# Using High Performance Networks as a Foundation for the New Superfacility Model

## Presented by ESnet / Berkeley Lab

## Introduction

**Scientists move hundreds of petabytes of data around the world every year, and petabytes every month. The technologies, such as detectors, supercomputers, networks, used to acquire, transfer, analyze and store data are becoming increasingly more important and critical to enable scientific discovery.**

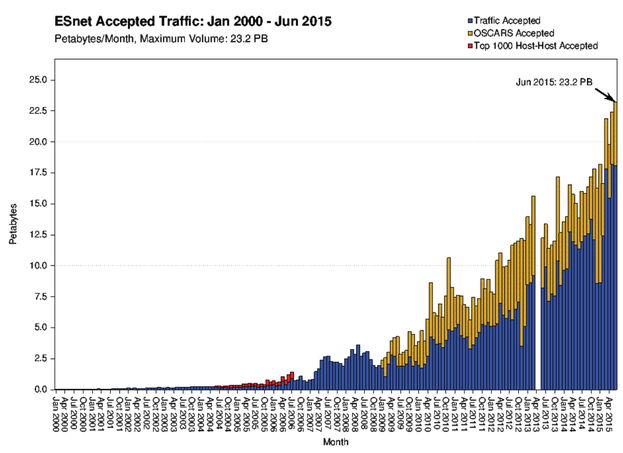


Figure 1: ESnet, the Department of Energy's designated science network, sees an exponential trend in data movement from month to month, with network traffic peaking at more than 120 gigabits per second (Gbps) at any point in a day.

## BUILDING A SUPERFACILITY

**Data volume and analysis needs for many experiments are growing faster than the experimental facility resources—smaller workstations and clusters cannot provide the resources needed for growing data-intensive experiments. To solve this problem, a handful of experimental facilities that are experiencing the greatest data growth are integrating remote supercomputing resources, advanced workflow and analysis tools, and high performance networks into their facility workflows. By tightly coupling these distributed resources, the facilities are creating the concept of a "superfacility" to solve today's most advanced science questions and problems.**

**The Advanced Light Source, a synchrotron at the Lawrence Berkeley National Laboratory, has been experimenting with a superfacility model for the last year. With the help of two supercomputing centers NERSC and OLCF, an advanced workflow and analysis tool SPOT Suite, advanced mathematical algorithms developed by CAMERA, a local high performance infrastructure using the Science DMZ, and two high performance networks ESnet and LBLnet, ALS users at beamline 7.3.3 can achieve near real-time analysis for their data, expediting scientific productivity. (See Figure 2 for an example.)**

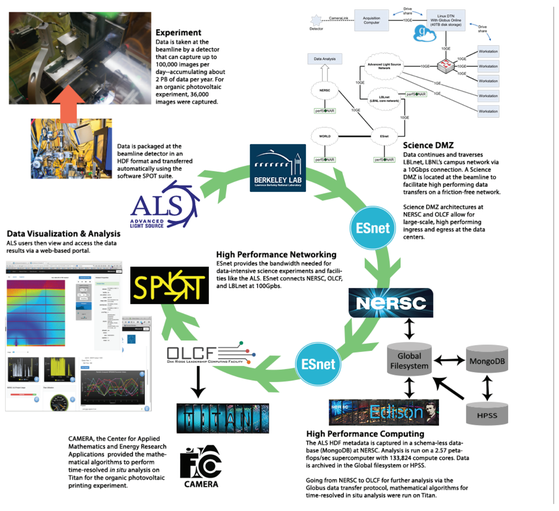


Figure 2: This workflow is an example of a superfacility used to analyze the crystallization process for organic photovoltaic printing. This superfacility represents a collaboration between various DOE-funded centers, user facilities and divisions: ALS, CAMERA, CRD (LBNL's Computational Research Division), ESnet, NERSC, and OLCF. [Images courtesy of Alex Hexemer (LBNL), Craig E. Tull (LBNL), and David Skinner (LBNL).]

## SETTING NETWORK EXPECTATIONS

**The following table shows approximately how long it should take to transfer 1 TB of data across varying network speeds:**

| | |
|---|---|
| **10 Mbps** network | **300 hrs (12.5 days)** |
| **100 Mbps** network | **30 hrs** |
| **1 Gbps** network | **3 hrs** |
| **10 Gbps** network | **20 mins** |

**These theoretical metrics describe the throughput for data transfers based on end-to-end network bandwidth noted in megabits and gigabits per second.**

ESnet
ENERGY SCIENCES NETWORK

BERKELEY LAB

## USING THE RIGHT TOOLS FOR TRANSFERRING DATASETS

**Using the right tools and protocols for data transfers can increase any science workflow performance. Some good protocols and tools for science datasets are:**

- **Globus: https://www.globus.org**
- **GridFTP from ANL has features needed to fill the network pipe**
- **bbcp: http://www.slac.stanford.edu/~abh/bbcp/**
- **lftp: http://lftp.yar.ru/**
- **axel: http://axel.alioth.debian.org/**
- **FDT (Fast Data Transfer) tool from Caltech: http://monalisa.cern.ch/FDT/**

## WHEN DATA TRANSFERS DON'T WORK WELL...

**There are many causes of poor network performance, but one of the most common causes is packet loss. Packet loss causes the most common network protocol, TCP, to adjust its rate of data transmission creating poor network performance. The graph in Figure 2 shows the large performance impact on data transfers with even a very small amount of packet loss.**

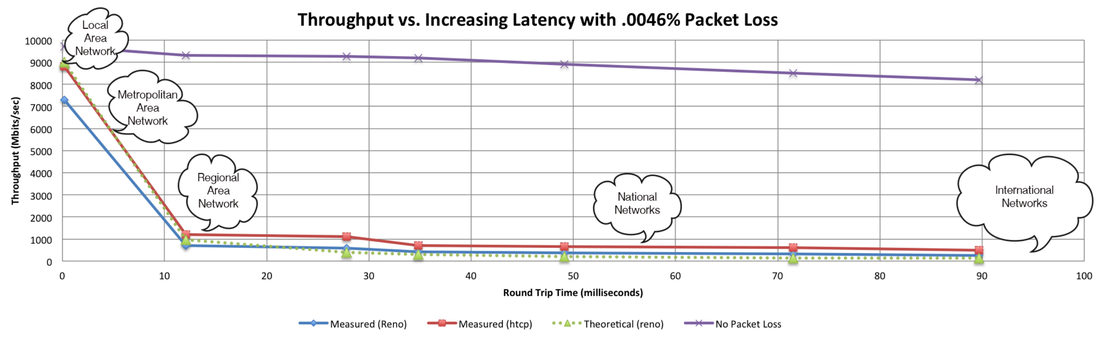


Figure 3: Graph displaying the behavior of network throughput when there is 1 in 22,000 packets dropped, or .0046% packet loss. Data traversing network paths more than 10 ms away will see extremely low data transfer performance.

# About ESnet

**ESnet is the Department of Energy's dedicated science network that provides high-performing, high-bandwidth, reliable network connections for scientists at national laboratories, universities and other research institutions, enabling them to collaborate on some of the world's most important scientific challenges including energy, climate science, and the origins of the universe. ESnet is funded by DOE Office of Science and is located within the Scientific Networking Division at Lawrence Berkeley National Laboratory.**

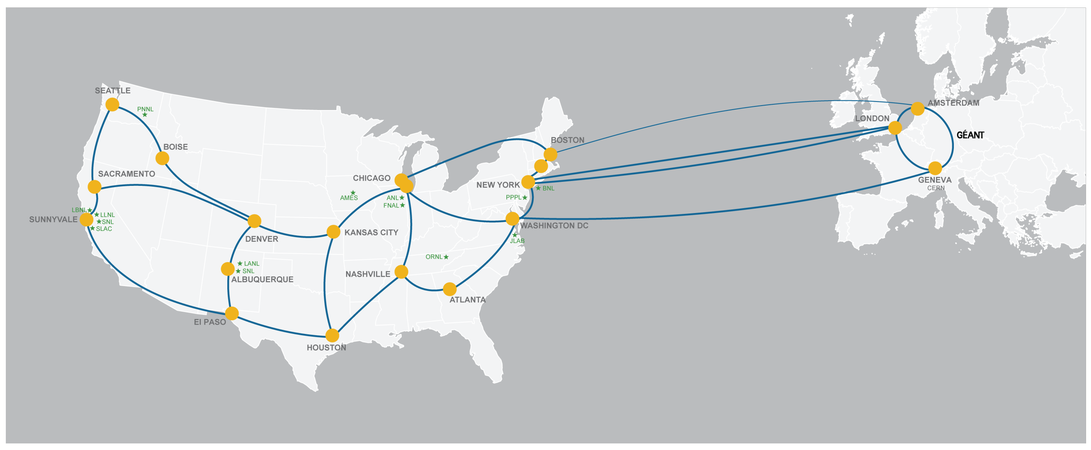


Figure 4: As of 2015, ESnet directly connects U.S. scientists to Europe, bringing the U.S. high energy physics community directly to CERN and the Large Hadron Collider with a reliable, high-bandwidth science network.

# Contacts and More Information

**For questions, email: engage@es.net**

**More information on transfer protocols and best practices visit Fasterdata Knowledge Base: fasterdata.es.net**

**If you are seeing any problems, call your local support or IT team!**

PACS 15-CS-1925