"An Integrated Science Cyberinfrastructure for Data-Intensive Research"

Panel

CISCO Executive Symposium

San Diego, CA

June 9, 2015

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

The Data-Intensive Discovery Era Requires High Performance Cyberinfrastructure

- Growth of Digital Data is Exponential
  - "Data Tsunami"
- Driven by Advances in Digital Detectors, Computing, Networking, & Storage Technologies
- Shared Internet Optimized for Megabyte-Size Objects
- Need Dedicated Photonic Cyberinfrastructure for Gigabyte/Terabyte Data Objects
- Finding Patterns in the Data is the New Imperative
  - Data-Driven Applications
  - Data Mining
  - Visual Analytics
  - Data Analysis Workflows



SDSC

Source: SDSC

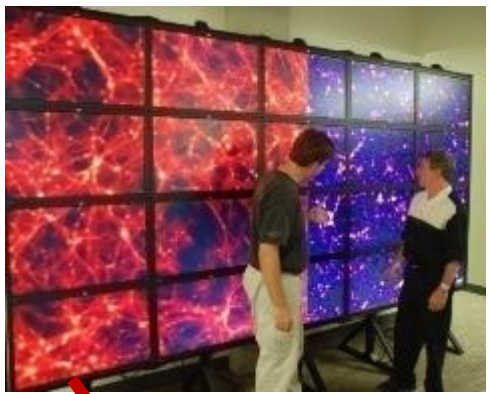
it²

Vision: Creating a "Big Data Freeway"

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

This Vision Has Been Building for Over Two Decades

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



End User OptIPortal



Instruments



HD/4k Telepresence

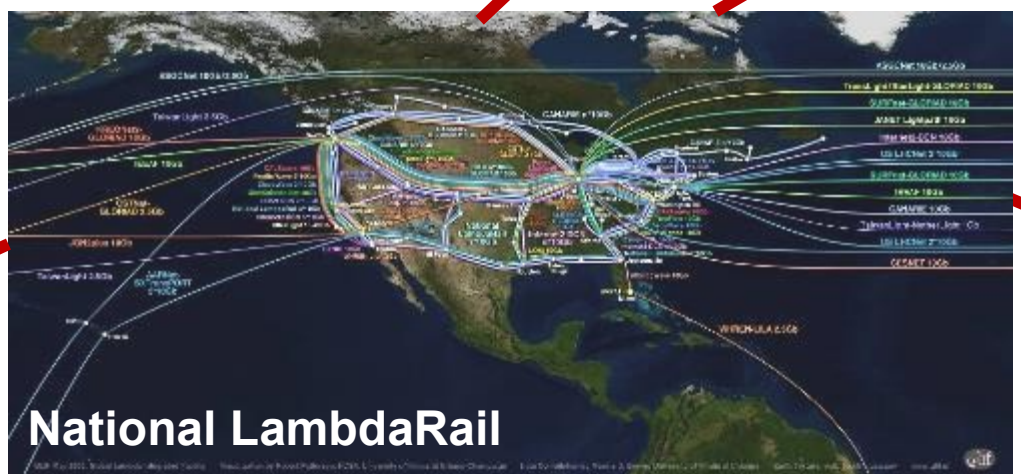


HD/4k Video Cams

HPC



10G Lightpath



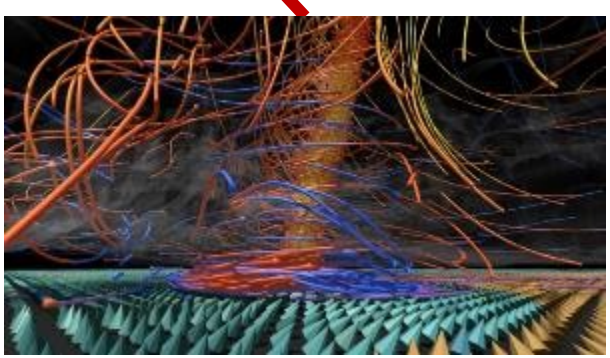
TeraGrid

OptIPuter

Campus Optical Switch



Data Repositories & Clusters

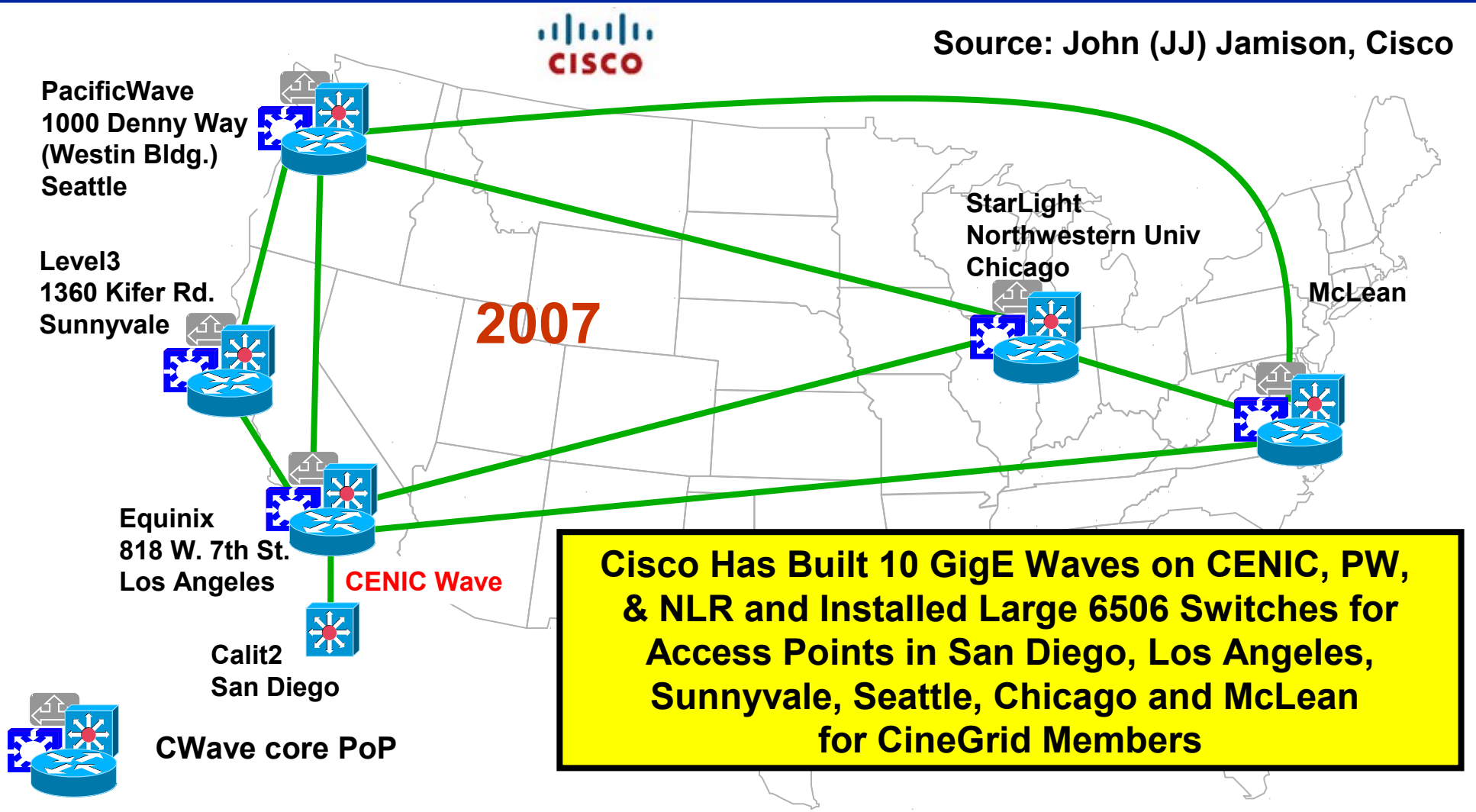


HD/4k Video Images

SDSC

it²

Cisco CWave for CineGrid: A New Cyberinfrastructure for High Resolution Media Streaming*



*CTWatch QUARTERLY May 2007

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

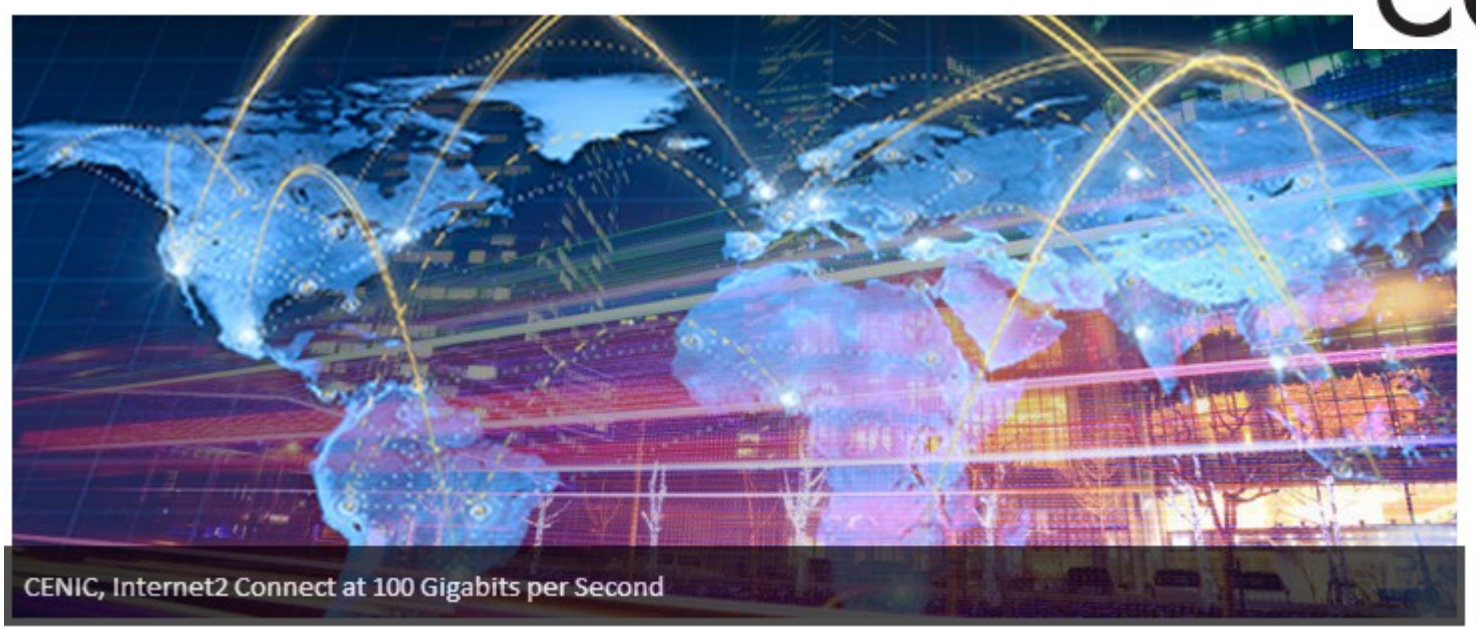
DOE



CENIC and ESnet Connected at 100 Gigabits per Second

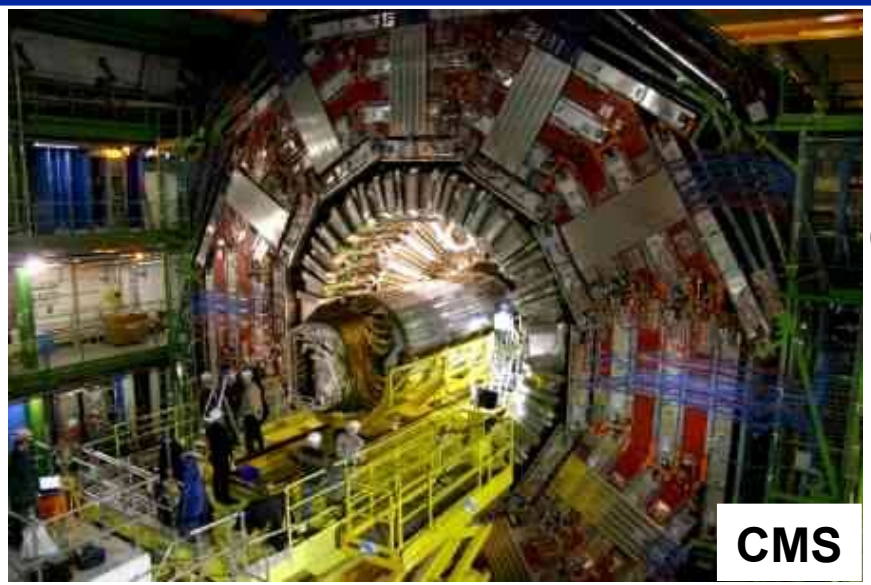
cenic

Internet2

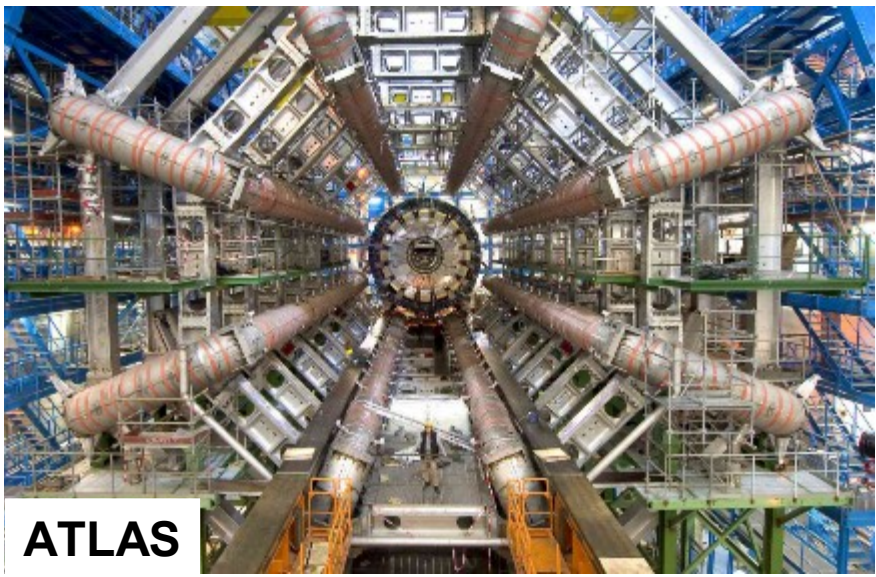

CENIC, Internet2 Connect at 100 Gigabits per Second

it²

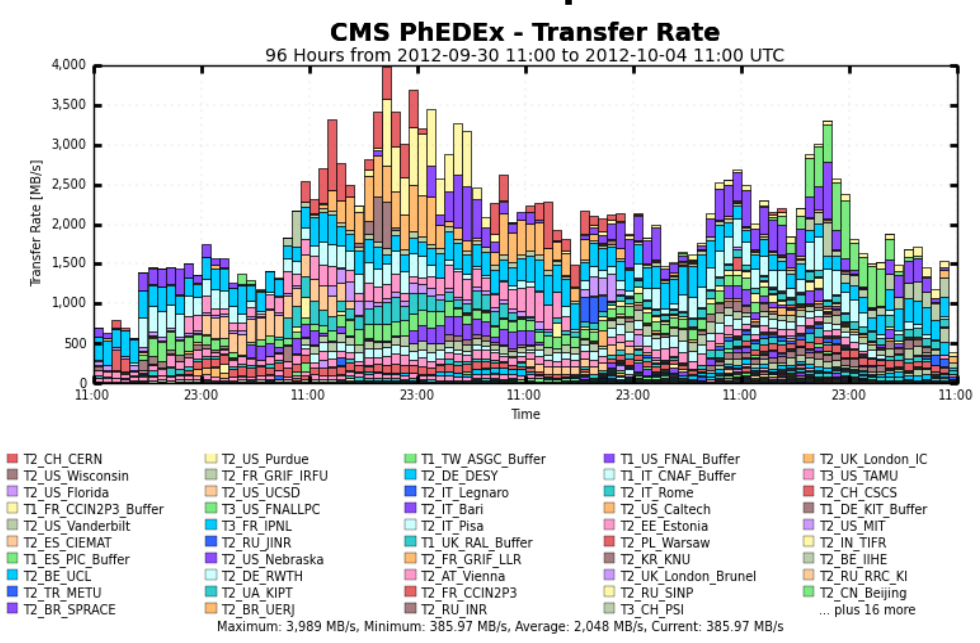
Particle Physics: Creating a 10-100 Gbps LambdaGrid to Support LHC Researchers

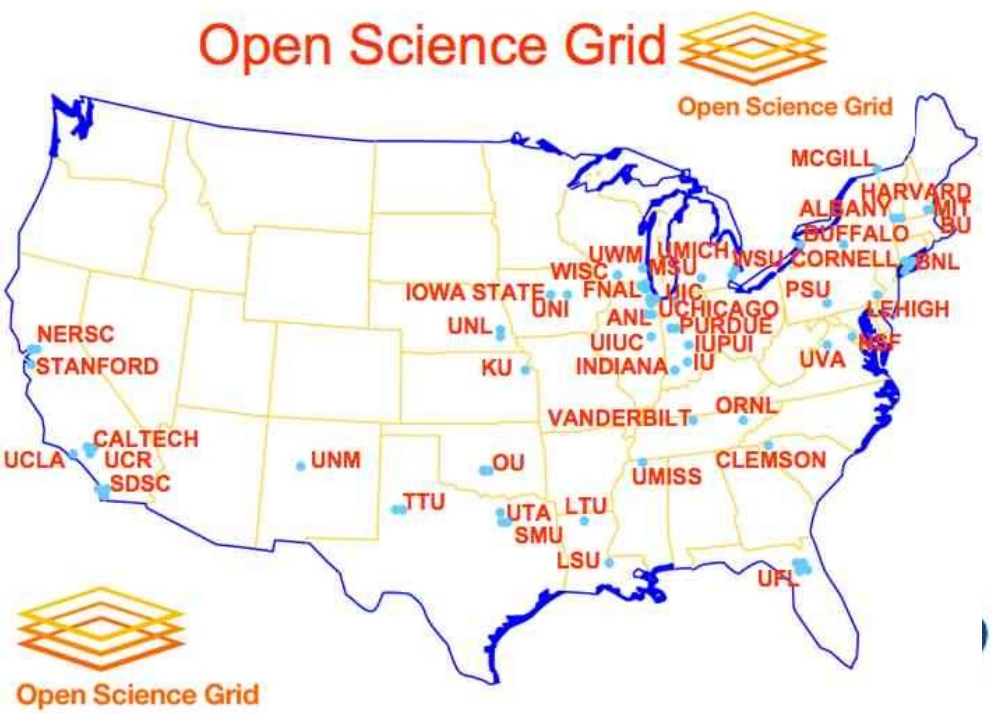


LHC Data Generated by CMS & ATLAS Detectors Analyzed on OSG

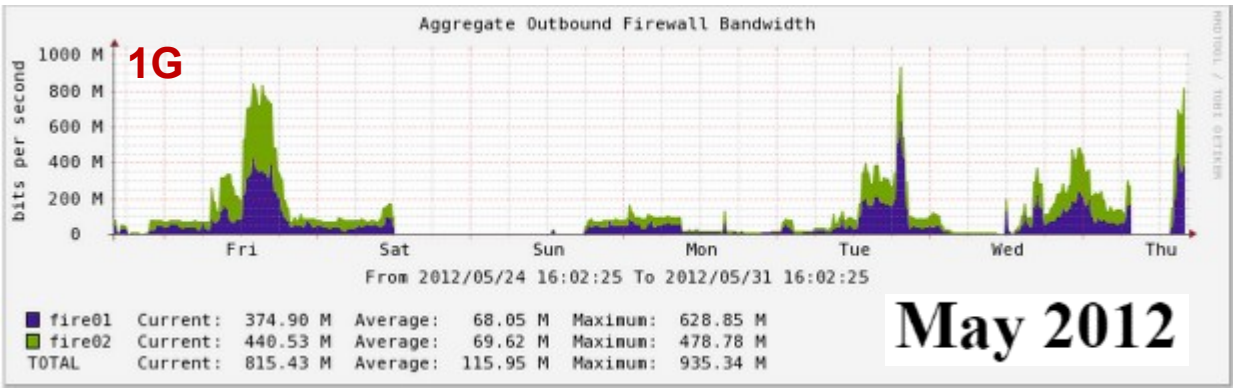


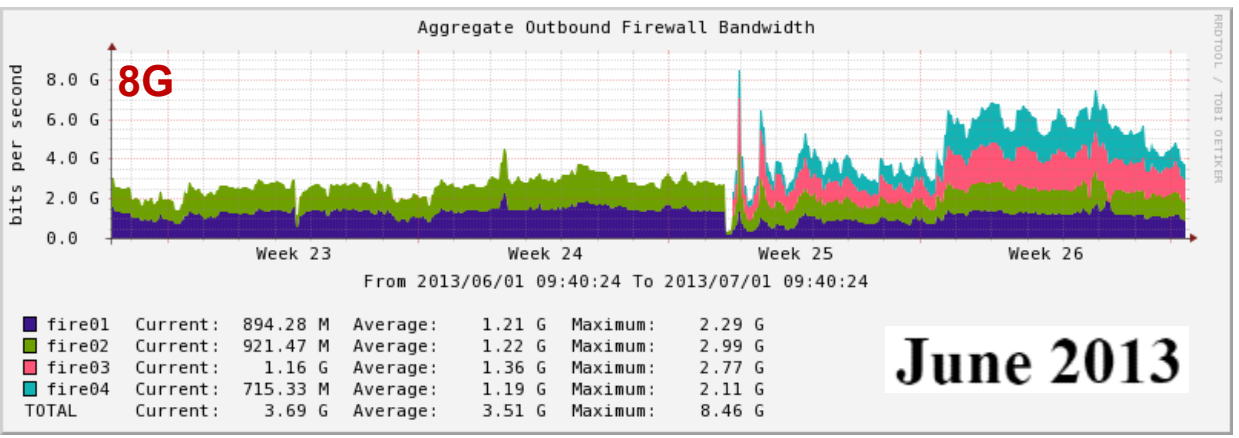
Flow Out of CERN for CMS Detector Peaks at 32 Gbps!

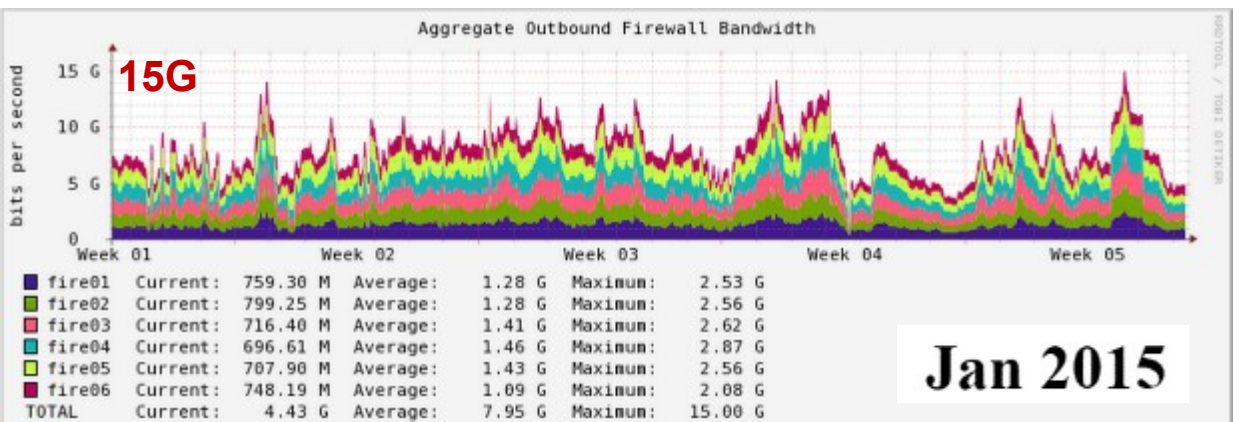




Cancer Genomics Hub (UCSC) is Housed in SDSC CoLo: Large Data Flows to End Users

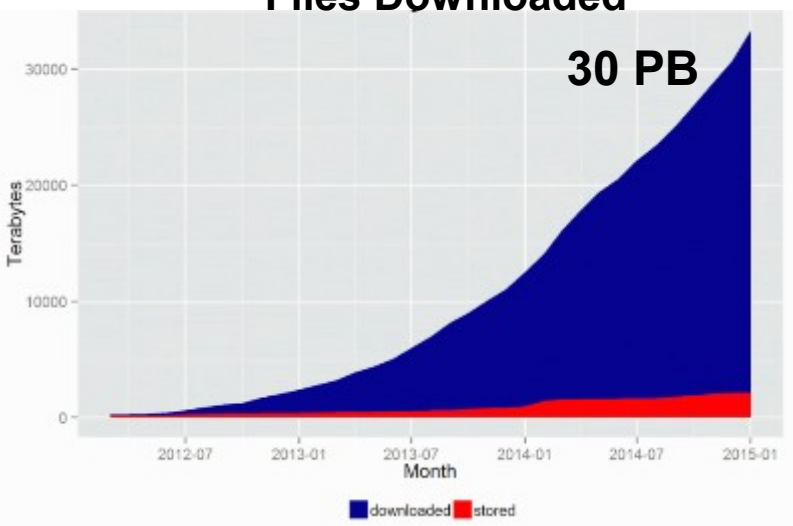








Cumulative TBs of CGH Files Downloaded



Data Source: David Haussler, Brad Smith, UCSC

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

800GB per night

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

Planning for climate change in California

substantial shifts on top of already high climate variability

SIO Campus Climate Researchers Need to Download Results from Remote Supercomputer Simulations to Make Regional Climate Change Forecasts

Dan Cayan

USGS Water Resources Discipline

Scripps Institution of Oceanography, UC San Diego

much support from Mary Tyree, Mike Dettinger, Guido Franco and other colleagues

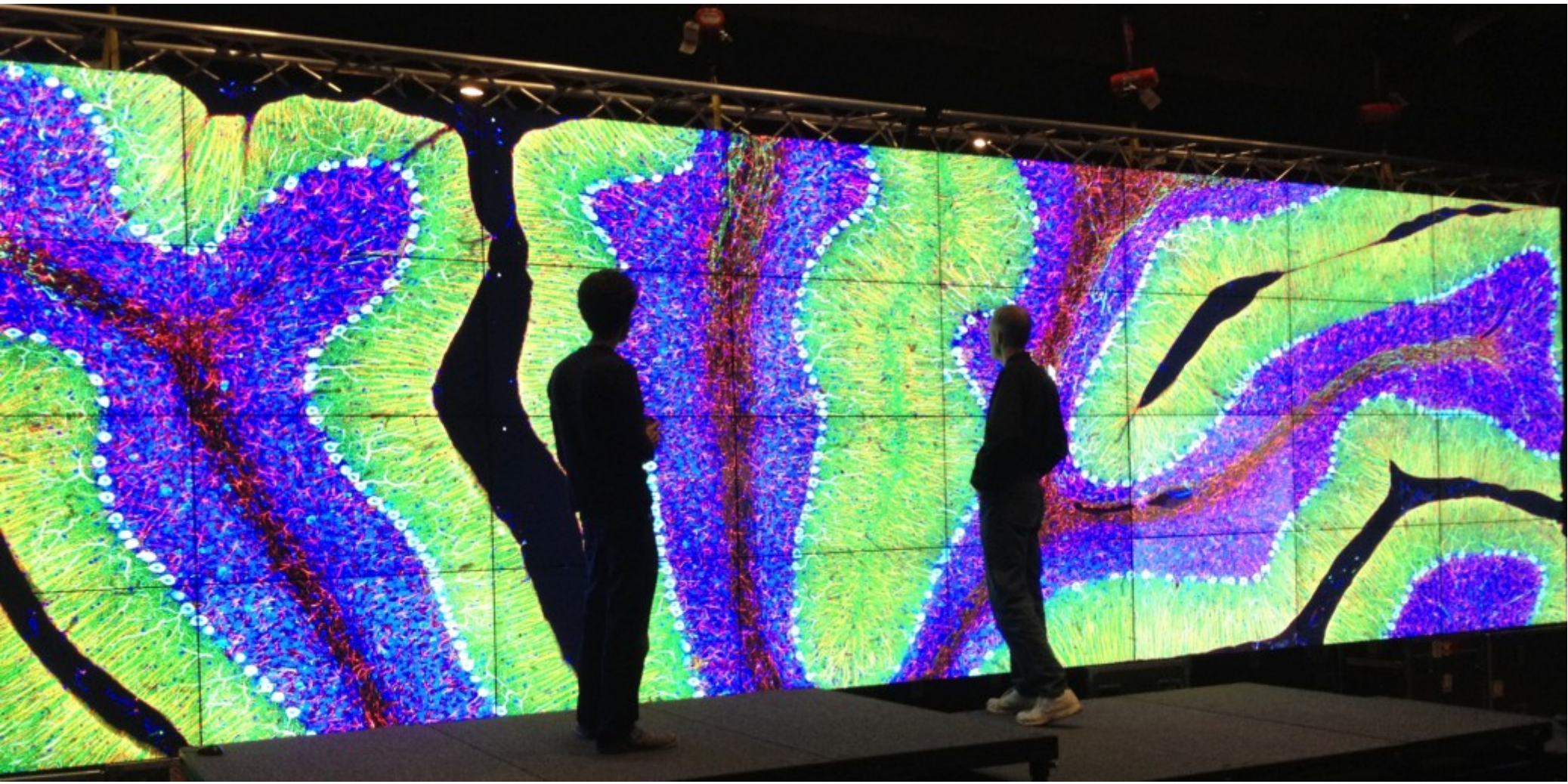
Sponsors:

California Energy Commission

NOAA RISA program

California DWR, DOE, NSF

Interactively Exploring Microscope Images of Brains: 40Gbps From NCMIR to Calit2 64Mpixel Wall





NCMIR National Center for Microscopy and Imaging Research

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



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

The White House Announcement Has Galvanized U.S. Campus CI Innovations

News

White House launches big data R&D effort

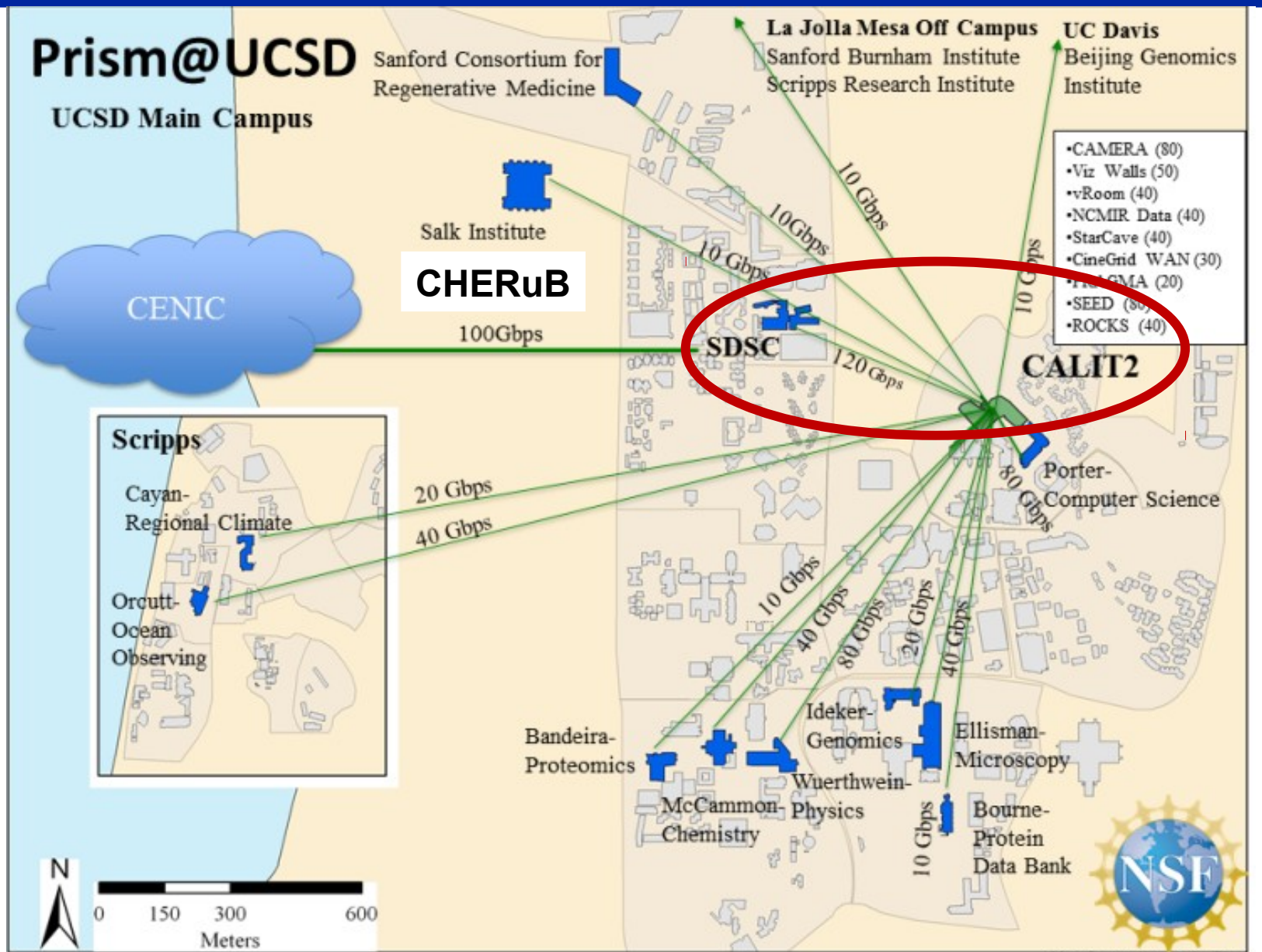
NSF and DARPA are among the agencies looking to harness the power of large volumes of data

By Grant Gross

March 29, 2012 12:34 PM ET

5 Comments + Briefcase What's this?

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



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

Making Critical High Performance Cyberinfrastructure Seamlessly Available to Users Where They Work



1-16

CENIC

CHERuB

10

Prism@UCSD

8







288

128

SDSC
Supercomputers



SDSC
SAN DIEGO SUPERCOMPUTER CENTER

384

384 x 10Gbps = 3.8Tbps

Oasis Data Store
>13,000 TB
> 800 Gbps

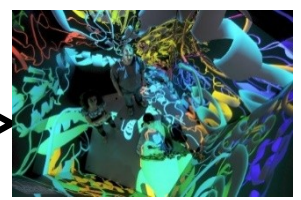
1

4

8

UCSD IDI Users

4





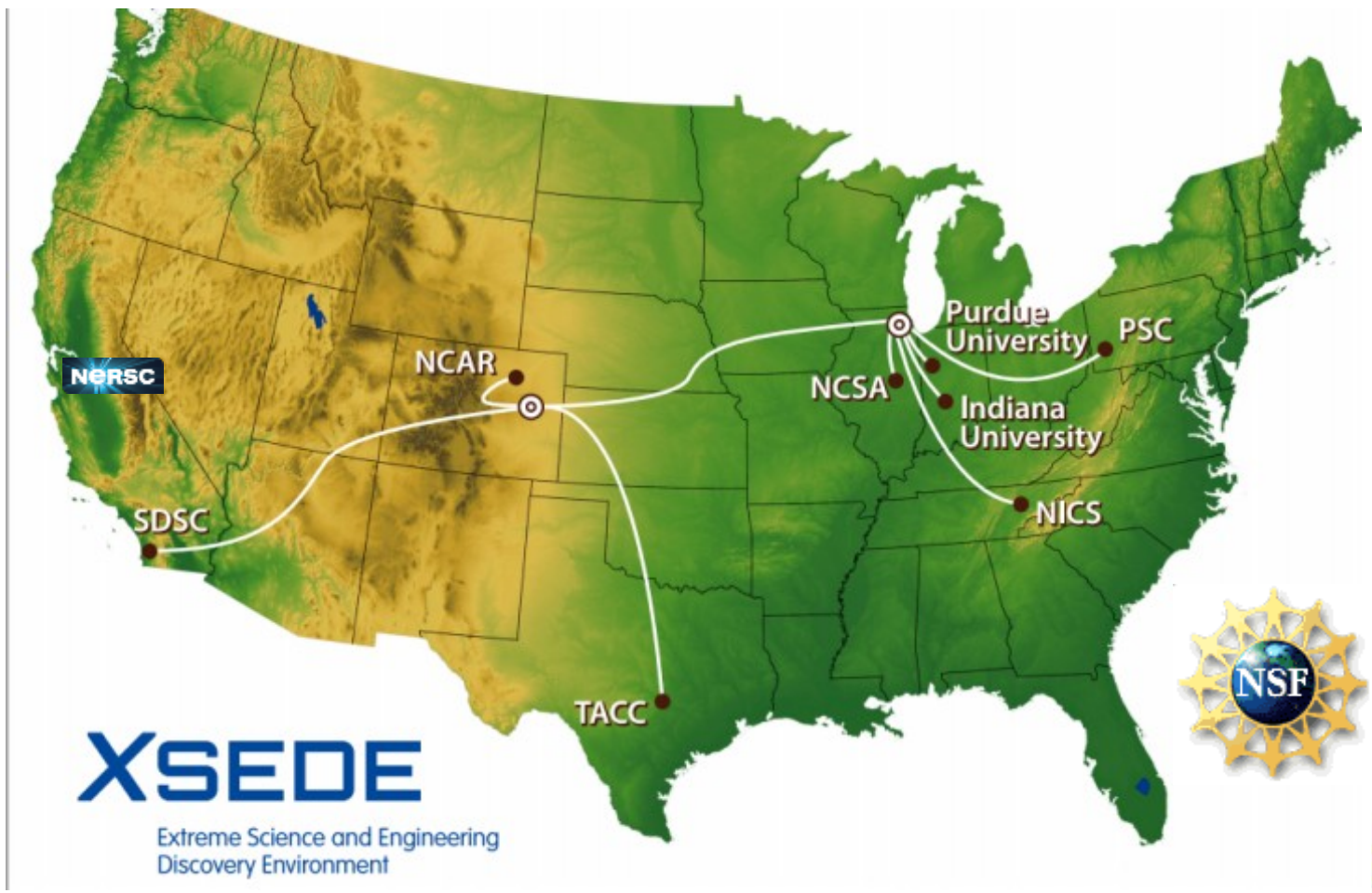


# of Parallel 10Gbps
Optical Light Paths

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

NERSC

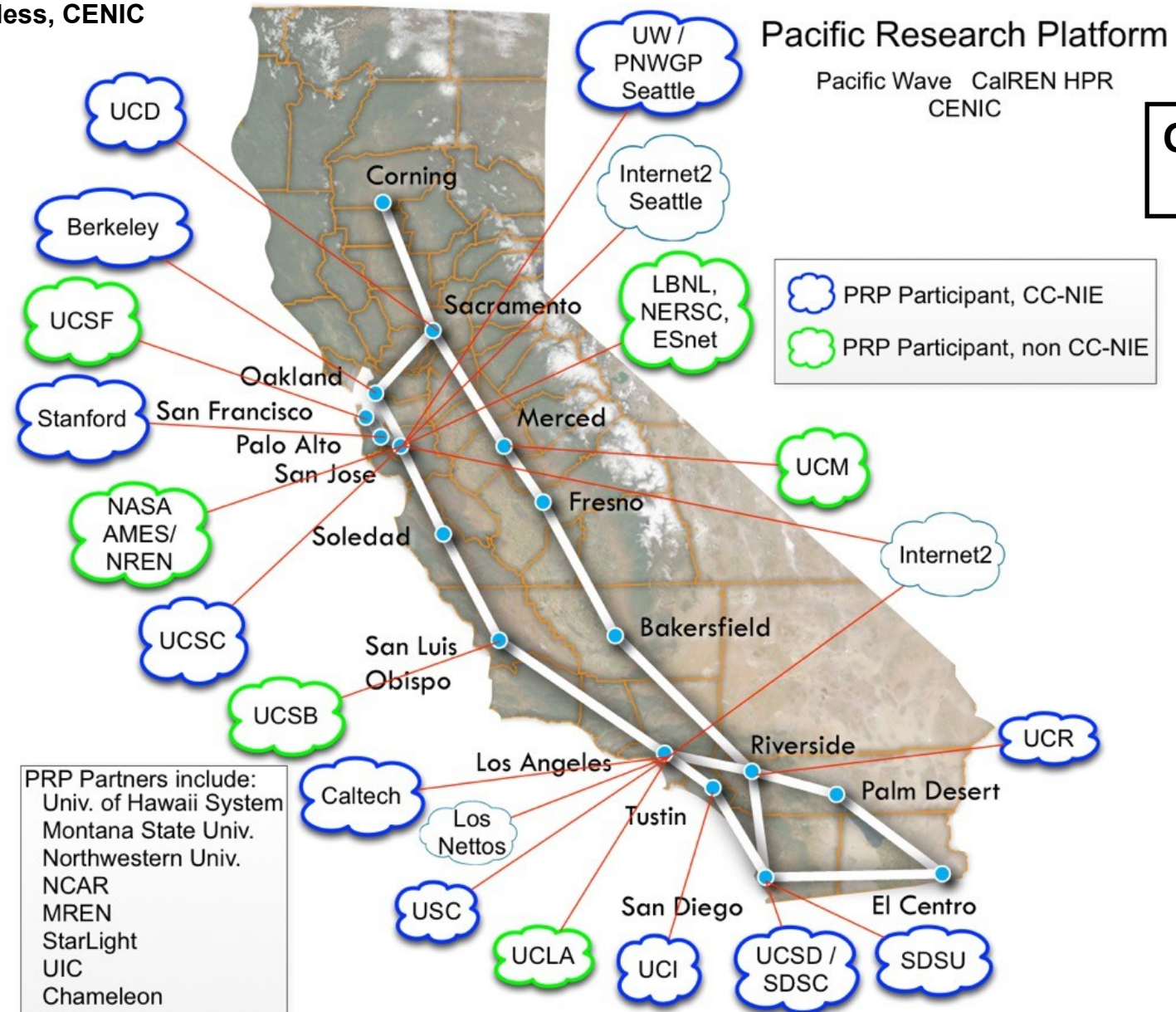
SDSC



it²

The Pacific Research Platform Creates a Regional Big Data Cyberinfrastructure

Map Source:
John Hess, CENIC



Optical Connections 10-100 Gbps

Organized by Calit2 and CITRIS

Ten Week Sprint to Demonstrate the West Coast Big Data Freeway System

Presented at CENIC 2015
March 9, 2015

As of 3/9/15, the Pacific Research Platform (PRPv0) as a facility, logs rather good performance:

From	To	Measured Bandwidth	Data Transfer Utility
San Diego State Univ.	UC Los Angeles	5Gb/s out of 10	GridFTP
UC Riverside	UC Los Angeles	9Gb/s out of 10	GridFTP
UC Berkeley	UC San Diego	9.6Gb/s out of 10	GridFTP
UC Davis	UC San Diego	9.6Gb/s out of 10	GridFTP
UC Irvine	UC Los Angeles	9.6Gb/s out of 10	GridFTP
UC Santa Cruz	UC San Diego	9.6Gb/s out of 10	FDT
Stanford	UC San Diego	12Gb/s out of 40	FDT
Univ. of Washington	UC San Diego	12Gb/s out of 40	FDT
UC Los Angeles	UC San Diego	36Gb/s out of 40	FDT
Caltech	UC San Diego	36Gb/s out of 40	FDT

Table I.2.1: Bandwidth of flash disk-to-flash disk file transfers shown between several sites for the existing experimental facility "PRPv0."

CENIC 2015 Conference
SHAKING THINGS UP
March 9-11 2015

The National Science Foundation Has Funded Over 100 Campuses to Build Data Freeways

