"An Integrated West Coast Science DMZ for Data-Intensive Research" … Building the Pacific Research Platform

Overview from CENIC 2015 Panel
Big Data, Big Network 3rd Workshop
Puerto Vallarta
April 23, 2015

Dr. Gregory Hidley, Technical Director
California Institute for Telecommunications and Information Technology
University of California, San Diego

CENIC 2015 Panel: Building the Pacific Research Platform

- **Presenters:**
  - **Larry Smarr, Calit2**
  - **Eli Dart, ESnet**
  - **John Haskins, UCSC**
  - **John Hess, CENIC**
  - **Erik McCroskey, UC Berkeley**
  - **Paul Murray, Stanford**
  - **Michael van Norman, UCLA**

Abstract: The Pacific Research Platform is a project to forward the work of advanced researchers and their access to technical infrastructure, with a vision of connecting all the National Science Foundation Cyberinfrastructure grants (NSF CC-NIE & CC-IIE) to research universities within the region, as well as the Department of Energy (DOE) labs and the San Diego Supercomputer Center (SDSC).

Larry Smarr, founding Director of Calit2, will present an overview of the project, followed by a panel discussion of regional inter-site connectivity challenges and opportunities.

- **LS had assistance today from:**
  - **Tom DeFanti, Research Scientist, Calit2's Qualcomm Institute, UC San Diego**
  - **John Graham, Senior Development Engineer, Calit2's Qualcomm Institute, UC San Diego**
  - **Richard Moore, Deputy Director, San Diego Supercomputer Center, UC San Diego**
  - **Phil Papadopoulos, CTO, San Diego Supercomputer Center, UC San Diego**

Vision: Creating a West Coast "Big Data Freeway" Connected by CENIC/Pacific Wave to I2 & GLIF

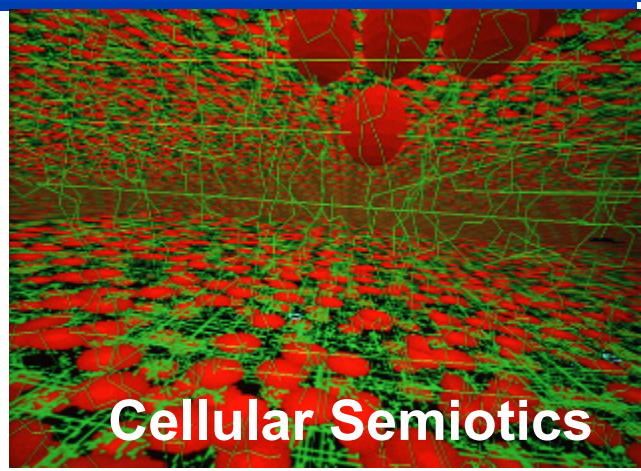
Use Lightpaths to Connect
All Data Generators and Consumers,
Creating a "Big Data" Plane
Integrated With High Performance Global Networks

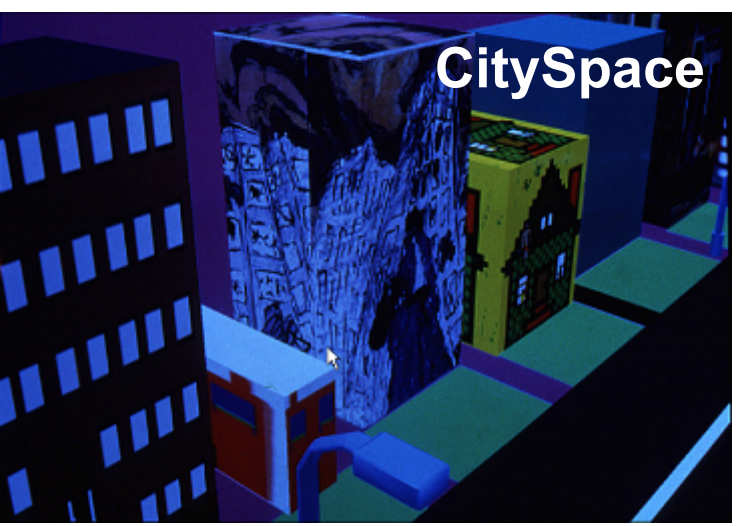
"The Bisection Bandwidth of a Cluster Interconnect, but Deployed on a 10-Campus Scale."

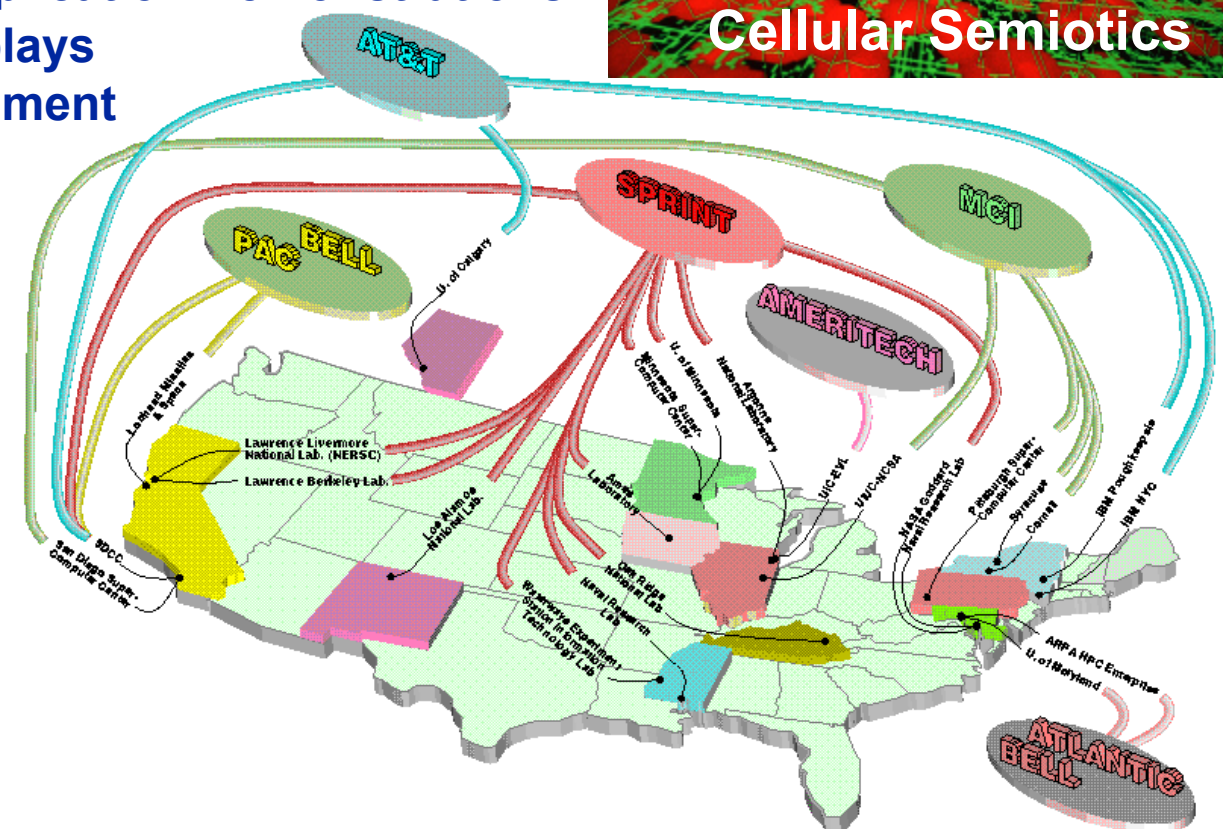
This Vision Has Been Building for Over Two Decades

I-WAY: Information Wide Area Year
Supercomputing '95

- **The First National 155 Mbps Research Network**
  - 65 Science Projects
  - Into the San Diego Convention Center
- **I-Way Featured:**
  - Networked Visualization Application Demonstrations
  - Large-Scale Immersive Displays
  - I-Soft Programming Environment







http://archive.ncsa.uiuc.edu/General/Training/SC95/GII.HPCC.html

UIC mcs NCSA

Global Connections
Between University Research Centers at 10Gbps

Maxine Brown, Tom DeFanti, Co-Chairs

iGrid 2005

THE GLOBAL LAMBDA INTEGRATED FACILITY

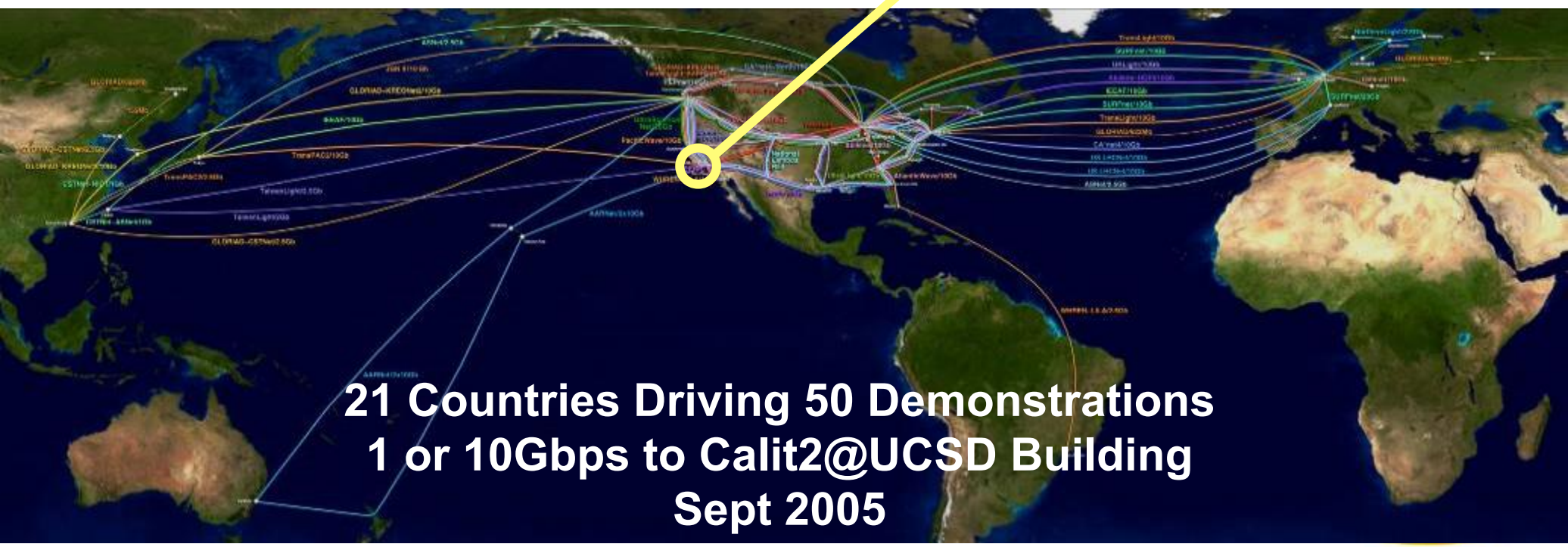
www.igrid2005.org

the UC Discovery Grant

September 26-30, 2005

Calit2 @ University of California, San Diego

California Institute for Telecommunications and Information Technology



21 Countries Driving 50 Demonstrations
1 or 10Gbps to Calit2@UCSD Building
Sept 2005

Academic Research "OptIPlatform" Cyberinfrastructure: A 10Gbps Lightpath Cloud

End User OptIPortal

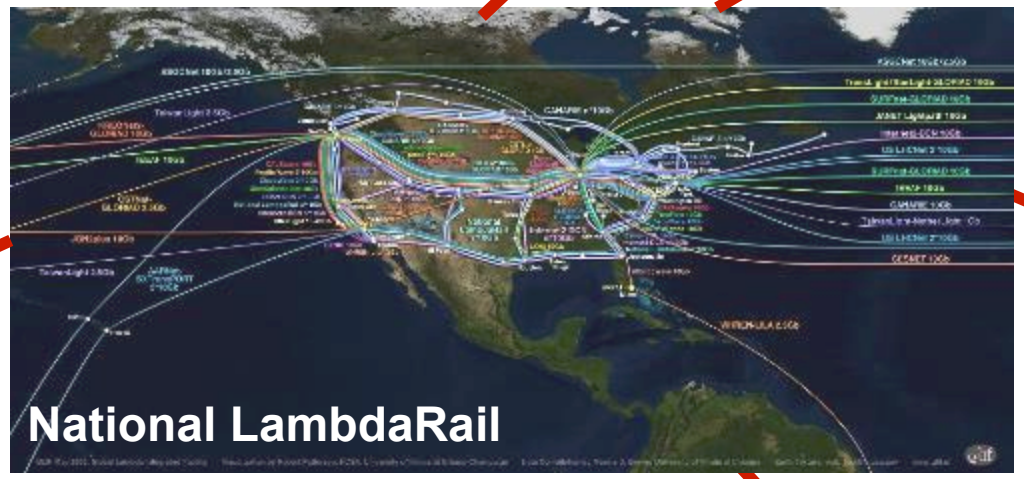
Instruments

HD/4k Telepresence

HD/4k Video Cams

HPC

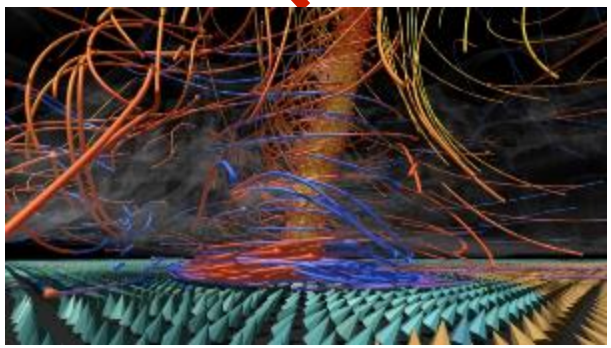
10G Lightpath



TeraGrid

Campus Optical Switch

Data Repositories & Clusters



HD/4k Video Images

LS 2009 Slide

SDSC

it²

CENIC is Rapidly Moving to Connect at 100 Gbps Across the State and Nation

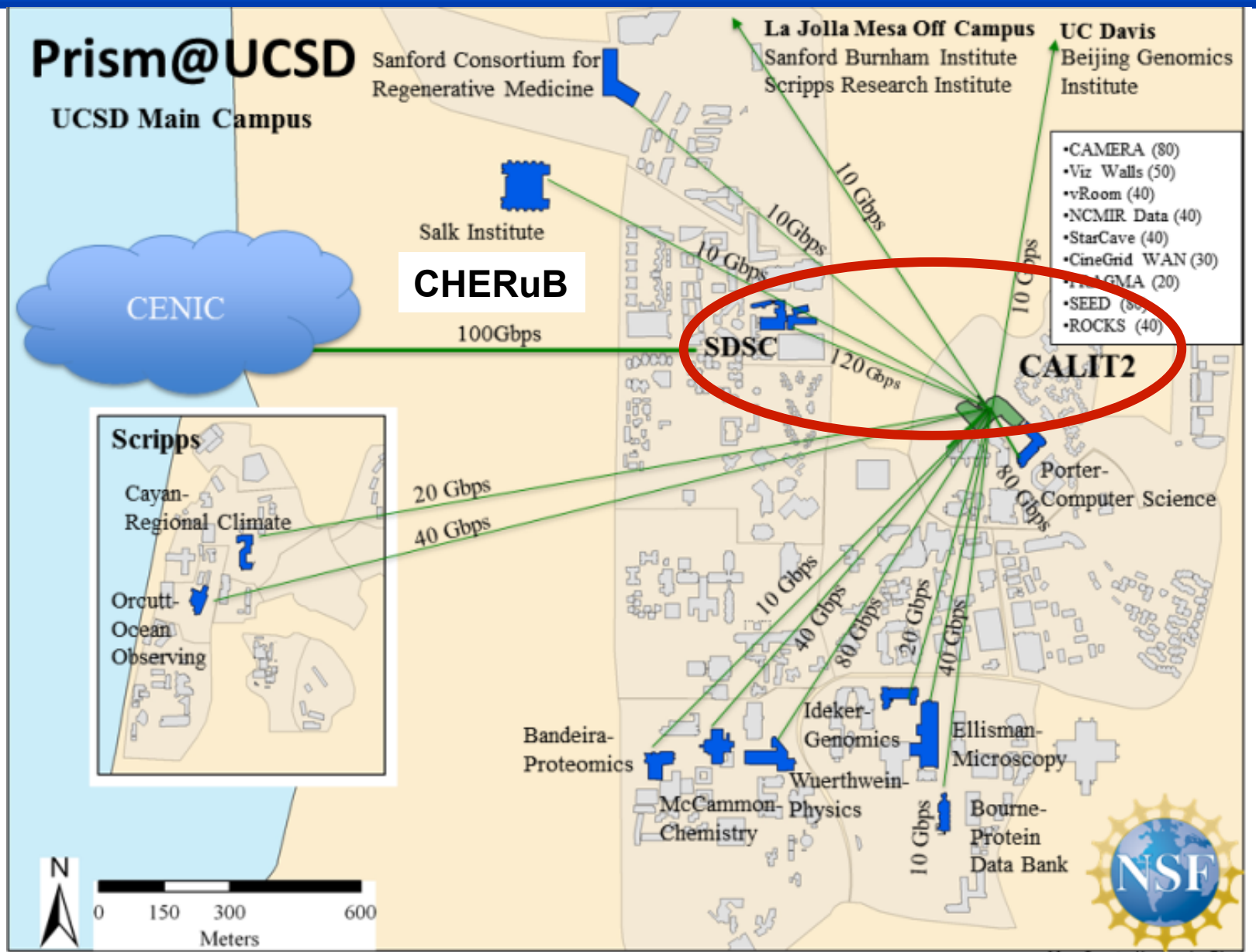
DOE

cenic

Internet2

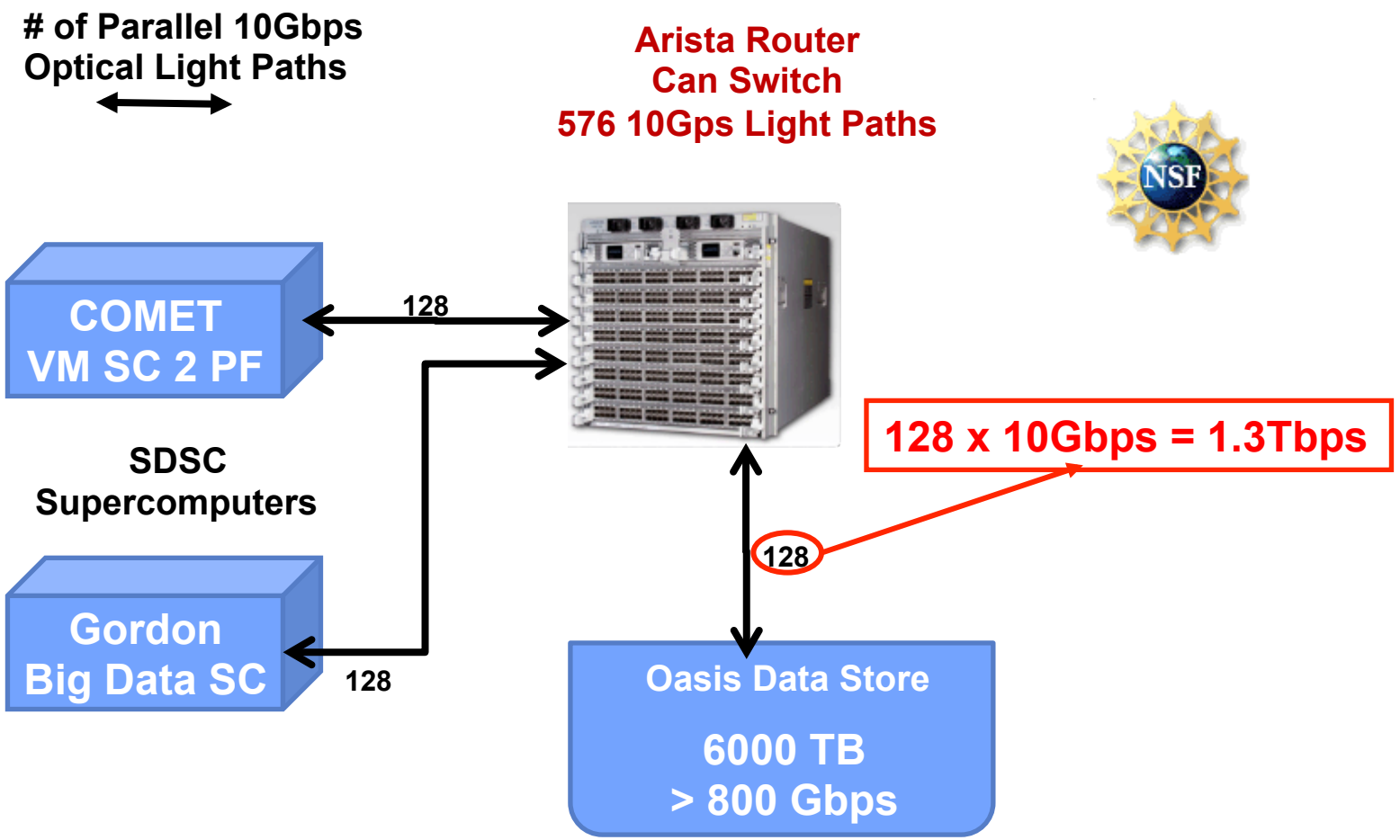
it²

Creating a "Big Data" Plane on Campus: NSF CC-NIE Funded Prism@UCSD and CHeruB



Prism@UCSD, Phil Papadopoulos, SDSC, Calit2, PI
CHERuB, Mike Norman, SDSC PI

SDSC Big Data Compute/Storage Facility - Interconnected at Over 1 Tbps



Source: Philip Papadopoulos, SDSC/Calit2

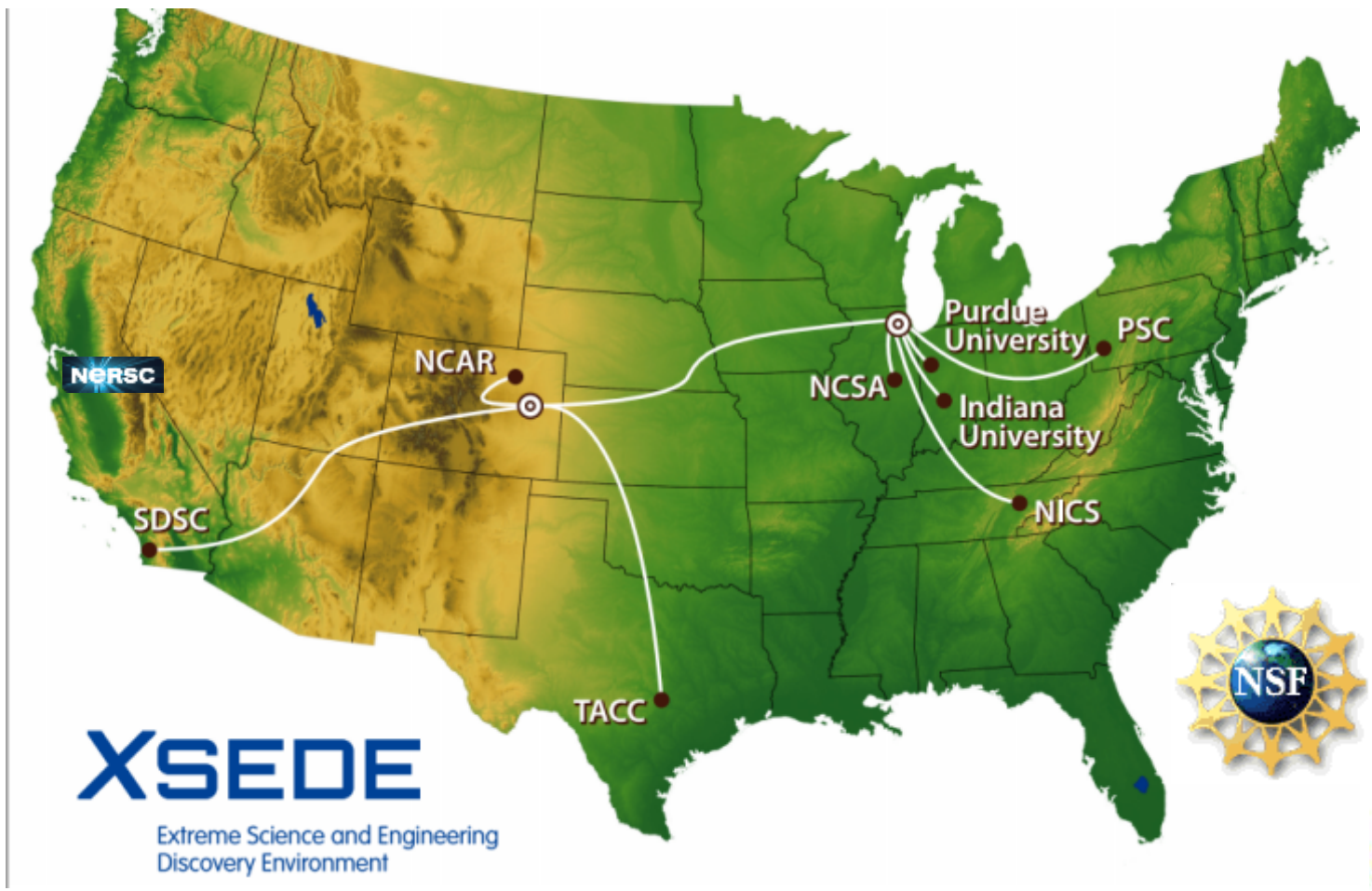
SDSC
SAN DIEGO SUPERCOMPUTER CENTER

it²

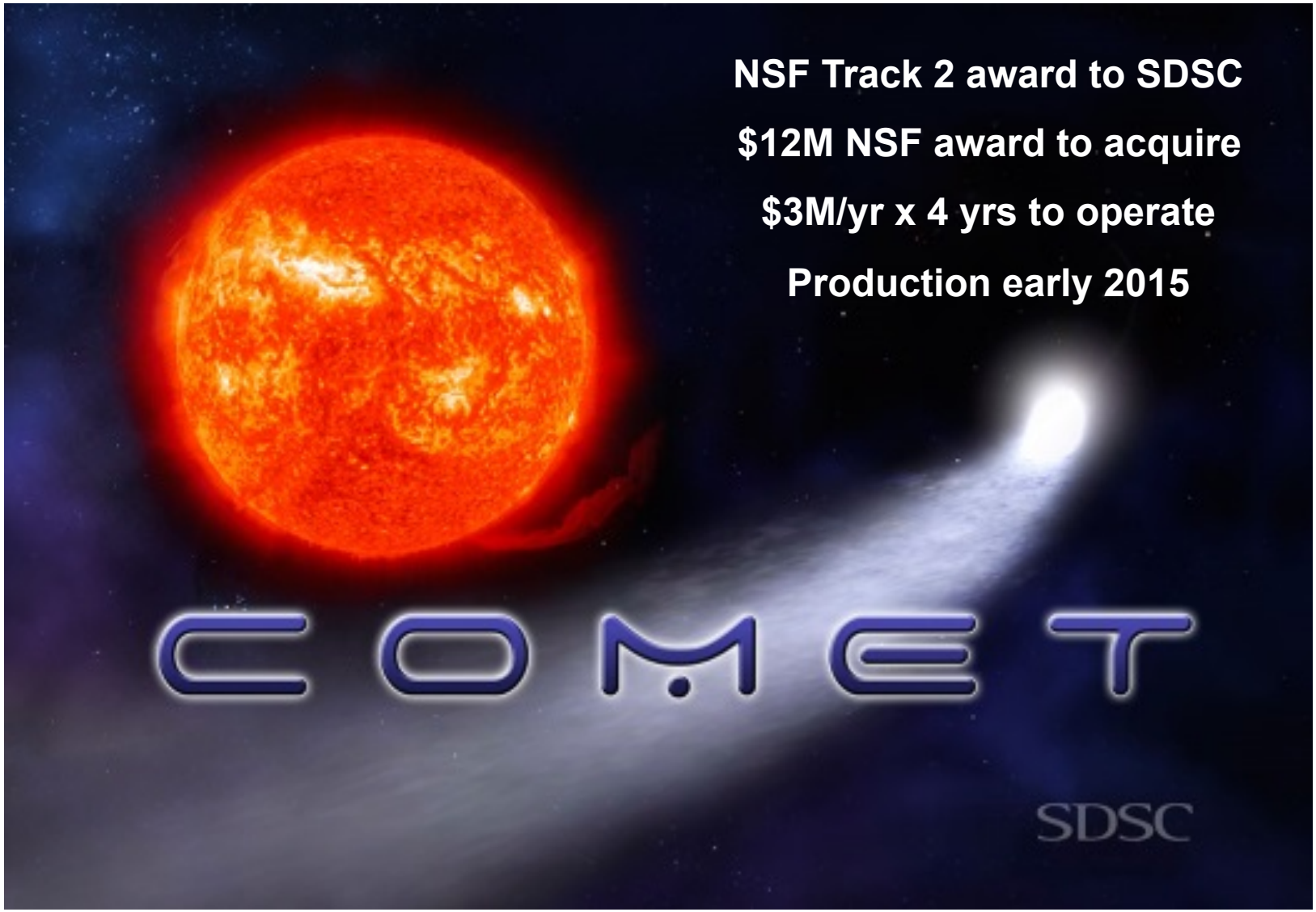
High Performance Computing and Storage Become Plug Ins to the "Big Data" Plane

NERSC

SDSC



SDSC's Comet is a ~2 PetaFLOPs System Architected for the "Long Tail of Science"



UCSD

NSF

SDSC

NERSC and ESnet
Offer High Performance Computing and Networking

EDISON ELECTRIFIES SCIENTIFIC COMPUTING

NERSC Flips Switch on New Flagship Supercomputer

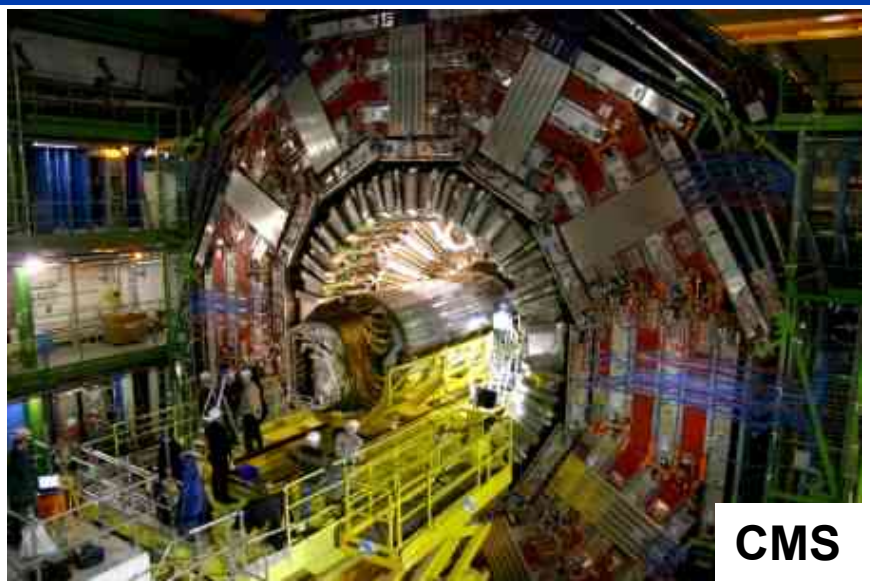
Cray XC30 2.4 Petaflops
Dedicated Feb. 5, 2014

Many Disciplines Beginning to Need Dedicated High Bandwidth on Campus

How to Utilize a CENIC 100G Campus Connection

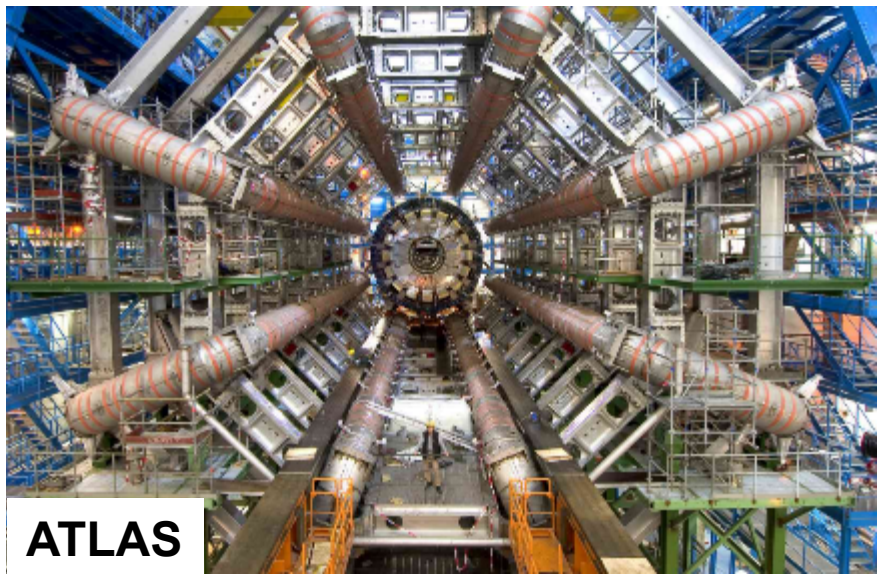
- **Remote Analysis of Large Data Sets**
  - **Particle Physics**
- **Connection to Remote Campus Compute & Storage Clusters**
  - **Microscopy and Next Gen Sequencers**
- **Providing Remote Access to Campus Data Repositories**
  - **Protein Data Bank and Mass Spectrometry**
- **Enabling Remote Collaborations**
  - **National and International**

Particle Physics: Creating a 10-100 Gbps LambdaGrid to Support LHC Researchers

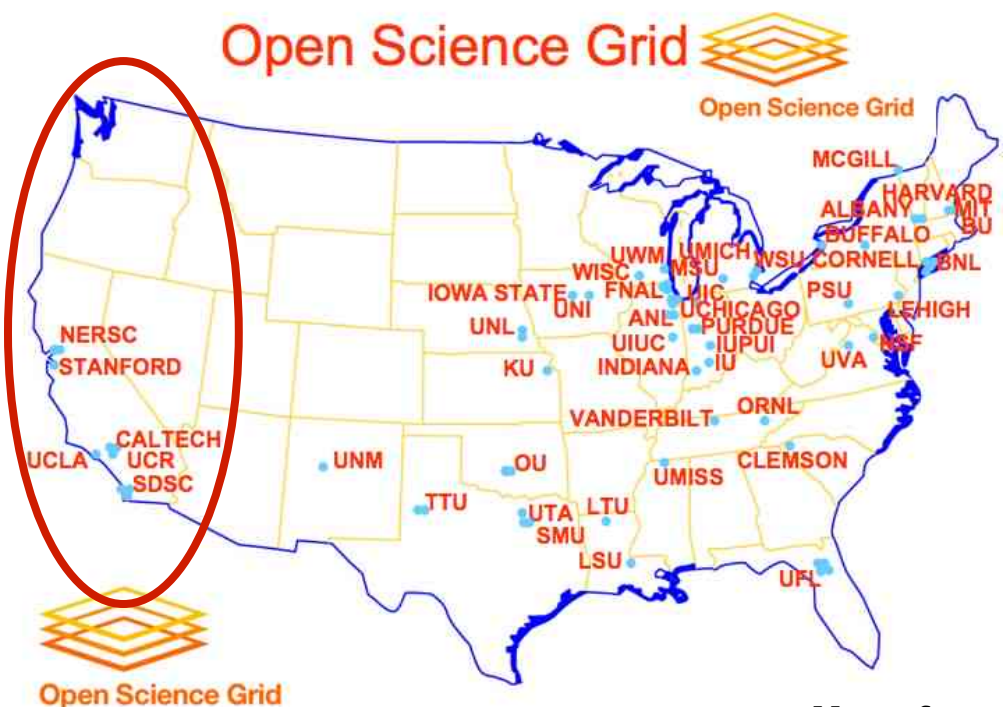


CMS

LHC Data Generated by CMS & ATLAS Detectors Analyzed on OSG

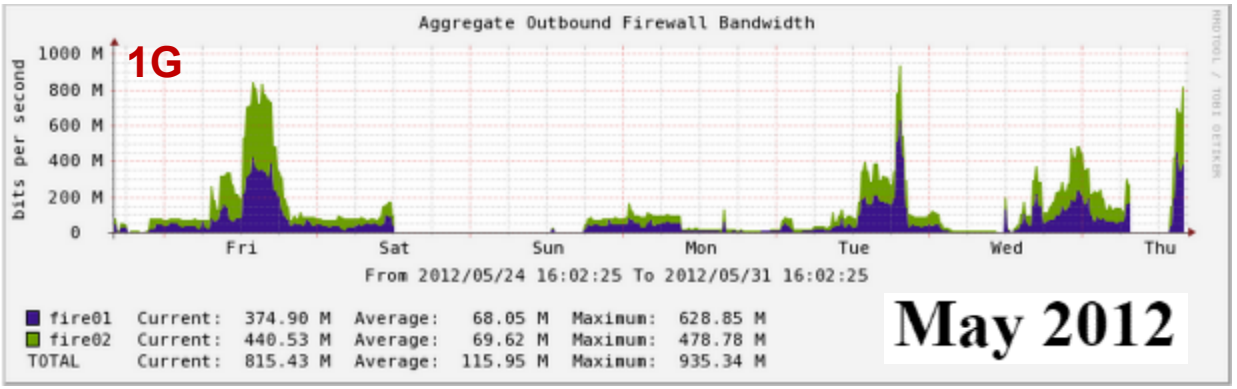


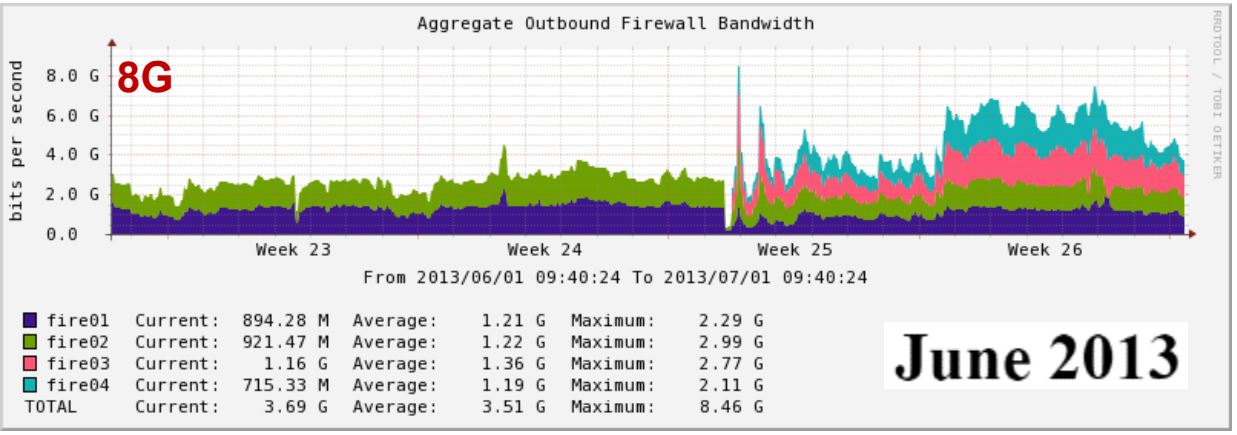
ATLAS

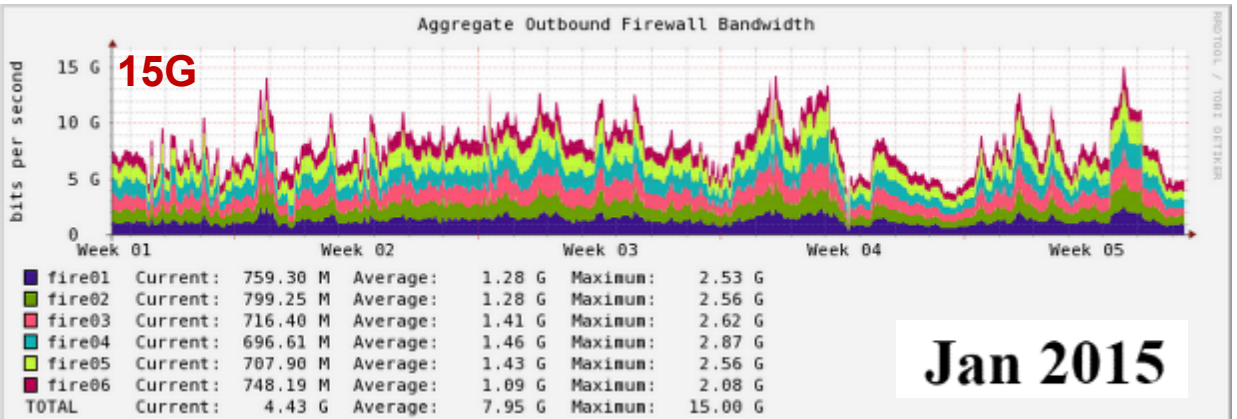


Maps from www.uslhc.us

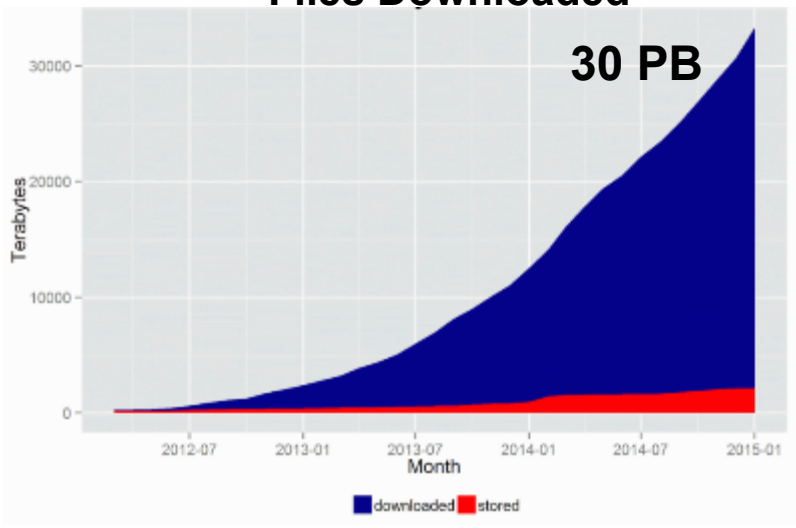
Cancer Genomics Hub (UCSC) is Housed in SDSC CoLo: Large Data Flows to End Users







Cumulative TBs of CGH Files Downloaded



Data Source: David Haussler, Brad Smith, UCSC

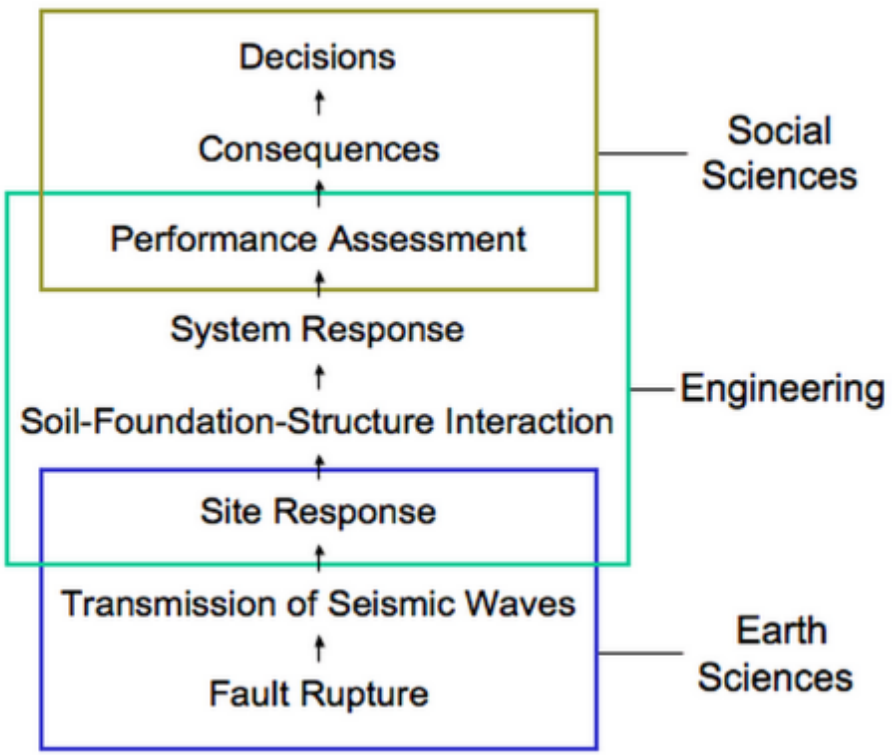
Earth Sciences: Pacific Earthquake Engineering Research Center

PEER

PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER

- University of California, Berkeley - Lead Institution
- California Institute of Technology
- Oregon State University
- Stanford University
- University of California, Davis
- University of California, Irvine
- University of California, Los Angeles
- University of California, San Diego
- University of Southern California
- University of Washington

Enabling Real-Time Coupling Between Shake Tables and Supercomputer Simulations



Automated Telescope Surveys Are Creating Huge Datasets

INTERMEDIATE PALOMAR TRANSIENT FACTORY

THE DARK ENERGY SURVEY

300 images per night.
100MB per raw image

30GB per night

120GB per night

When processed at NERSC Increased by 4x

250 images per night.
530MB per raw image

150 GB per night

600GB per night

Source: Peter Nugent, Division Deputy for Scientific Engagement, LBL
Professor of Astronomy, UC Berkeley

Using Supernetworks to Couple End User to Remote Supercomputers and Visualization Servers

Source: Mike Norman, Rick Wagner, SDSC

Demoed SC09

Argonne NL

DOE Eureka
100 Dual Quad Core Xeon Servers
200 NVIDIA Quadro FX GPUs in 50 Quadro Plex S4 1U enclosures
3.2 TB RAM

Real-Time Interactive Volume Rendering Streamed from ANL to SDSC

ESnet
10 Gb/s fiber optic network

SDSC

Calit2/SDSC OptIPortal1
20 30" (2560 x 1600 pixel) LCD panels
10 NVIDIA Quadro FX 4600 graphics cards > 80 megapixels
10 Gb/s network throughout

NICS ORNL

NSF TeraGrid Kraken
Cray XT5
8,256 Compute Nodes
99,072 Compute Cores
129 TB RAM

SDSC

*ANL * Calit2 * LBNL * NICS * ORNL * SDSC
www.calit2.net/newsroom/release.php?id=1624

Collaboration Between EVL's CAVE2 and Calit2's VROOM Over 10Gb Wavelength

Source: NTT Sponsored ON*VECTOR Workshop at Calit2 March 6, 2013

DOE Esnet's Science DMZ: A Scalable Network Design Model for Optimizing Science Data Transfers

- **A Science DMZ integrates 4 key concepts into a unified whole:**
  - A network architecture designed for high-performance applications, with the science network distinct from the general-purpose network
  - The use of dedicated systems for data transfer
  - Performance measurement and network testing systems that are regularly used to characterize and troubleshoot the network
  - Security policies and enforcement mechanisms that are tailored for high performance science environments

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

NSF Funding Has Enabled Science DMZs at Over 100 U.S. Campuses

- **2011 ACCI Strategic Recommendation to the NSF #3:**
  - NSF should create a new program funding high-speed (currently 10 Gbps) connections from campuses to the nearest landing point for a national network backbone. The design of these connections must include support for dynamic network provisioning services and must be engineered to support rapid movement of large scientific data sets."
  - - pg. 6, NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging, Final Report, March 2011
    - www.nsf.gov/od/oci/taskforces/TaskForceReport_CampusBridging.pdf
  - Led to Office of Cyberinfrastructure CC-NIE RFP March 1, 2012
- **NSF's Campus Cyberinfrastructure – Network Infrastructure & Engineering (CC-NIE) Program**
  - >130 Grants Awarded So Far (New Solicitation Open)
    - Roughly $500k per Campus

Next Logical Step-Interconnect Campus Science DMZs

Science DMZ Data Transfer Nodes Can Be Inexpensive PCs Optimized for Big Data

- **FIONA – Flash I/O Node Appliance**
  - Combination of Desktop and Server Building Blocks
  - US$5K - US$7K
  - Desktop Flash up to 16TB
  - RAID Drives up to 48TB
  - 10GbE/40GbE Adapter
  - Tested speed 40Gbs
  - Developed Under UCSD CC-NIE Prism Award by UCSD's
    - Phil Papadopoulos
    - Tom DeFanti
    - Joe Keefe

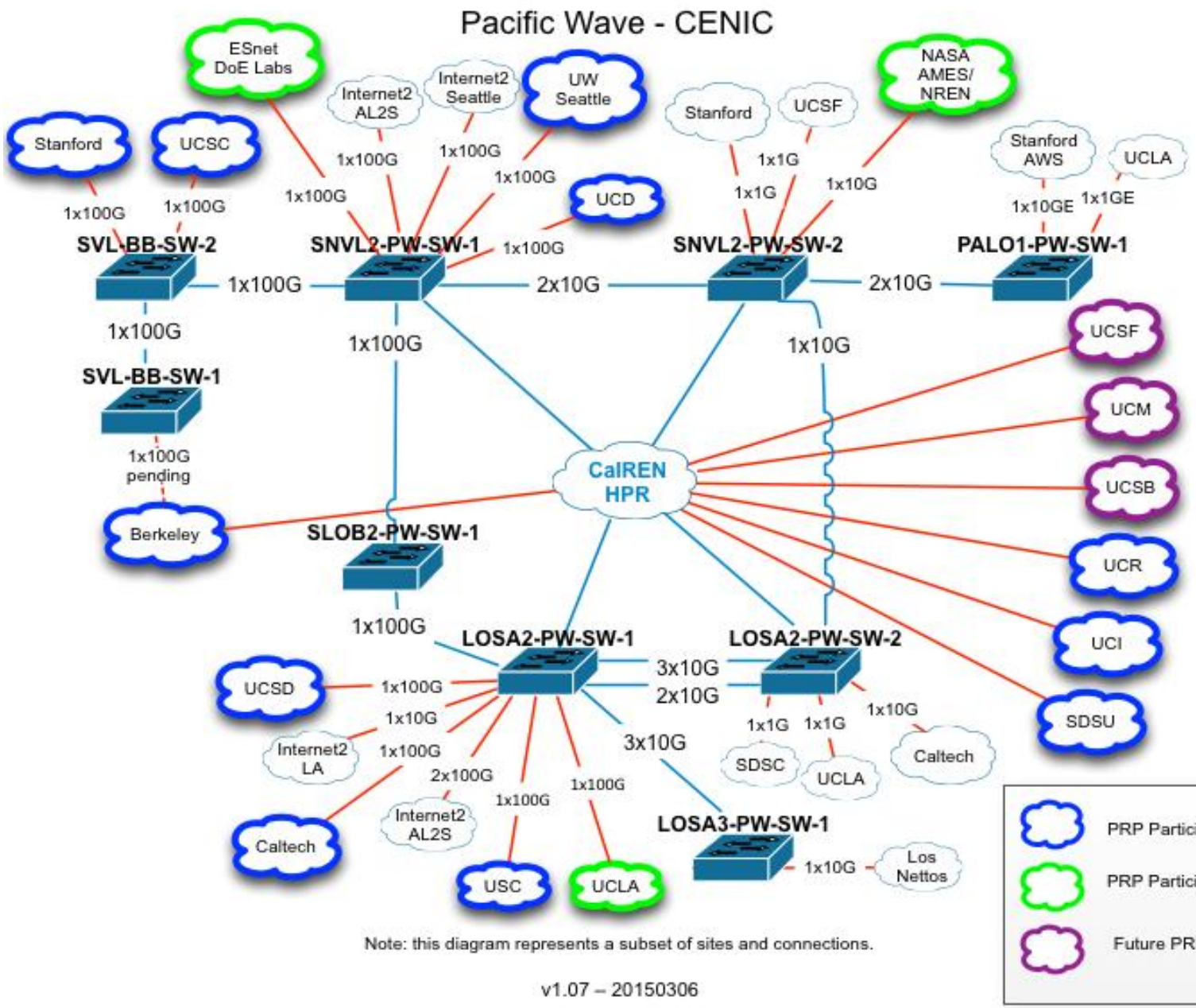
SDSC

For More on Science DMZ DTNs See:
https://fasterdata.es.net/science-dmz/DTN/

Audacious Goal: Build a West Coast Science DMZ

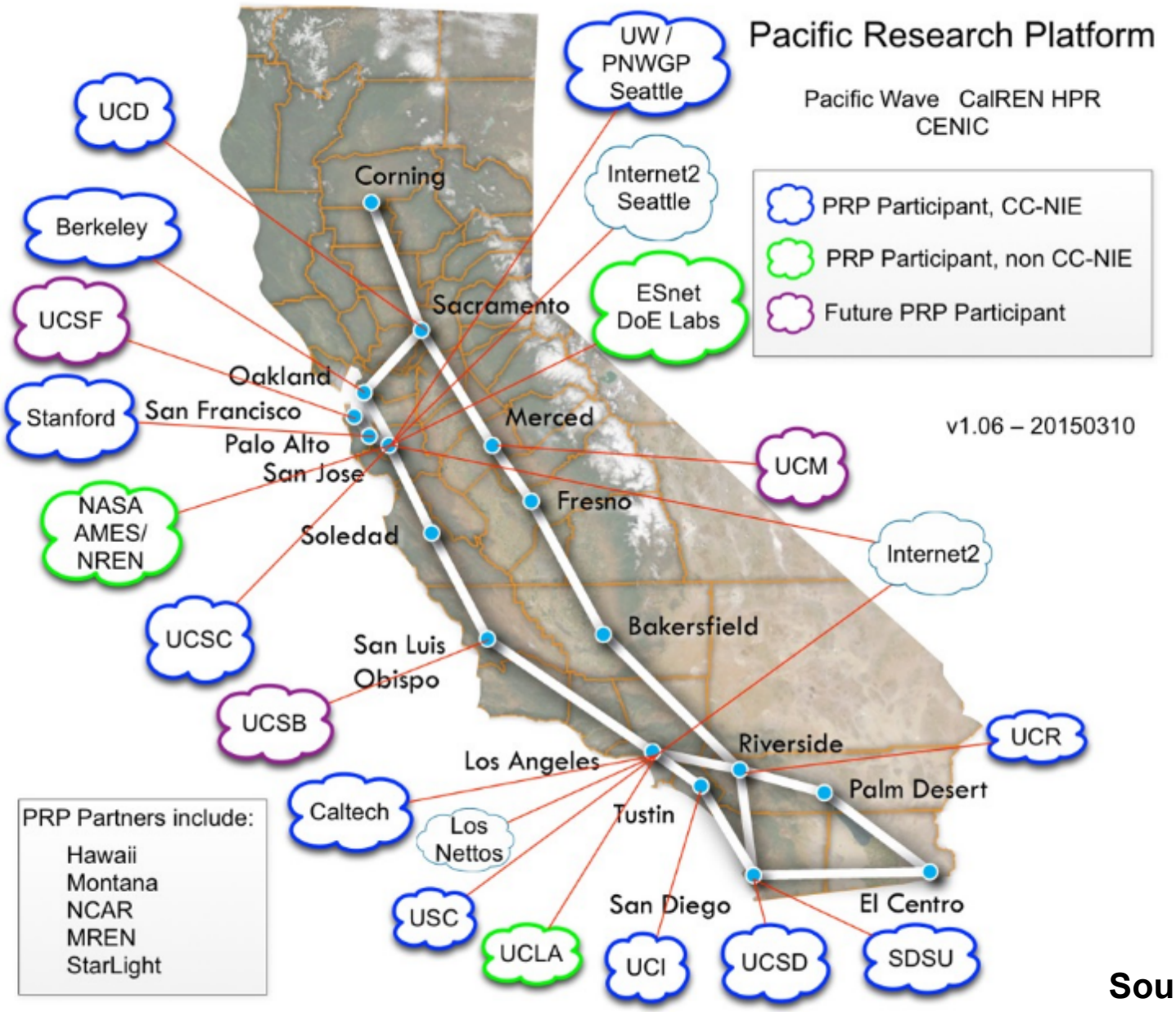
- **Why Did We Think This Was Possible?**
  - **Esnet Designed Science DMZs to be:**
    - **Scalable and incrementally deployable,**
    - **Easily adaptable to incorporate emerging technologies such as:**
      - 100 Gigabit Ethernet services,
      - virtual circuits, and
      - software-defined networking capabilities
  - **Many Campuses on the West Coast Created Science DMZs**
  - **CENIC/Pacific Wave is Upgrading to 100G Services**
  - **UCSD's FIONAs Are Rapidly Deployable Inexpensive DTNs**
- **So Can We Use CENIC/PW to Interconnect Many Science DMZs?**

CENIC/Pacific Wave is the Optical Backplane of the Pacific Research Platform (PRP)



NSF Has Invested Over $9M in CC-NIE Campus Awards

The Pacific Wave Platform Creates a Regional Science DMZ



Note: this diagram represents a subset of sites and connections.

Source: John Hess, CENIC

Thanks to:

Caltech
CENIC / Pacific Wave
ESnet / LBNL
San Diego State University
SDSC
Stanford University
University of Washington
USC
UC Berkeley
UC Davis
UC Irvine
UC Los Angeles
UC Riverside
UC San Diego
UC Santa Cruz

Pacific Research Platform – Panel Discussion

March 9, 2015

CENIC 2015

Pacific Research Platform Strategic Arc

- High performance network backplane for data-intensive science
  - This is qualitatively different than the commodity Internet
  - High performance data movement provides capabilities that are otherwise unavailable to scientists
  - Linking the Science DMZs across the West Coast is building something new
  - This capability is extensible, both regionally and nationally
- Goal - scientists at CENIC institutions can get the data they need, where they need it, when they need it

What did we do?

Concentrated on the regional aspects of the problem. There are lots of parts to the research data movement problem. This experiment mostly looked at the inter-campus piece.

If it looks a bit rough, this has all happened in about 10 weeks of work.

Collaborated among lots of network and HPC staff at lots of sites to

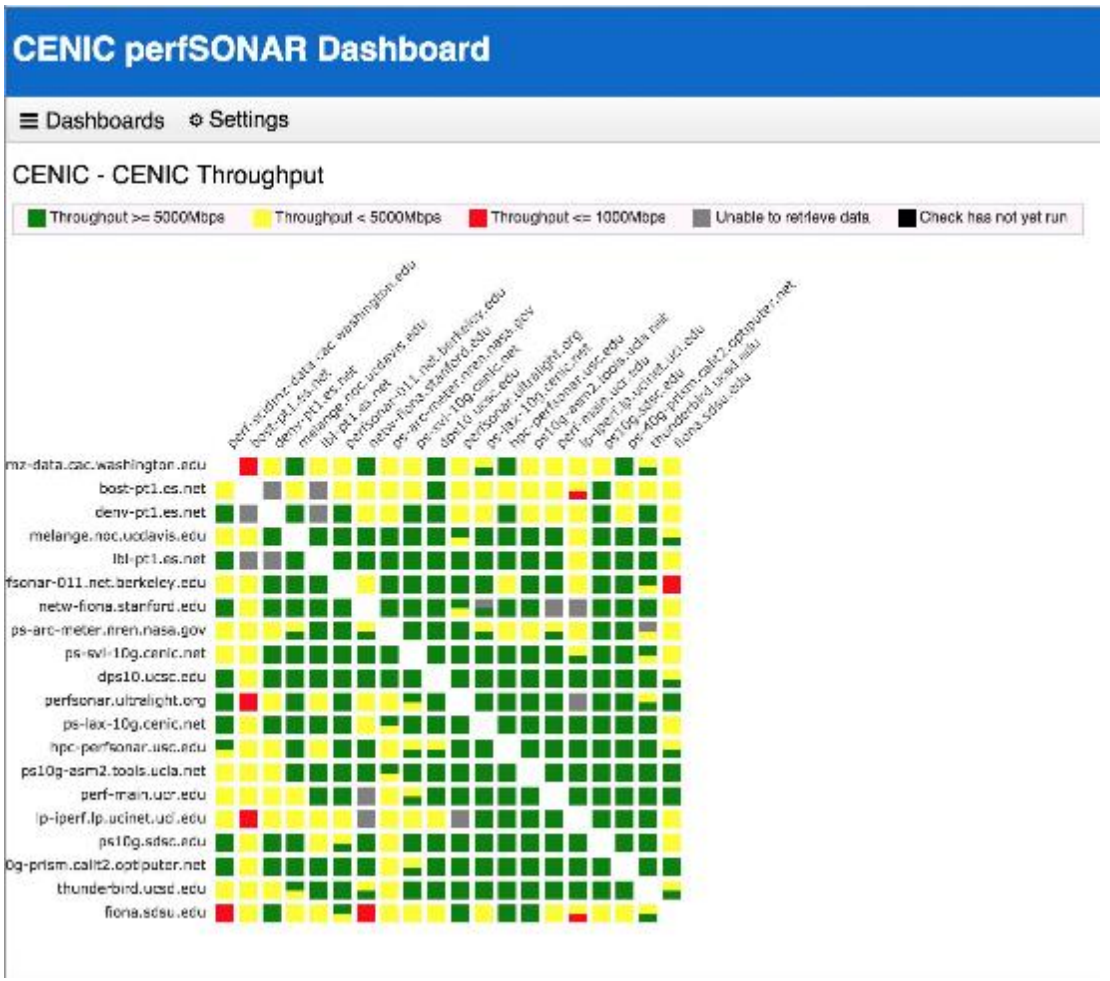
- *Build mesh of perfSONAR instances.*
- *Implement MaDDash -- Measurement and Debugging Dashboard.*
- *Deploy Data Transfer Nodes (DTN)*
- *Perform GridFTP file transfers to quantify throughput of reference data sets.*

What did we do?

- *Constructed a temporary network using 100G links to demonstrate the potential of networks with burst capacity greater than that of a single DTN.*
- *Partial ad-hoc BGP peering mesh between some test points to make use of 100G paths.*
- *Identified some specific optimizations needed.*
- *Fixed a few problems in pursuit of gathering illustrative data for this preso.*
- *Identified anomalies for further investigation.*

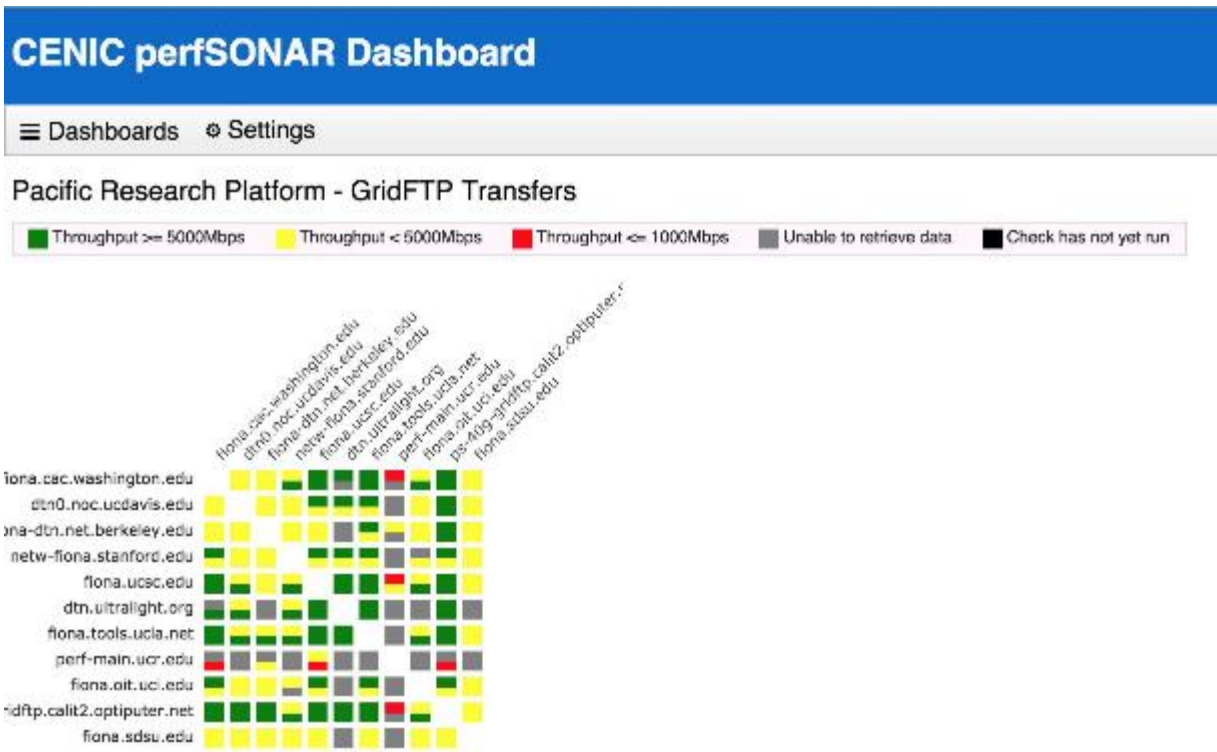
MaDDash of perfSONAR throughput and loss

- Test nodes ordered by geographic latitude
- Performance for nodes that are close is better than for nodes that are far away
- Network problems that manifest over a distance may not manifest locally

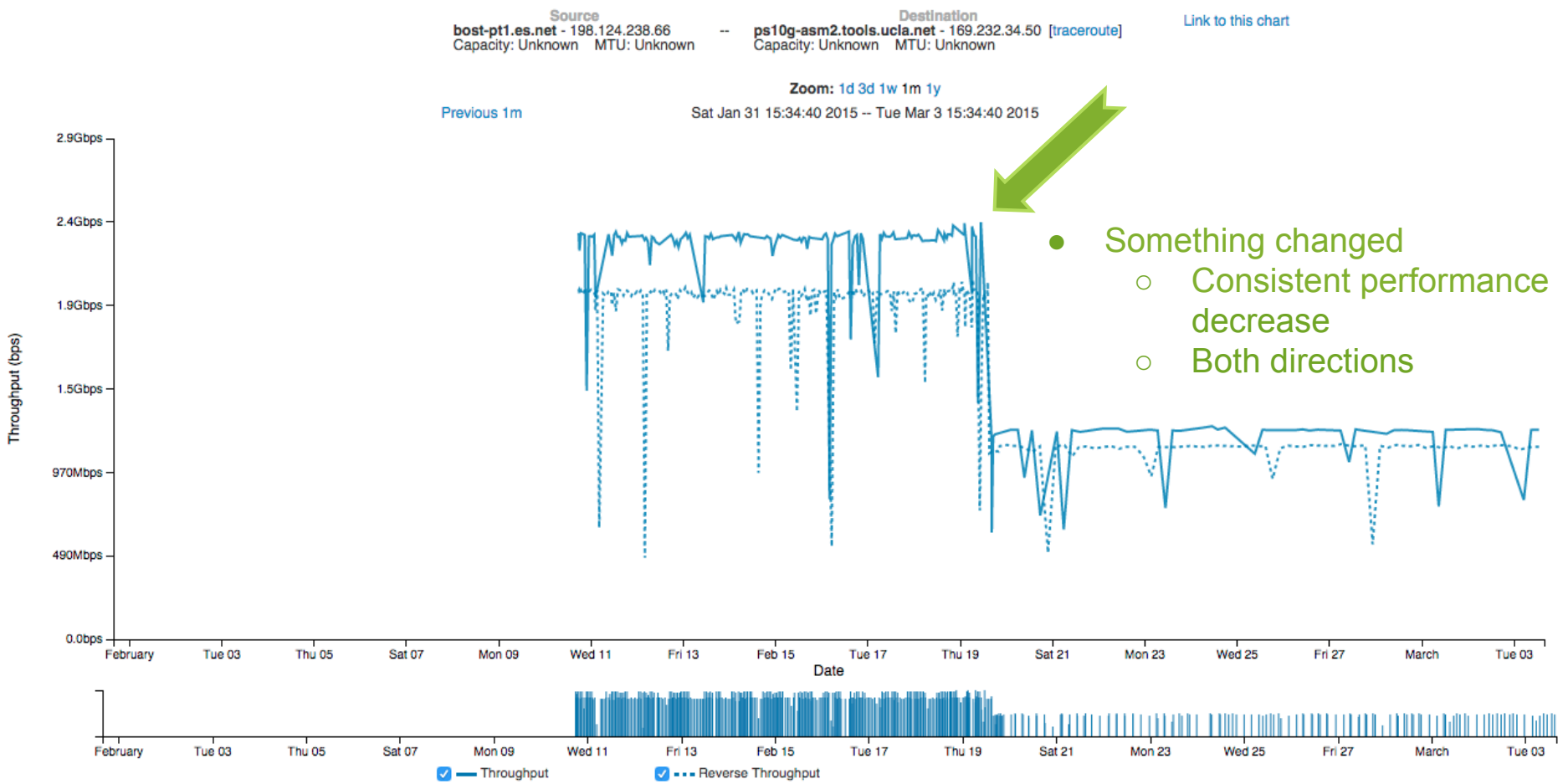


MaDDash of GridFTP transfers

- DTNs loaded with Globus Connect Server suite to obtain GridFTP tools.
- cron-scheduled transfers using globus-url-copy.
- ESnet-contributed script parses GridFTP transfer log and loads results in an esmond measurement archive.



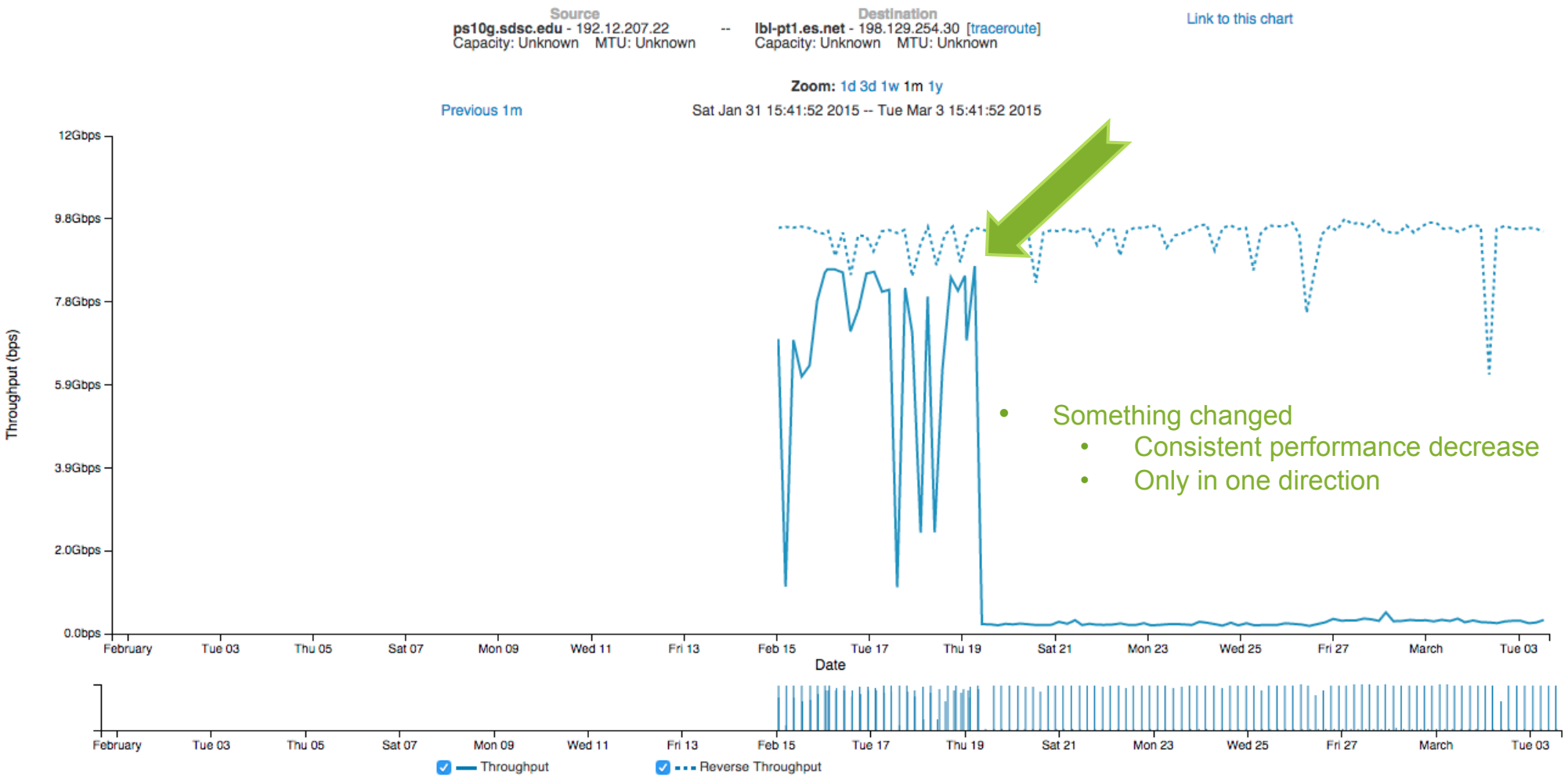
bost-pt1.es.net -- ps10g-asm2.tools.ucla.net



- Something changed
  - Consistent performance decrease
  - Both directions

CENIC

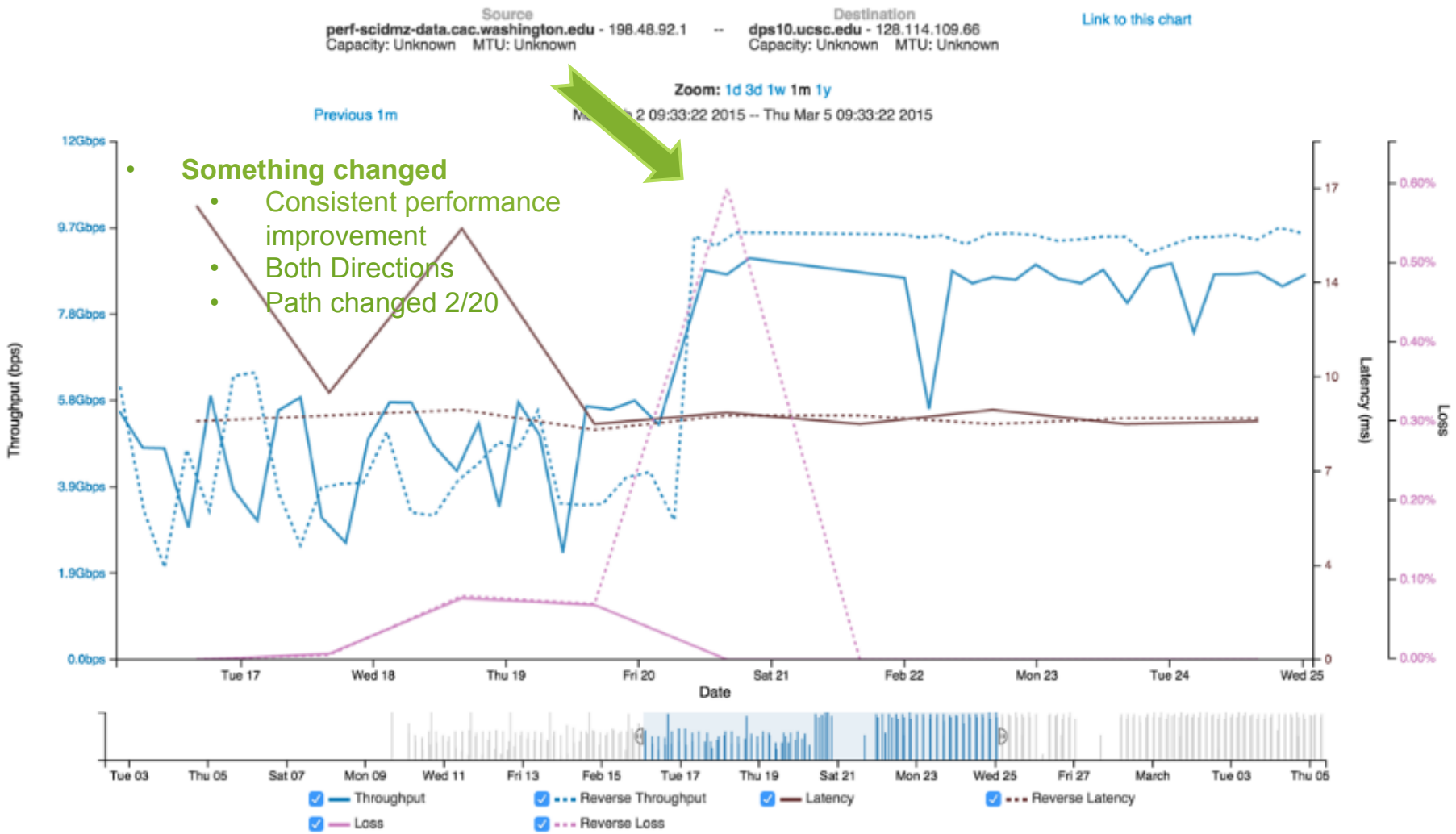
ps10g.sdsc.edu—lbl-pt1.es.net



- Something changed
  - Consistent performance decrease
  - Only in one direction

CENIC

perf-scidmz.cac.washington.edu -- dps10.ucsc.edu



CENIC

What did we learn

Coordinating this effort was quite a bit of work, and there's still a lot to do.

Traffic doesn't always go where you think it does.

Familiarity with measurement toolkits such as perfSONAR (bwctl / iperf3, owamp) and MaDDash.

We need people's time to continue the effort.

Next Steps or Near Future

- Future of CENIC High Performance Research Network (HPR)
  - Migrate to 100 Gbps Layer3 on HPR.
  - Evolve into persistent infrastructure
- Enhance and maintain perfSONAR test infrastructure across R&E sites.
- Engagement with scientists to map their research to the Pacific Research Platform

Links

- ESnet fasterdata knowledge base
  - http://fasterdata.es.net/
- Science DMZ paper
  - http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf
- Science DMZ email list
  - *To subscribe, send email to sympa@lists.lbl.gov*
  - *subject "subscribe esnet-sciencedmz"*
- perfSONAR
  - http://fasterdata.es.net/performance-testing/perfsonar/
  - http://www.perfsonar.net
- perfSONAR dashboard
  - http://ps-dashboard.es.net/

MaDDash the movie

CENIC perfSONAR Dashboard

☰ Dashboards ⚙ Settings

Last page refresh time: February 22, 2015 23:10:56 PM PST

CENIC - CENIC Throughput

CENIC