ESnet
ENERGY SCIENCES NETWORK

Future Landscape of Science Networking

Eli Dart, Network Engineer
ESnet Science Engagement
Lawrence Berkeley National Laboratory

Snowmass Computational Frontier Workshop
Virtual
August 11, 2020

U.S. DEPARTMENT OF ENERGY
Office of Science
BERKELEY LAB

Outline

- Networking stack in relation to ESnet6
- Some current research efforts within ESnet
- Looking forward

Many additional ESnet contributors (Inder Monga, Chin Guok, Dale Carder, Tom Lehman, Yatish Kumar)

Networking stack evolution

- Rough stack layer groupings
  - Optical
  - Frames and Packets
  - Applications
  - ESnet6 focused on all three of these

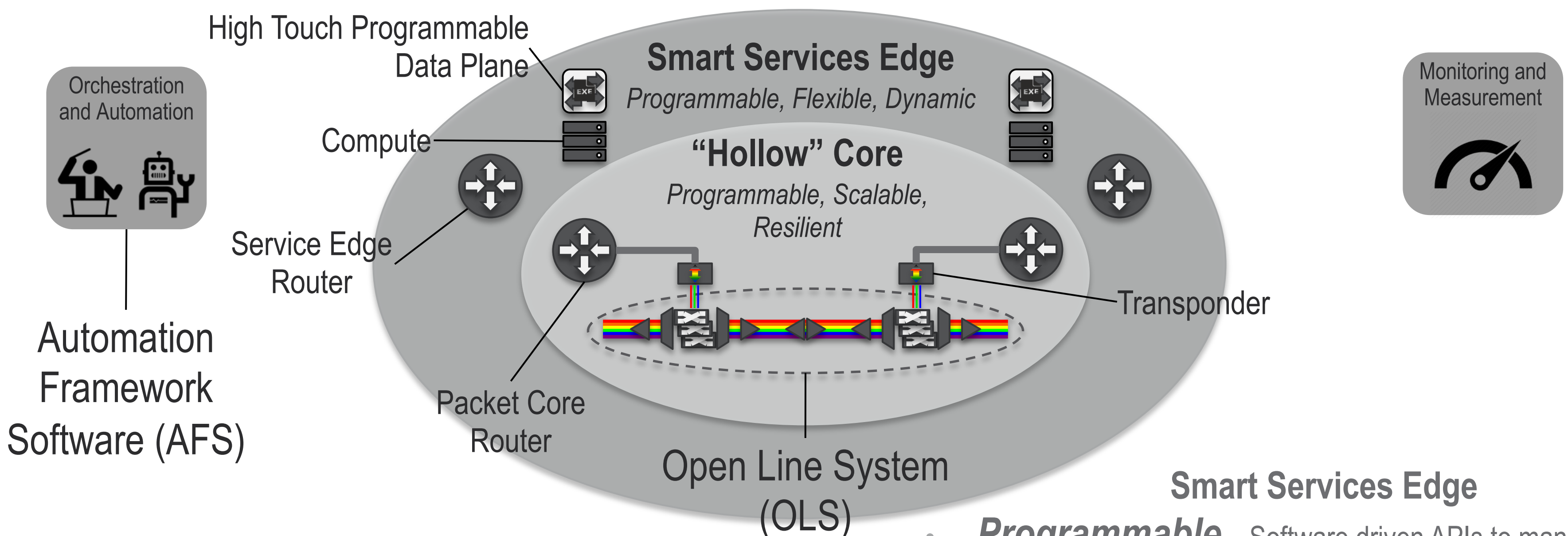
- Optical – open line systems, spectrum
  - Already revolutionizing R&E networks, including ESnet and GEANT
  - Decreased costs and increased flexibility at the optical layer
  - Overlay networks at optical layer in addition to frame/packet layer

Networking Stack Evolution (2)

- Frames and Packets
  - TCP/IP stack is here to stay
    - Working outside of TCP/IP is going to be challenging, esp. for hosts and apps
    - Some progress likely on TCP (BBRv2 and successors)
  - Virtual circuits enhanced
    - SENSE to address end to end issues
    - Traffic isolation, bandwidth/service guarantees using converged network infrastructure
  - LHCONE continuing forward
- Application layer – wide open
  - In-network caching
  - Remote I/O vs. file/object transfer
  - Workflow, orchestration, federation
  - Integration with lower layers is key research area

ESnet

ESnet6 ("Hollow-Core") Architecture Overview



Smart Services Edge

- ***Programmable*** – Software driven APIs to manage edge router/switch and retrieve telemetry information.
- ***Flexible*** - Data plane programmable switches (e.g. FPGA, NPU) in conjunction with compute resources to prototype new services (driven by Software Defined Networks (SDN))
- ***Dynamic*** – Dynamic instantiation of services using SDN paradigms (e.g. Network Function Virtualization (NFV, Virtual Network Functions (VNF), service chaining).

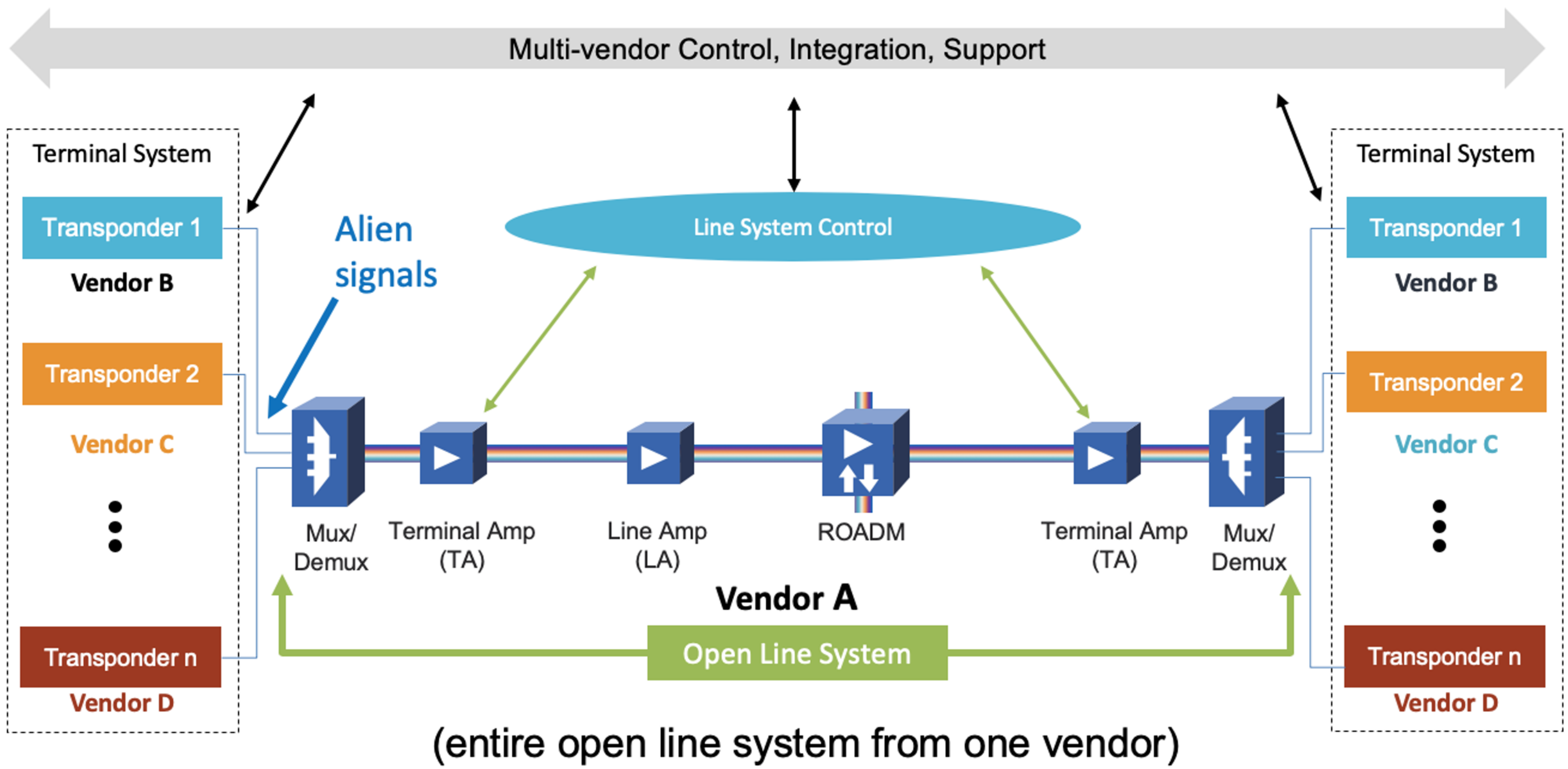
"Hollow" Core

- ***Programmable*** – Software driven APIs to allocate core bandwidth as needed, and monitor status and performance.
- ***Scalable*** – Increased capacity scale and flexibility by leveraging latest technology (e.g. FlexGrid spectral partitioning, tunable wave modulation).
- ***Resilient*** – Protection and restoration functions using next generation Traffic Engineering (TE) protocols (e.g. Segment Routing (SR)).

ESnet

ESnet6 Optical Network

Open Line System with Disaggregated Transponders



FABRIC

- A National-Scale Programmable Experimental Network Infrastructure
- Revolutionize Internet architectures with @scale integration of in-network compute, storage, accelerators with high-speed optical

This work is funded by NSF grant: CNS-1935966

FABRIC

- Built on ESnet6 Open Line System – example of optical spectrum capabilities
- Possible deployment of FABRIC nodes in Europe – funding requested

ESnet

FABRIC: Broad Research Infrastructure

FABRIC Enables New Internet and Science Applications

- Stateful network architectures, distributed applications that directly program the network

FABRIC Advances Cybersecurity

- At-scale realistic research facilitated by peering with production networks

FABRIC Integrates HPC, Wireless, and IoT

- A diverse environment connecting PAWR testbeds, NSF Clouds, HPC centers and instruments

FABRIC Integrates Machine Learning & Artificial Intelligence

- Support for in-network GPU-accelerated data analysis and control

FABRIC Helps Train the Next Generation of Computer Science Researchers

ESnet

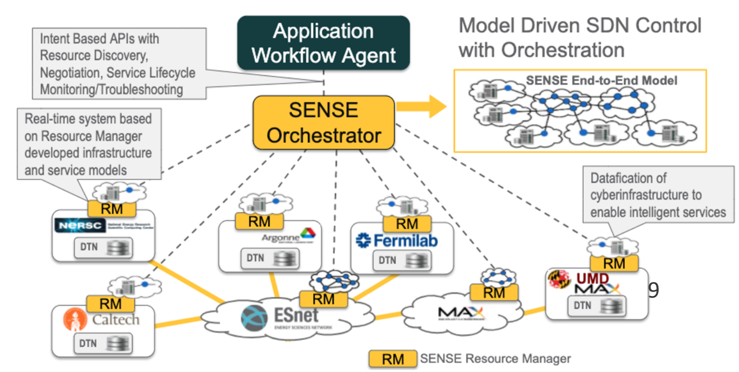
SENSE: SDN for End-to-end Networked Science at the Exascale

This work was supported by the Director, Office of Science. Office of Advanced Scientific Computing Research. Mathematical, Information, and Computational Sciences Division under U.S. Department of Energy Contract No. DE-AC02-05CH11231.

- **Motivation and Objective**
  Building an intent-driven, multi-domain, orchestration framework to allow distributed science workflows to intuitively provision and manage end-to-end network services including site-edge data caching.

- **Significance and Impact**
  Policy-guided end-to-end orchestration of network resources to enable real time coordination of network, compute, and storage resources -- this allows for science application workflows to interact with the network as a first class resource to drive end-to-end service provisioning, workflow optimization based on end-to-end resource discovery and performance information, and real-time interaction and negotiation based decision making.

- **Research Details**
  - Model Driven SDN Control with Orchestration.
  - Intent based science application facing APIs with resource discovery and negotiation.
  - Automated end-to-end troubleshooting and debugging.
  - Datification of cyberinfrastructure to enable intelligent services.



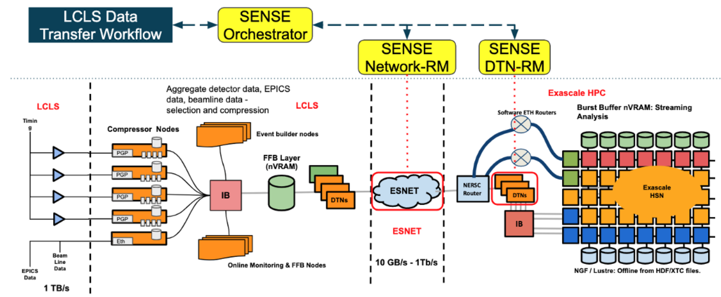
SENSE Architecture

SENSE: SDN for End-to-end Networked Science at the Exascale

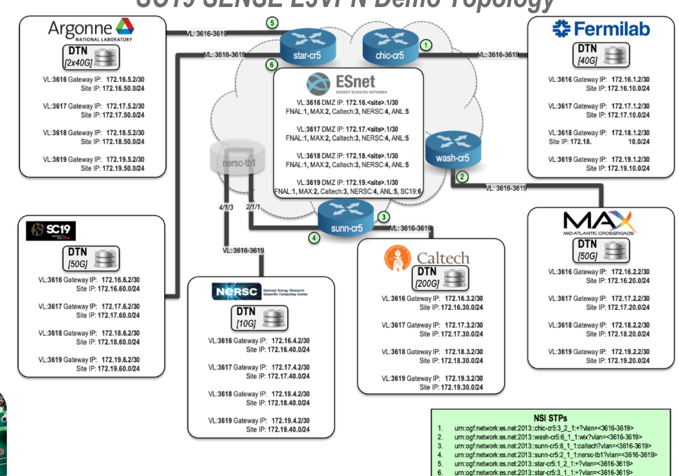
Ongoing Activities

- Integration of SENSE resource orchestration into prototype ExaFEL data workflow to demonstrate support for real-time feedback capabilities.
- Adoption of SENSE as part of LBNL Superfacility effort.
- Development of L3VPN service model in addition to L2 Services.
- Integration of SENSE and AutoGOLE to enable true end-to-end multi-resource services.
- Deployment of SENSE (beyond the established SENSE testbed in ANL, Caltech, ESnet, FNAL, MAX, NERSC) at Starlight (Chicago, IL), MOXY (Montreal, Quebec), and NetherLight (Amsterdam, Netherlands), CERN, CENIC, Tokyo, others.

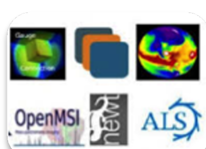
Automation Prototype of ExaFEL data workflow using SENSE



SC19 SENSE L3VPN Demo Topology

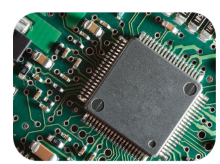


User Engagement



Data Lifecycle

Automated Resource Allocation



Computing at the Edge

Superfacility Initiative

ESnet

Caching: Motivation and Opportunities

- **Observations**
  - Data sharing between facilities and sites becomes more frequent
  - Significant portion of the popular dataset is transferred multiple times to different users and to the same user
  - New forms of data sharing other than files such as data streaming
  - Needs for near real-time data analysis, understanding, feedback and action for remote experimental science
- **Motivation**
  - Reduce data access limitations for scientists (in a smart way that does not require unaffordable network capacity)
  - Assist and accelerate data management approaches for large science collaborations that build their own software and data infrastructure
  - Incorporate elements of data management as a service within the network that make sense, and offer these common services to science applications
  - Provide value-added services by offering/integrating network reservation capabilities within the data management offerings
- **In-network temporary data caching opportunities for data sharing**
  - Provide opportunistic caching in the network to reduce the latency of data access by scientists
  - Leverage capacity on ESnet to pre-fetch the data, stage for multiple locations, and be shared by multiple user accesses
  - Explore if caching can reduce trans-atlantic bandwidth requirements

Xcache Pilot Experiment and Ongoing Efforts

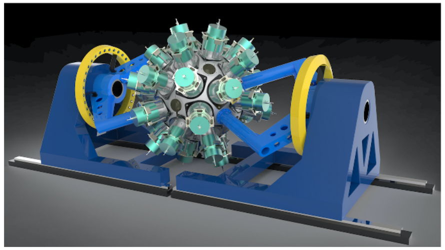
- **Apr. 2019 - July 2019, with US ATLAS and ESnet**
  - To understand how much data is being shared within the network
  - To determine how much network traffic bandwidth is saved consequently by relocating data
  - To study how much efficiency the temporary network cache gives to the application performance
  - To study how network cache storage helps network traffic performance and the overall application performance
- **Prototyping and engagement**
  - OSG/US-CMS (Nov. 2019 – present)
  - To study how network cache storage helps network traffic performance and the overall application performance
  - Accumulate experience on how the DOE scientific experiments and simulations share data among their users

Additional Research and Collaboration

- ESnet is engaged in multiple projects related to networking
- I will briefly discuss three of these (there are more)
  - Goal is to consider what is possible
  - Think about ways to work more closely together

GRETA: Gamma Ray Energy Tracking Array

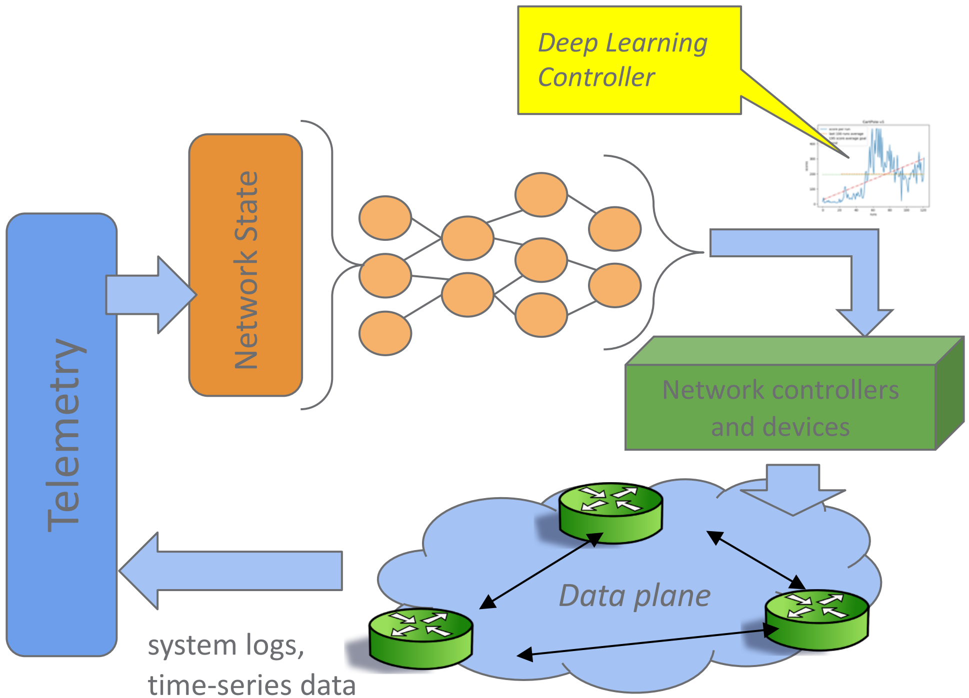
- Gamma Ray Energy Tracking Array - http://greta.lbl.gov
  - New instrument for low-energy nuclear science
  - GRETA marks a major advance in the development of γ-ray detector systems
  - Order-of-magnitude gains in sensitivity compared to existing arrays.
- ESnet contributions
  - ESnet is supplying the Architect for the computing environment
  - ESnet is developing two key high-performance software components
    - Forward Buffer
    - Global Event Builder
- Key ideas
  - Network-based data acquisition system
  - High-speed data processing software
  - Integration of high-speed networking from the beginning



GRETA conceptual rendering

ESnet

Deep Learning Research in Network Engineering



Research Questions:

Q1: Reduce traffic hotspots

Q2: Improve network performance for workflows/applications

Q3: Self-Learning Network controllers (more transfers, less congestion, etc.)

Quantum networking

- DOE Quantum Internet Blueprint Workshop
- Priority Research Opportunities:
  - Provide the Foundational Building Blocks for a Quantum Internet.
  - Integrate Multiple Quantum Networking Devices.
  - Create Repeating, Switching, and Routing for Quantum Entanglement.
  - Enable Error Correction of Quantum Networking Functions.
- ESnet participating, contributing, supporting

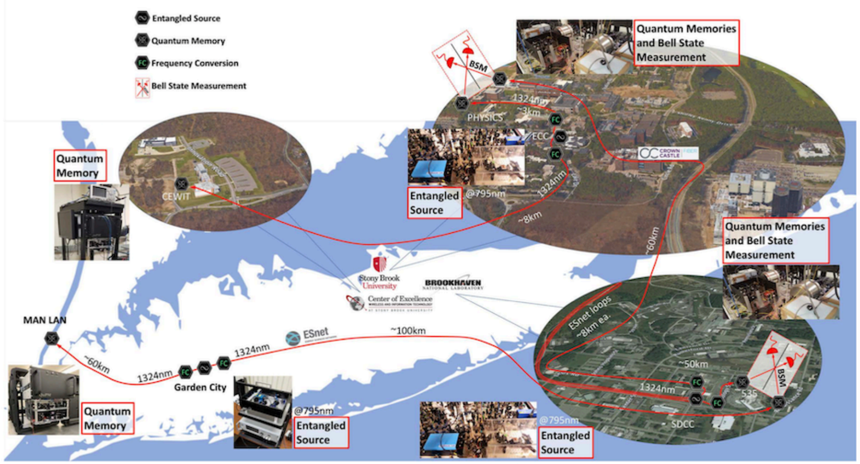


Figure 3.3. Long Island quantum network extended to New York City. The network will use a chain of quantum repeaters, extended across Long Island via three entangled sources, six quantum memories, and two entanglement swapping stations. Using ESnet's existing fiber infrastructure, the network will connect SBU to New York City via Brookhaven Lab with intermediate stations on the two campuses and in Garden City, N.Y. This is expected to be the first quantum repeater network of its kind in the world.

BNL/Stonybrook/ORNL/LANL quantum network testbed – ESnet supporting

- Many collaborations doing work
  - BNL/SBU/ORNL/LANL
  - FNAL/Caltech/ATT
  - MIT/Harvard/BBN
  - International as well (Canada, China, Austria, others)

ESnet

ESnet and Physics

- ESnet is planning for a Terabit future
- ESnet is already working on a variety of projects
- Clear value in integrating ESnet6 capabilities with LHC stack and services
- How shall we do this?
  - Architectural aspects
  - Service development and integration
  - Data challenges, testbeds, iteration
- Resources are not unlimited
  - Network bandwidth, people (among others)
  - Where is the greatest value?

Architectural Elements

- Some things can be improved by adopting useful architectures
- Interfaces to different things
  - Detectors, network, compute, storage (components)
  - HPC facilities, university clusters, experiments (facilities)
  - These will probably require different but interoperable interfaces
- Data staging vs. streaming data access (e.g. XRootD)
  - What do network-facing systems look like for each?
    - DTN clusters for data transfers
    - High performance network access from compute nodes for XRootD
- What are the right interfaces?
  - APIs, network services, …
  - Caching in the network, at sites, in clouds?

ESnet

Storage is Key

- We heard yesterday about challenges for storage
  - Tape: does it have a future?
  - Speed + scale for non-tape storage
  - Which locations will deploy significant storage?
- Need to track tradeoffs between networking and storage
  - If network bandwidth is free and infinite, some things make sense
  - If storage is infinitely fast and large, some things make sense
  - Neither of these are likely to be true (obviously!)
- Compute wherever you can, deliver data to compute from storage
  - What does this mean for the future?
  - Move compute to data or data to compute (networks support both!)
  - HPC vs. HTC
  - It is likely that we will need more than one model

ESnet

Engagement, Collaboration, Iteration

- Collaborating on R&D efforts benefits us all
- We need to work together over time
  - Human structures to enable long-term collaboration are more important than specific technologies
  - Data challenges drive the innovation, development, and deployment cycles
    - Very important!
- There are multiple current efforts
  - SENSE, LHCONE, caching, blueprinting, packet marking, …
  - New/evolving data-layer tools in the LHC stack
- What new structures do we need?

ESnet

Thanks!

Eli Dart

Energy Sciences Network (ESnet)

Lawrence Berkeley National Laboratory

http://my.es.net/

http://www.es.net/

http://fasterdata.es.net/