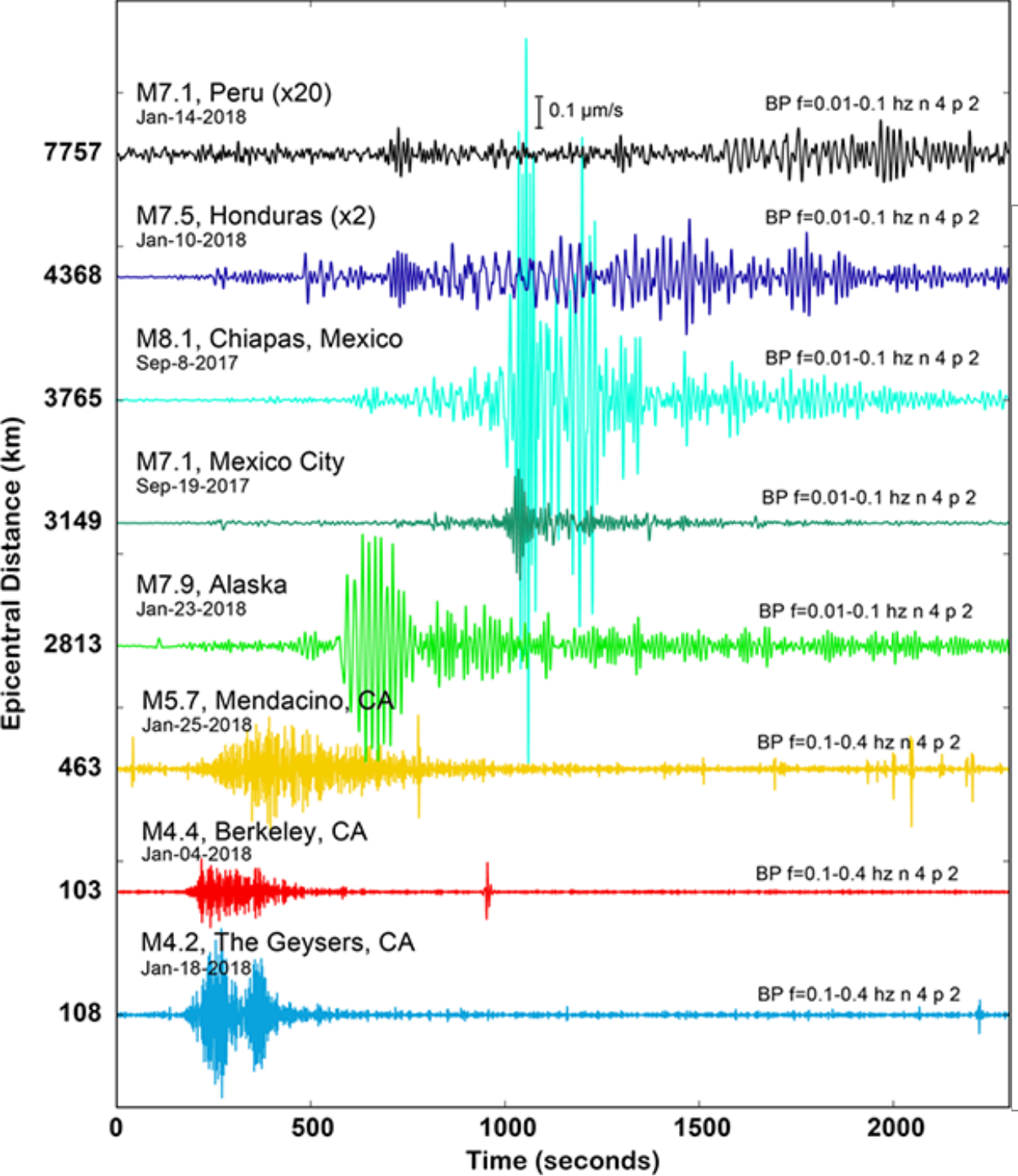
ESnet

# Collaboration between Earth Science and Networking



ESnet

# Try crazy ideas!



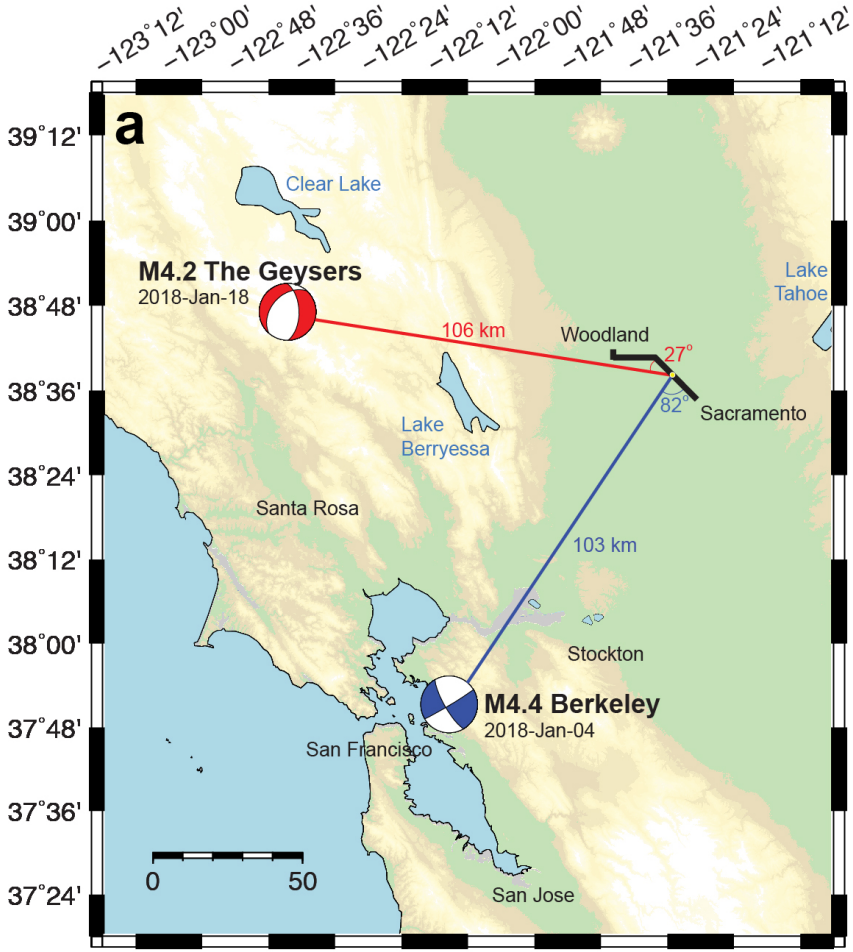

Dark Fiber Lays Groundwork for Long-Distance Earthquake Detection and Groundwater Mapping (Nature's Scientific Reports, 2019)

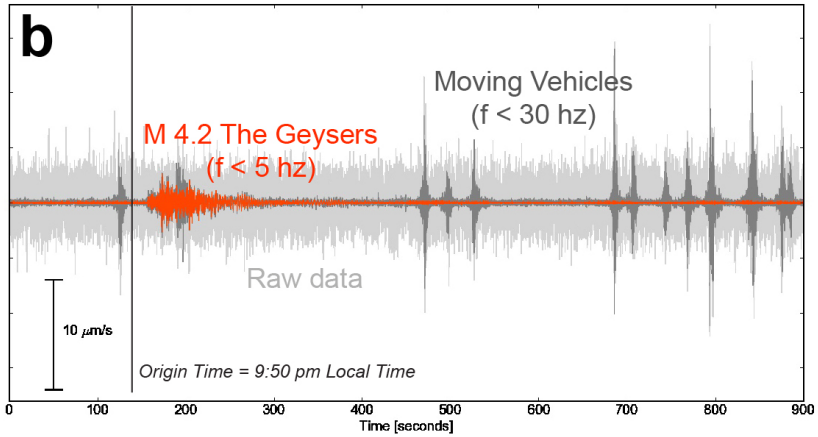
https://newscenter.lbl.gov/2019/02/05/dark-fiber-lays-groundwork-for-long-distance-earthquake-detection-and-groundwater-mapping/

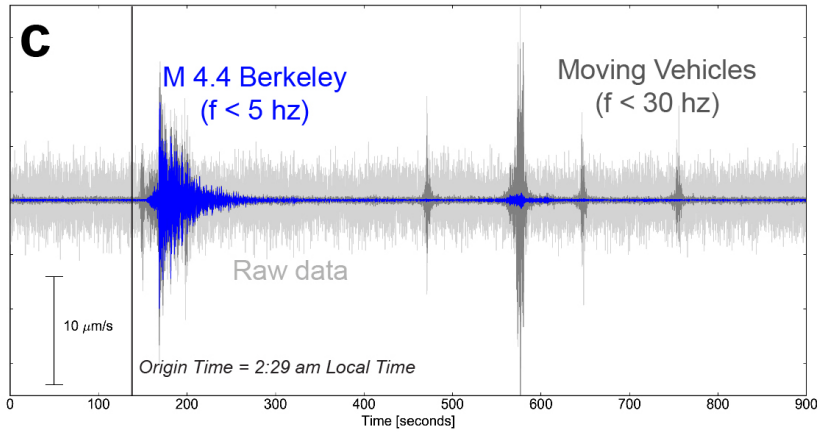
https://www.nature.com/articles/s41598-018-36675-8

# Impact of Filtering on Event Signatures

- Challenge : presence of significant random and correlated noise sources
- Vehicle traffic useful for Vs inversion, bad for EQ seismology
- Initial lowpass (30 hz top) removes some optical noise
- Below 5 Hz, EQs relatively clear, challenge is distinguishing cars from small EQs

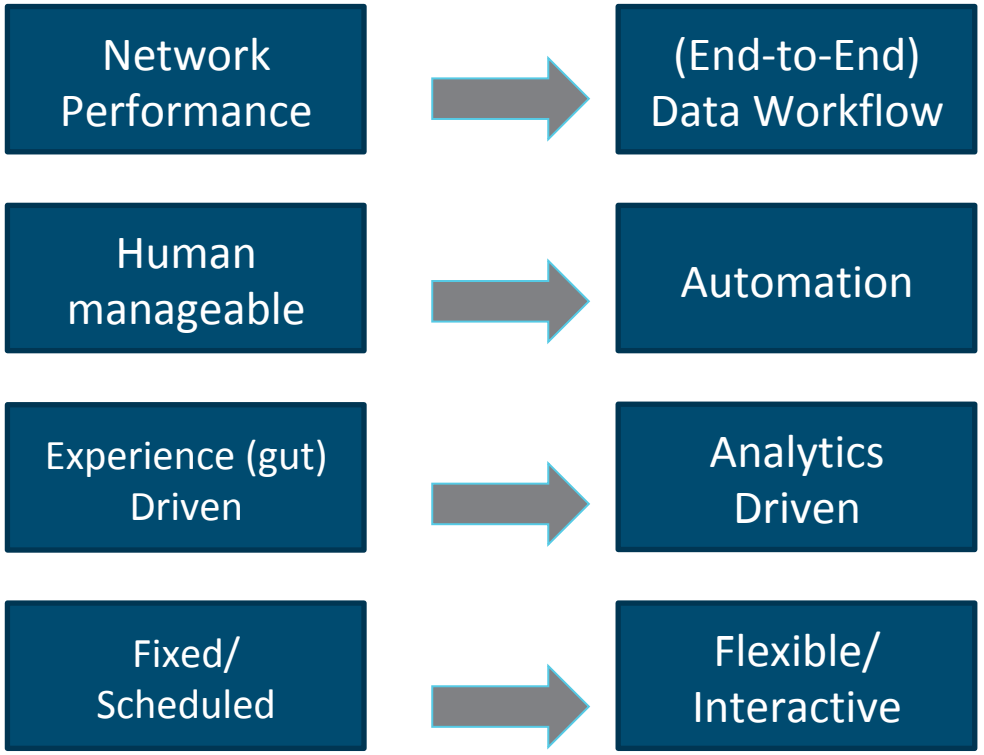


# What's next?

# Science Data 'Tsunami' driving network transformation




## Initiatives in progress

*SENSE*

ESnet6 Orchestration

Machine Learning
*Netbeam*
ESnet6 Telemetry

ESnet6 Architecture

ESnet

# ESnet6: ESnet's next-generation network

## Mission Need

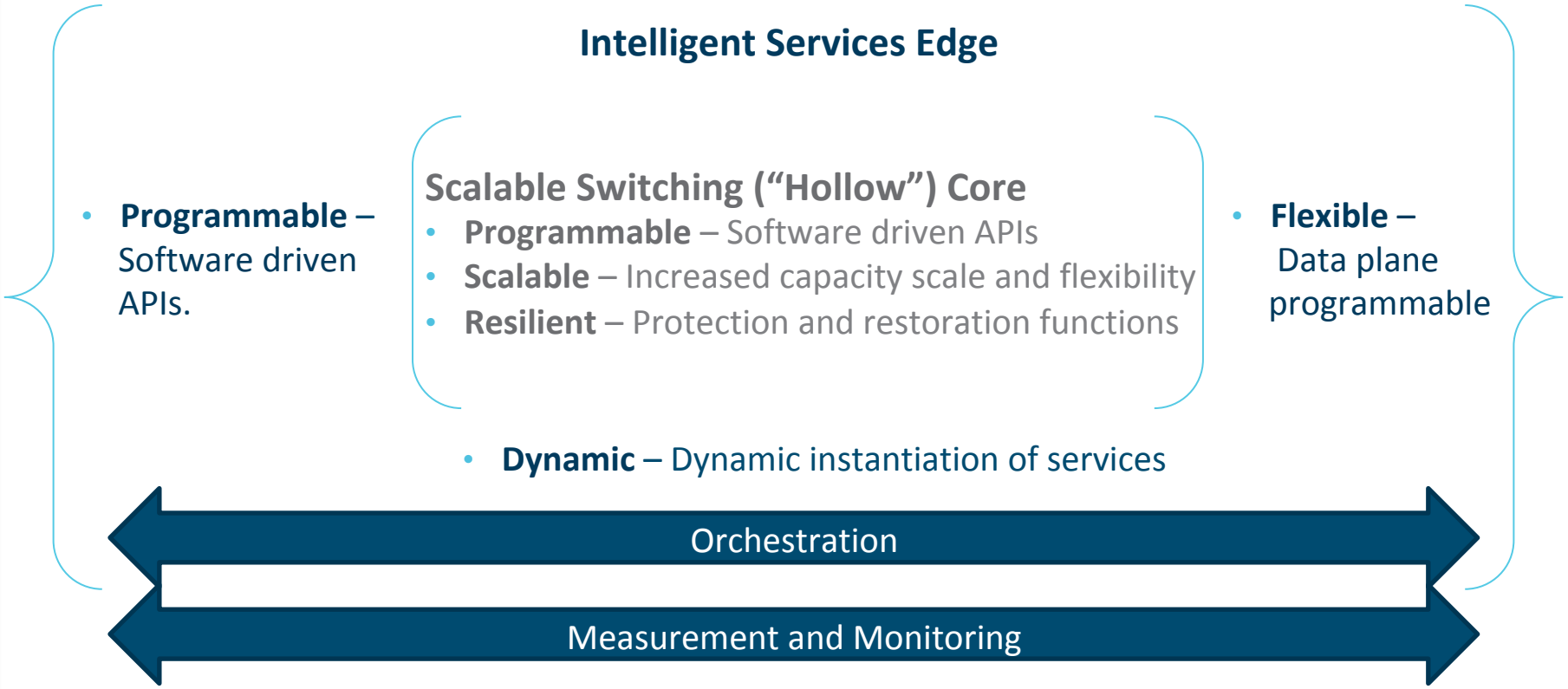
1. Capacity
2. Reliability and cyber-resiliency
3. Flexibility


- Innovative architecture on nationwide dark fiber
- Automation and programmability planned as key features

CD 0 Dec. 2016 → CD 1/3A Aug. 2018 → CD 2/3 *Dec. 2019* → CD 4 *Dec. 2023*

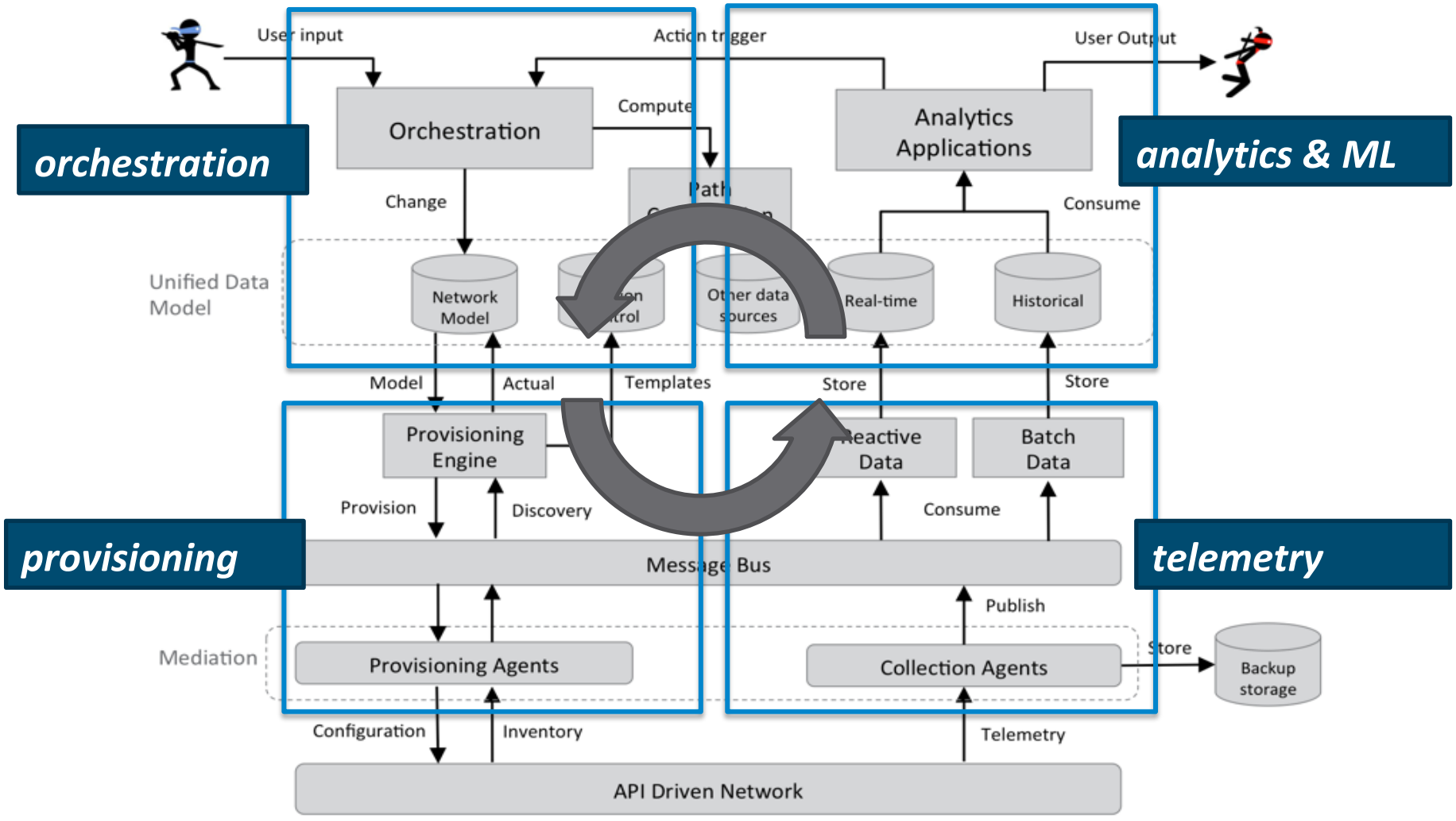
ESnet

# ESnet6 abstract architecture



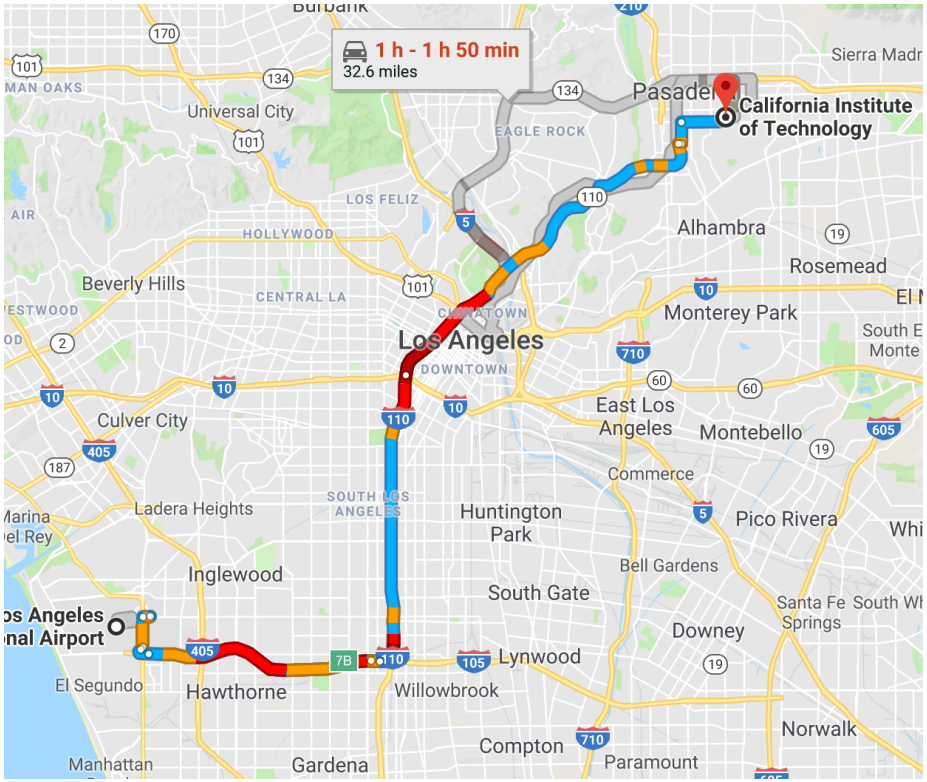

ESnet

# ESnet6 software automation architecture

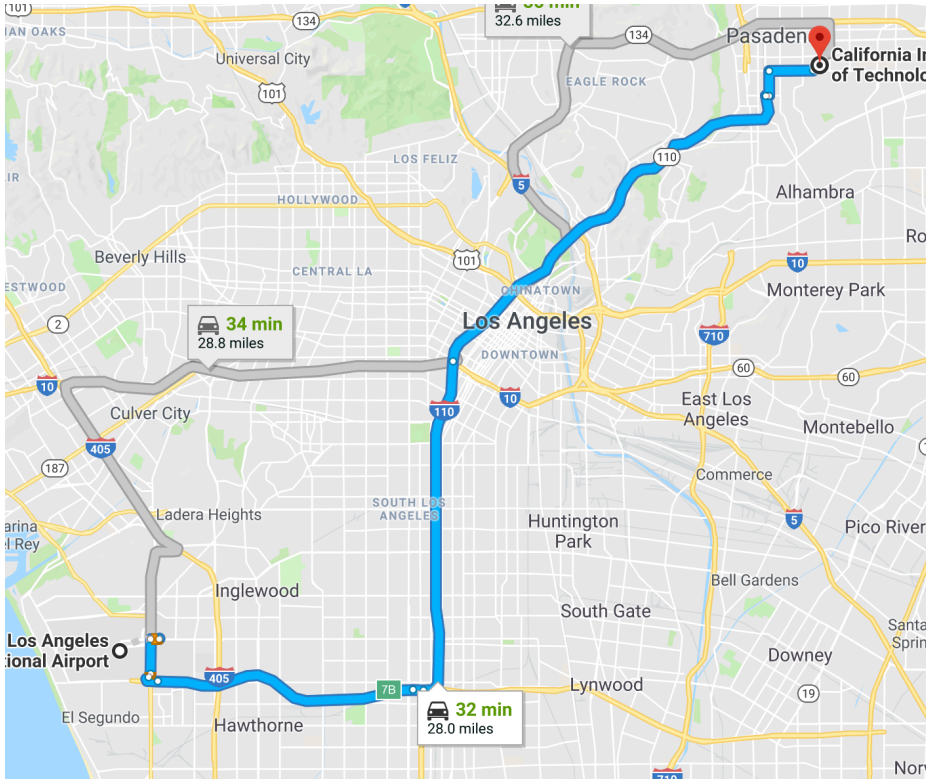



ESnet

# When will I get home?




LAX– Caltech, 6 pm:
1 hr – 1hr 50 min




LAX– Caltech, 11 pm:
32 min

# High-precision telemetry: deep insight into flows

Jupiter with the naked eye

Jupiter Close Up

**<u>Per flow, high-precision telemetry</u>**

- Per packet-metadata tracking (e.g. timestamp, ingress location, etc)
- 10 ns precision in timing

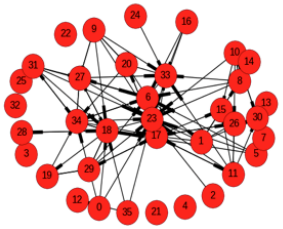
Use high-fidelity data to get better insights and analytics:

- Packet Microbursts
- Path deviations ( RTT and Delay )
- Security / anomaly detection
- Head of Queue Blocking
- Many others…

ESnet

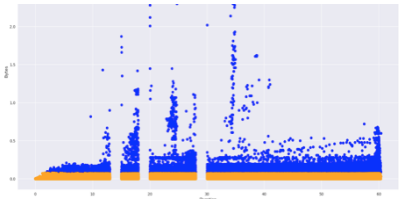
# Machine Learning applied to network telemetry data – learn, understand and optimize

*Understanding which sites are busiest at different times*



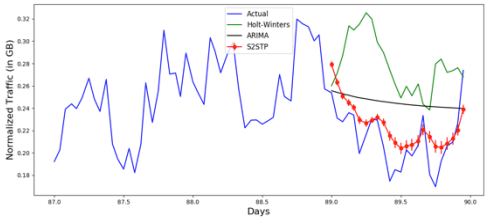
Markov Decision Processes and Bayesian approaches

*High-Speed classifying of big and small flows to redirect packet routes*



Gaussian Mixture models, other feature extraction methods (PCA, k-means)

*Prevent congestion and links failures by anticipating traffic 24 hours in advance*



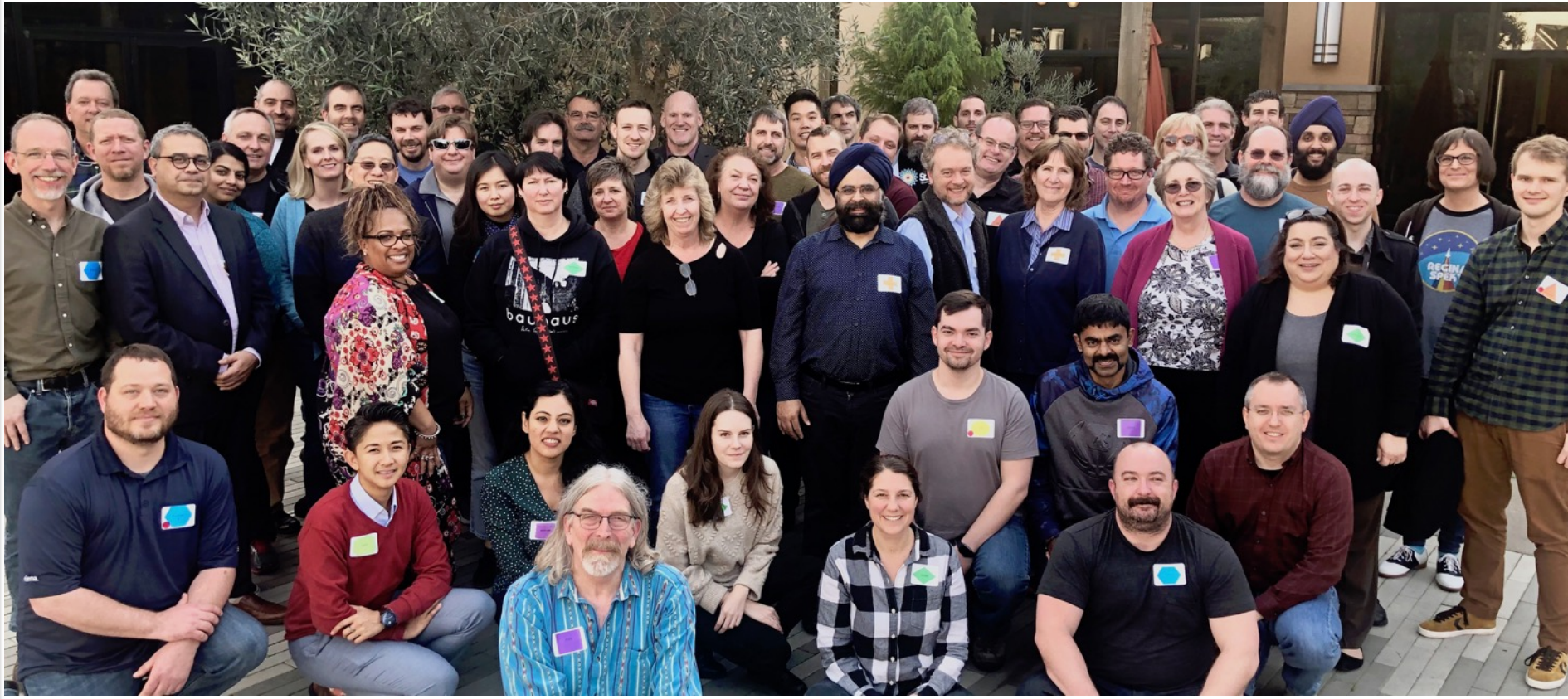

LSTM-autoencoder models, other classical time-series models (ARIMA, Holt-Winters, Box-Jenkins)

- Advancing network research and operations (with DL and non-DL approaches)
- Scaling solutions to our network complexity



ESnet

# Acknowledgements to the ESnet team!



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 **network design and application interaction** is extremely important to accomplish the end-to-end vision

ESnet

# Thank you.

imonga@es.net

ESnet