"Set My Data Free: High-Performance CI for Data-Intensive Research"

KeynoteSpeaker

Cyberinfrastructure Days

University of Michigan

Ann Arbor, MI

November 3, 2010

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

Follow me on Twitter: lsmarr

Abstract

As the need for large datasets and high-volume transfer grows, the shared Internet is becoming a bottleneck for cutting-edge research in universities. What are needed instead are large-bandwidth "data freeways." In this talk, I will describe some of the state-of-the-art uses of high-performance CI and how universities can evolve to support free movement of large datasets.

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

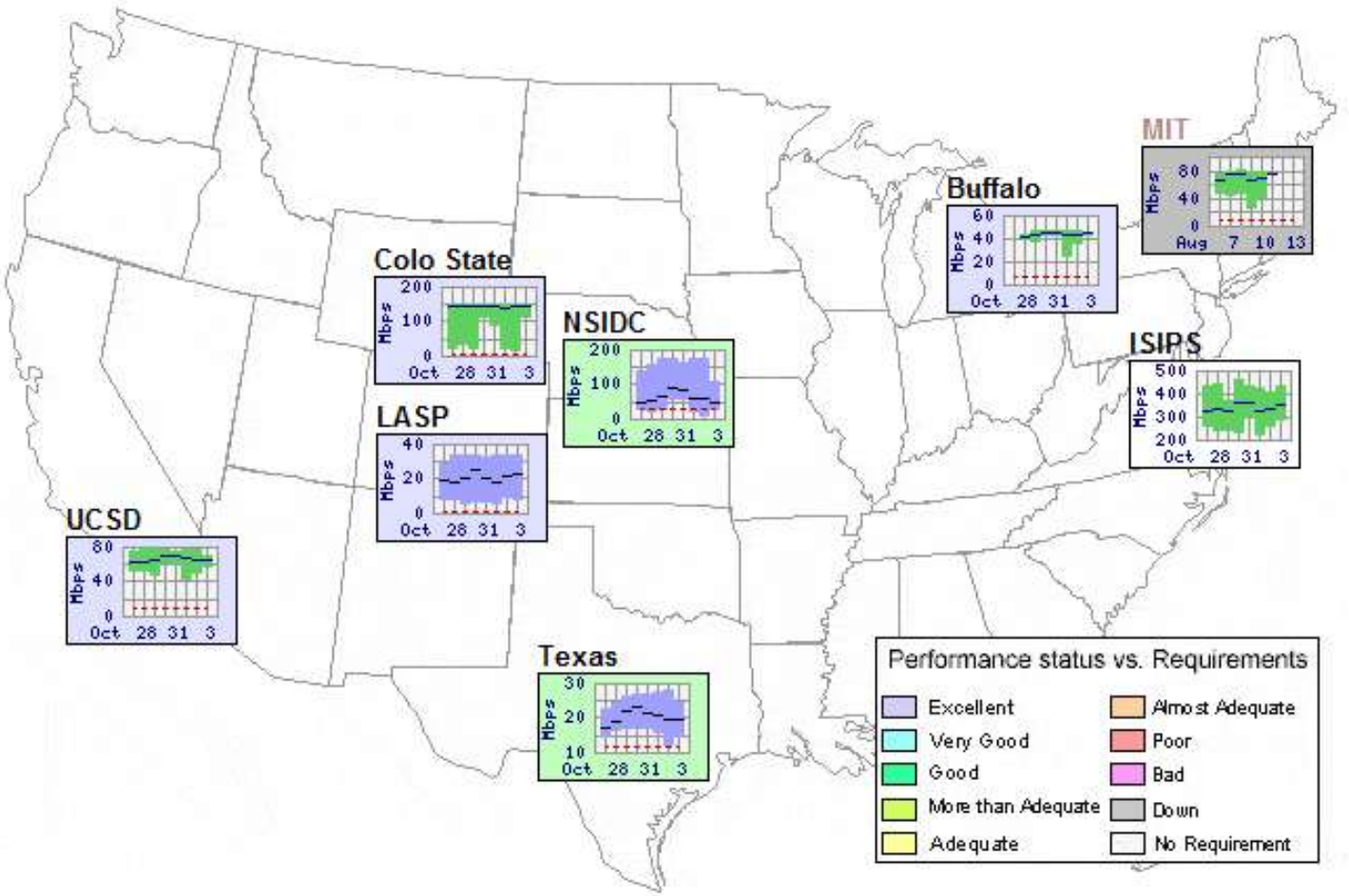


Source: SDSC

Large Data Challenge: Average Throughput to End User on Shared Internet is 10-100 Mbps



Tested
October 2010



Transferring 1 TB:
--10 Mbps = 10 Days
--10 Gbps = 15 Minutes

http://ensight.eos.nasa.gov/Missions/icesat/index.shtml

The Large Hadron Collider Uses a Global Fiber Infrastructure To Connect Its Users

CERN Unveils Global Grid For Particle Physics Research

The network can pull in the IT power of more than 140 computer centers in 33 countries to crunch an expected 15 million GB of data every year.

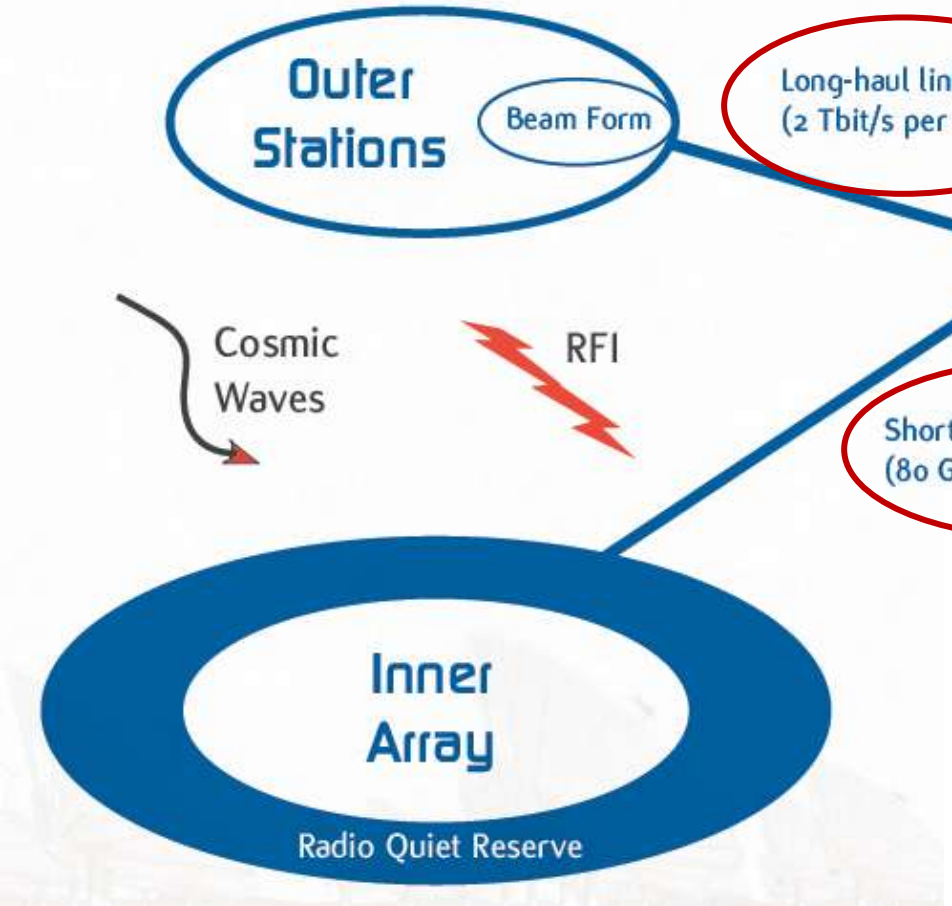
By Antone Gonsalves
InformationWeek
October 3, 2008 04:57 PM



- The grid relies on optical fiber networks to distribute data from CERN to 11 major computer centers in Europe, North America, and Asia
- The grid is capable of routinely processing 250,000 jobs a day
- The data flow will be ~6 Gigabits/sec or 15 million gigabytes a year for 10 to 15 years

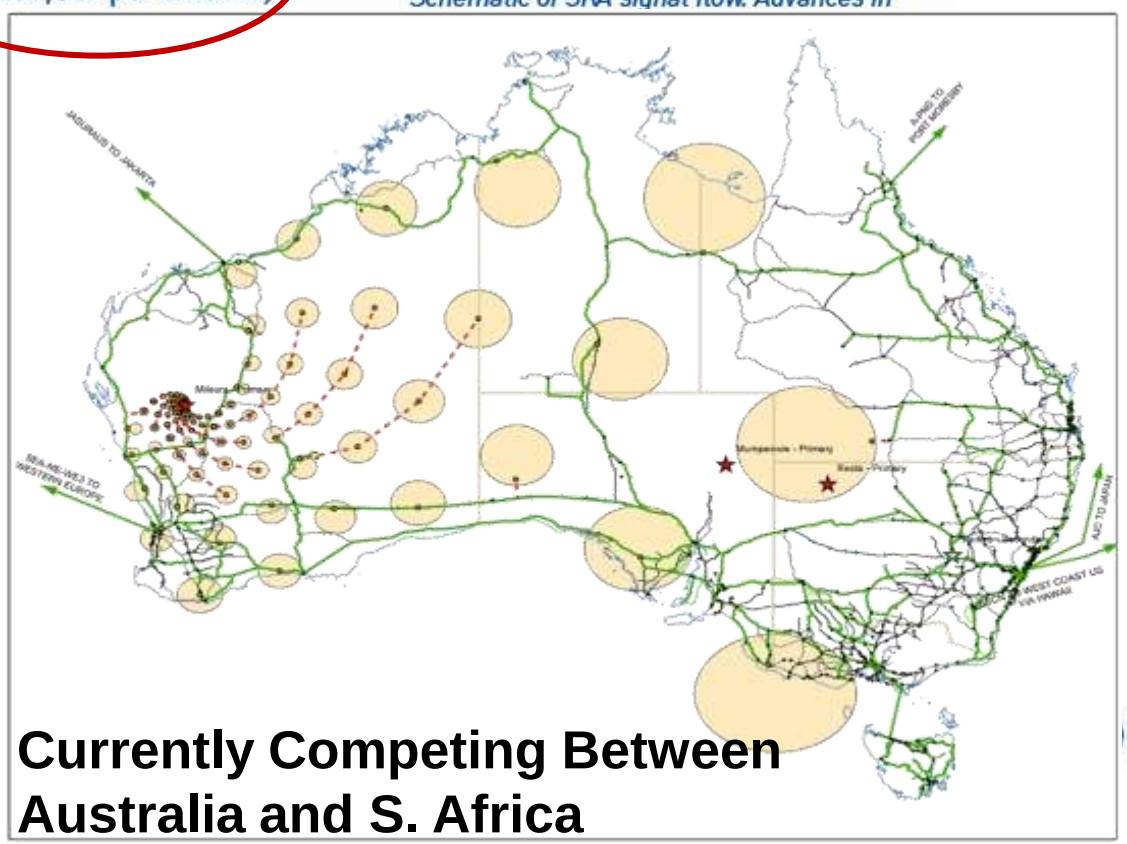
Next Great Planetary Instrument:
The Square Kilometer Array Requires Dedicated Fiber

www.skatelescope.org



Transfers Of 1 TByte Images World-wide Will Be Needed Every Minute!

Schematic of SKA signal flow. Advances in



Currently Competing Between Australia and S. Africa

Grand Challenges in Data-Intensive Sciences

October 26-28, 2010

San Diego Supercomputer Center , UC San Diego

Confirmed conference *topics* and speakers :

- ***Needs and Opportunities in Observational Astronomy*** - **Alex Szalay, JHU**
- ***Transient Sky Surveys*** – **Peter Nugent, LBNL**
- ***Large Data-Intensive Graph Problems*** – **John Gilbert, UCSB**
- ***Algorithms for Massive Data Sets*** – **Michael Mahoney, Stanford U.**
- ***Needs and Opportunities in Seismic Modeling and Earthquake Preparedness*** - **Tom Jordan, USC**
- ***Needs and Opportunities in Fluid Dynamics Modeling and Flow Field Data Analysis*** – **Parviz Moin, Stanford U.**
- ***Needs and Emerging Opportunities in Neuroscience*** – **Mark Ellisman, UCSD**
- ***Data-Driven Science in the Globally Networked World*** – **Larry Smarr, UCSD**

SDSC

Petascale High Performance Computing Generates TB Datasets to Analyze

it²

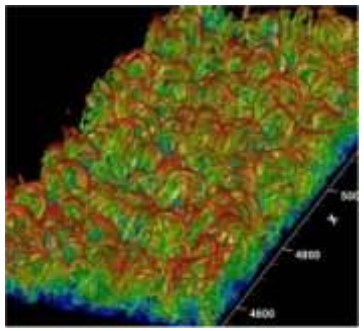
Growth of Turbulence Data Over Three Decades (Assuming Double Precision and Collocated Points)

Year	Authors	Simulation	Points	Size
1972	Orszag & Patterson	Isotropic Turbulence	32^3	1 MB
1987	Kim, Moin & Moser	Plane Channel Flow	192x160x128	120 MB
1988	Spalart	Turbulent Boundary Layer	432x80x320	340 MB
1994	Le & Moin	Backward-Facing Step	768x64x192	288 MB
2000	Freund, Lele & Moin	Compressible Turbulent Jet	640x270x128	845 MB
2003	Earth Simulator	Isotropic Turbulence	4096^3	0.8 TB*
2006	Hoyas & Jiménez	Plane Channel Flow	6144x633x4608	550 GB
2008	Wu & Moin	Turbulent Pipe Flow	$256x512^2$	2.1 GB
2009	Larsson & Lele	Isotropic Shock-Turbulence	$1080x384^2$	6.1 GB
2010	Wu & Moin	Turbulent Boundary Layer	8192x500x256	40 GB

Turbulent Boundary Layer:

One-Periodic Direction

100x Larger Data Sets in 20 Years



SDSC

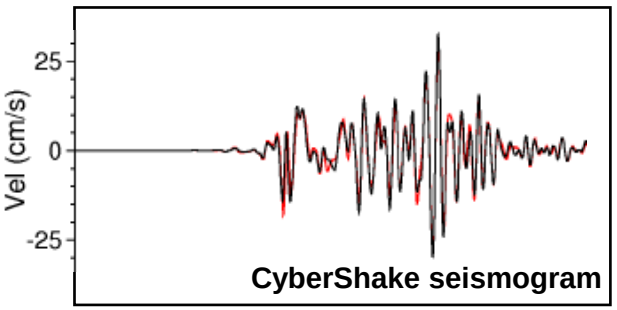
Source: Parviz Moin, Stanford

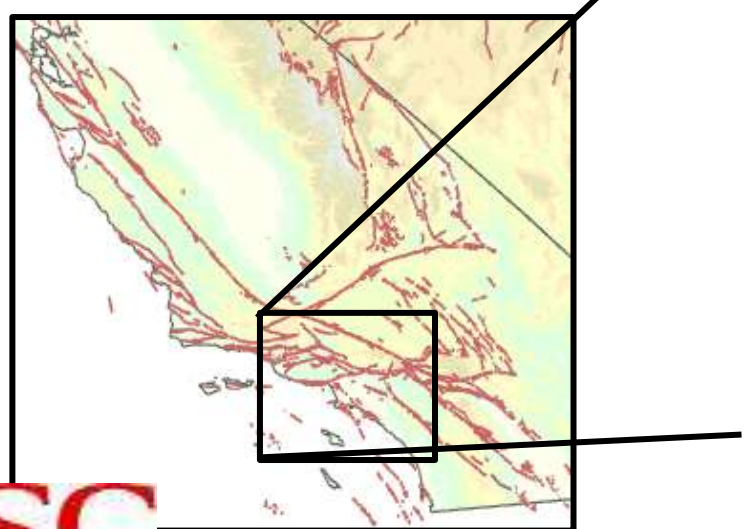
it²

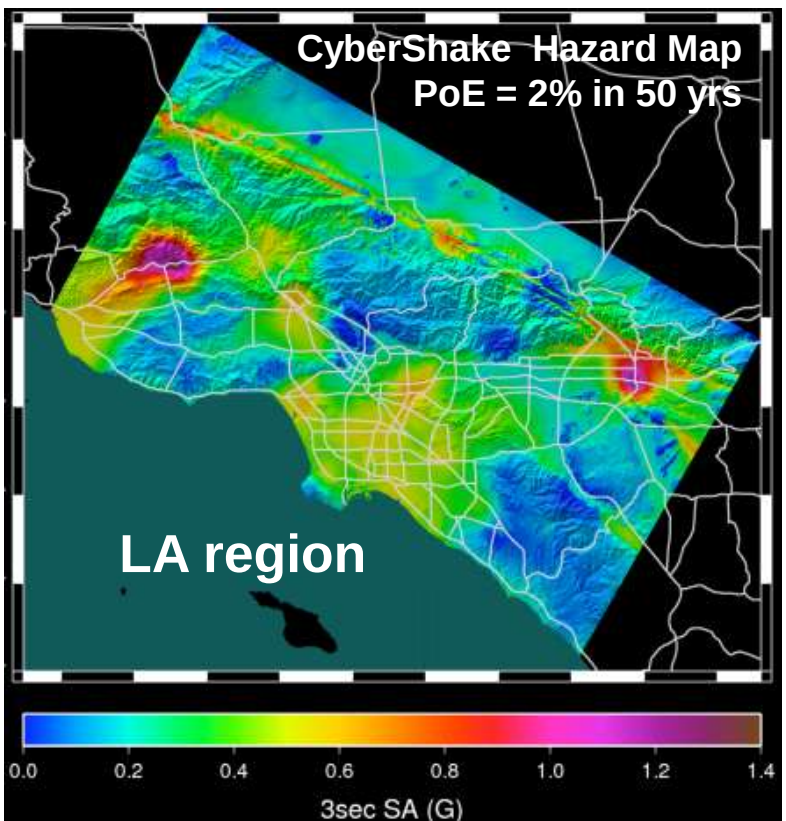
CyberShake 1.0 Hazard Model
Need to Analyze Terabytes of Computed Data

- **CyberShake 1.0 Computation**
  - 440,000 Simulations per Site
  - 5.5 Million CPU hrs (50-Day Run on *Ranger* Using 4,400 cores)
  - 189 Million Jobs
  - 165 TB of Total Output Data
  - 10.6 TB of Stored Data
  - 2.1 TB of Archived Data

Source: Thomas H. Jordan, USC, Director, Southern California Earthquake Center







SDSC

it²

Large-Scale PetaApps Climate Change Run Generates Terabyte Per Day of Computed Data

- **155 Year Control Run**
  - 0.1° Ocean model [3600 x 2400 x 42]
  - 0.1° Sea-ice model [3600 x 2400 x 20]
  - 0.5° Atmosphere [576 x 384 x 26]
  - 0.5° Land [576 x 384]
- **Statistics**
  - ~18M CPU Hours
  - 5844 Cores for 4-5 Months
  - ~100 TB of Data Generated
  - 0.5 to 1 TB per Wall Clock Day Generated

100x Current Production

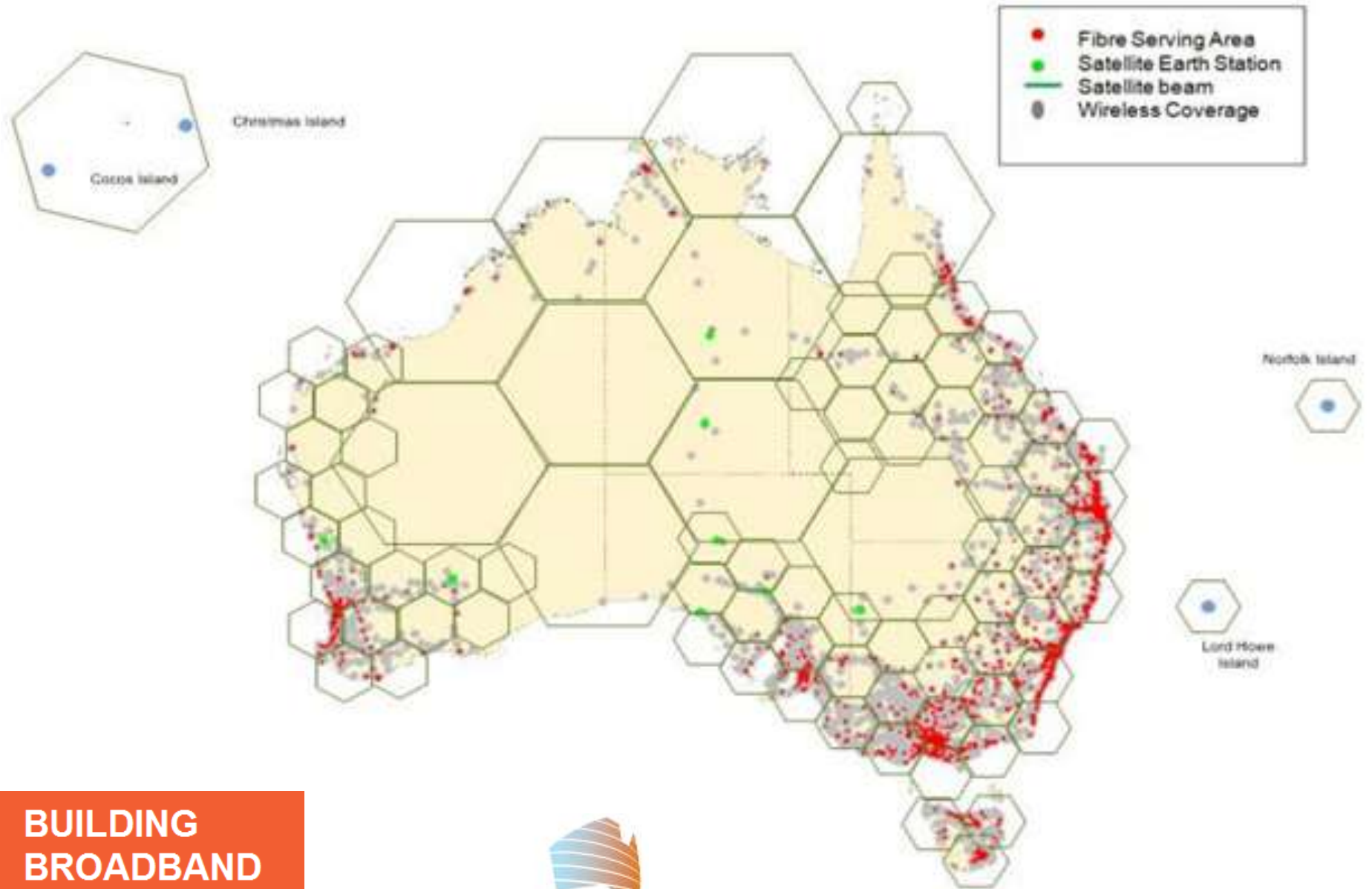
4x current production

Source: John M. Dennis, Matthew Woitaszek, UCAR

The Required Components of High Performance Cyberinfrastructure

- High Performance Optical Networks
- Scalable Visualization and Analysis
- Multi-Site Collaborative Systems
- End-to-End Wide Area CI
- Data-Intensive Campus Research CI

Australia—The Broadband Nation:
Universal Coverage with Fiber, Wireless, Satellite



BUILDING BROADBAND FOR ALL AUSTRALIANS

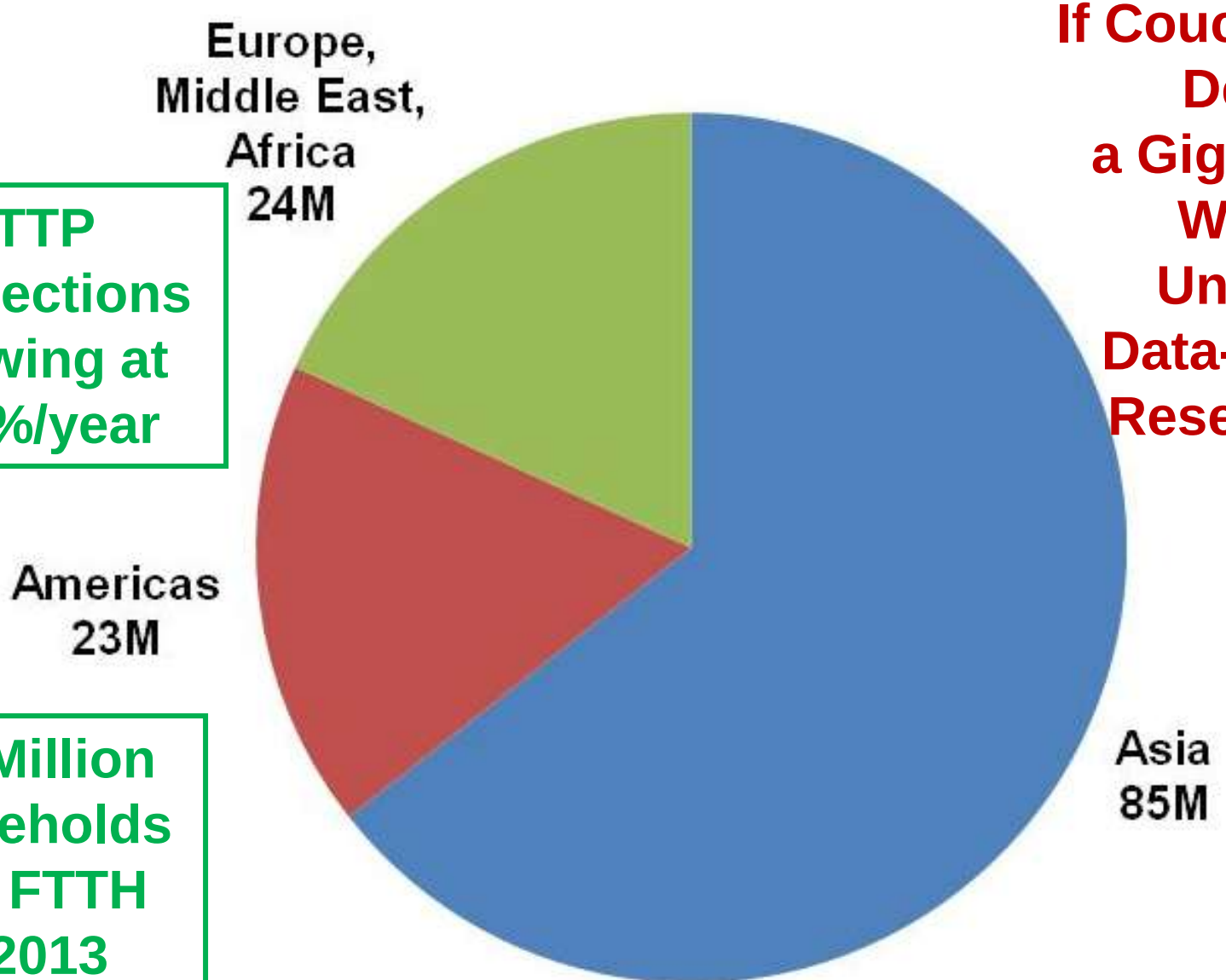
NBNCo Limited

www.nbnco.com.au

Globally Fiber to the Premise is Growing Rapidly, Mostly in Asia

If Couch Potatoes Deserve a Gigabit Fiber, Why Not University Data-Intensive Researchers?

FTTP Connections Growing at ~30%/year

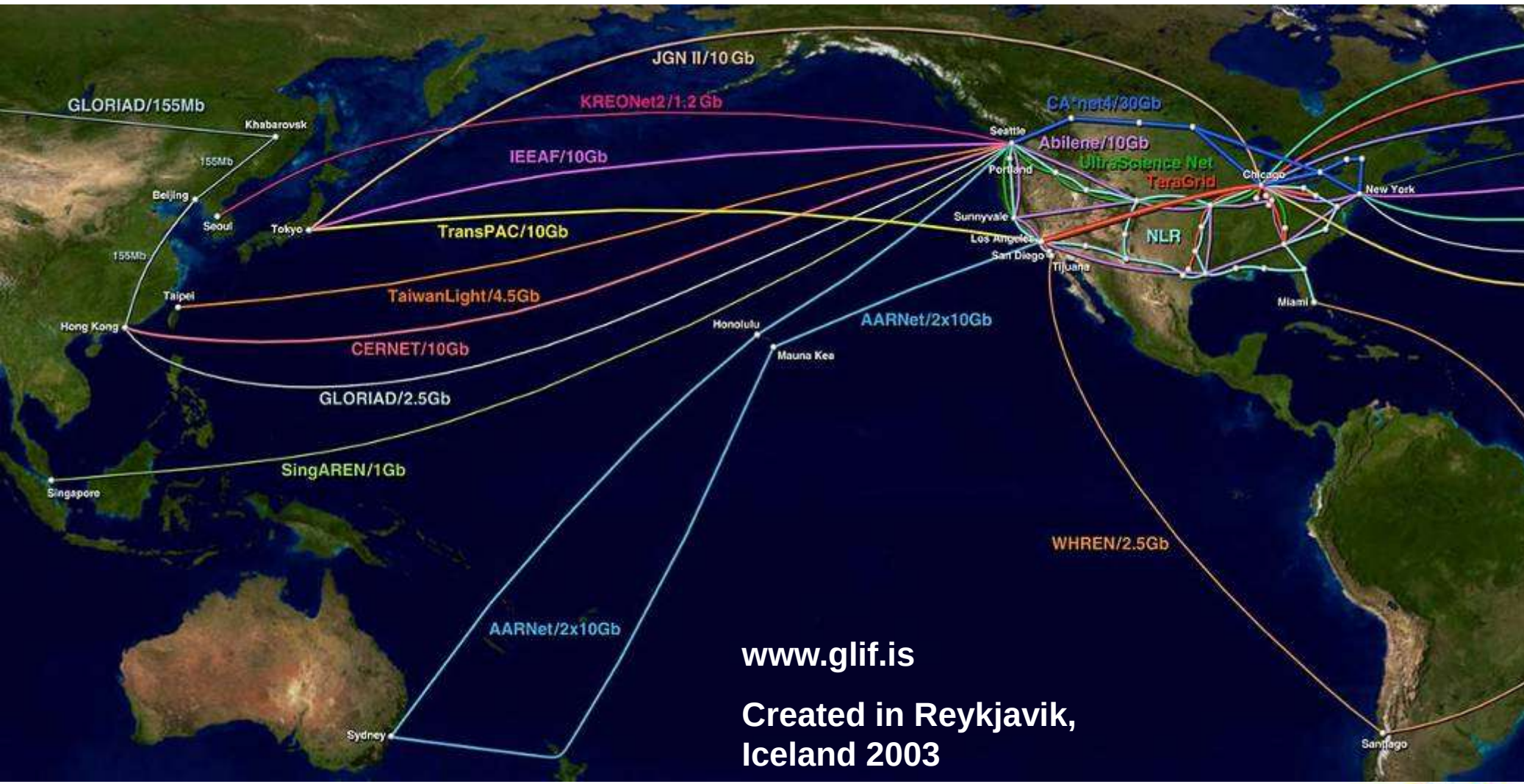


130 Million Households with FTTH in 2013

Source: Heavy Reading (www.heavyreading.com), the market research division of Light Reading (www.lightreading.com).

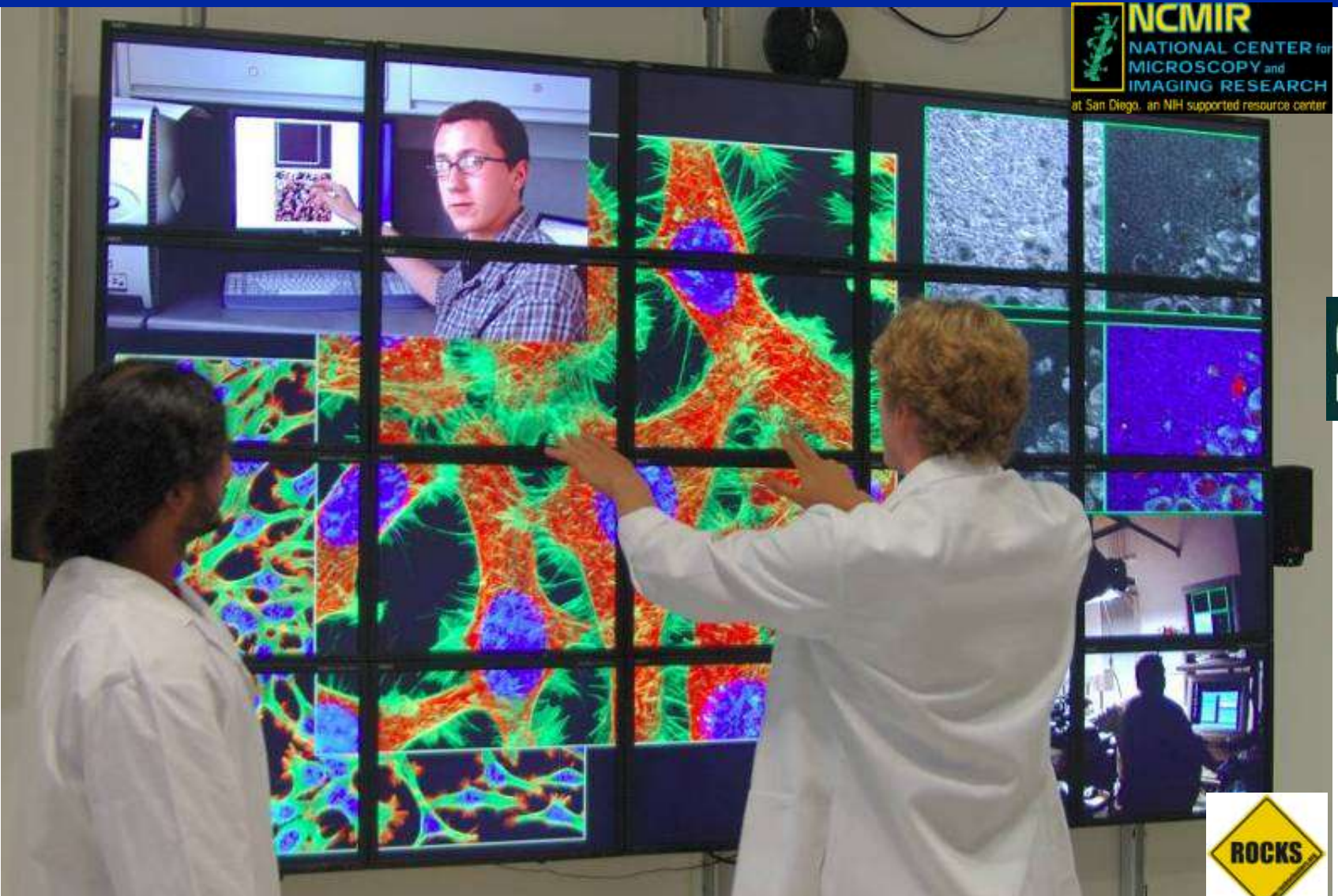
The Global Lambda Integrated Facility--
Creating a Planetary-Scale High Bandwidth Collaboratory

Research Innovation Labs Linked by 10G GLIF



Visualization courtesy of
Bob Patterson, NCSA.

The OptIPuter Project: Creating High Resolution Portals Over Dedicated Optical Channels to Global Science Data



Scalable Adaptive Graphics Environment (SAGE)

evl electronic visualization laboratory

NSF

Picture Source: Mark Ellisman, David Lee, Jason Leigh

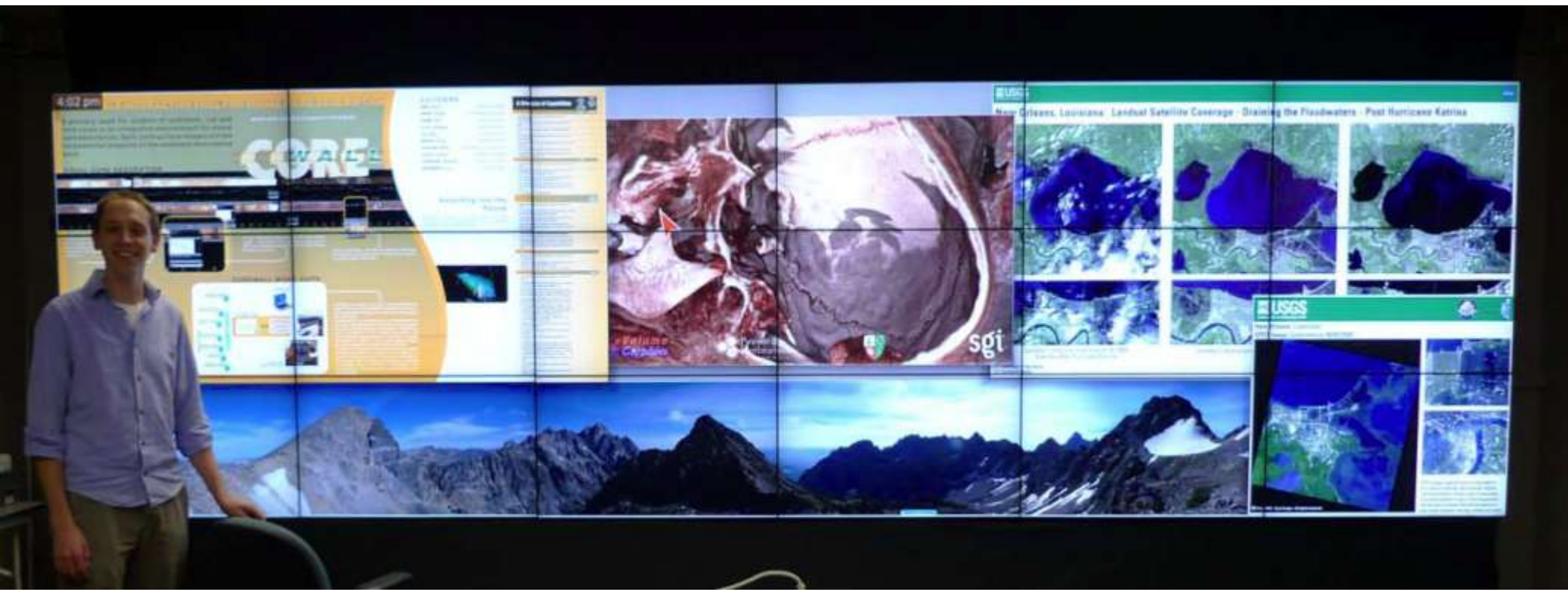
Calit2 (UCSD, UCI), SDSC, and UIC Leads—Larry Smarr PI

Univ. Partners: NCSA, USC, SDSU, NW, TA&M, UvA, SARA, KISTI, AIST

Industry: IBM, Sun, Telcordia, Chiaro, Calient, Glimmerglass, Lucent

Nearly Seamless AESOP OptIPortal

46” NEC Ultra-Narrow Bezel 720p LCD Monitors



Source: Tom DeFanti, Calit2@UCSD;

3D Stereo Head Tracked OptIPortal: NexCAVE

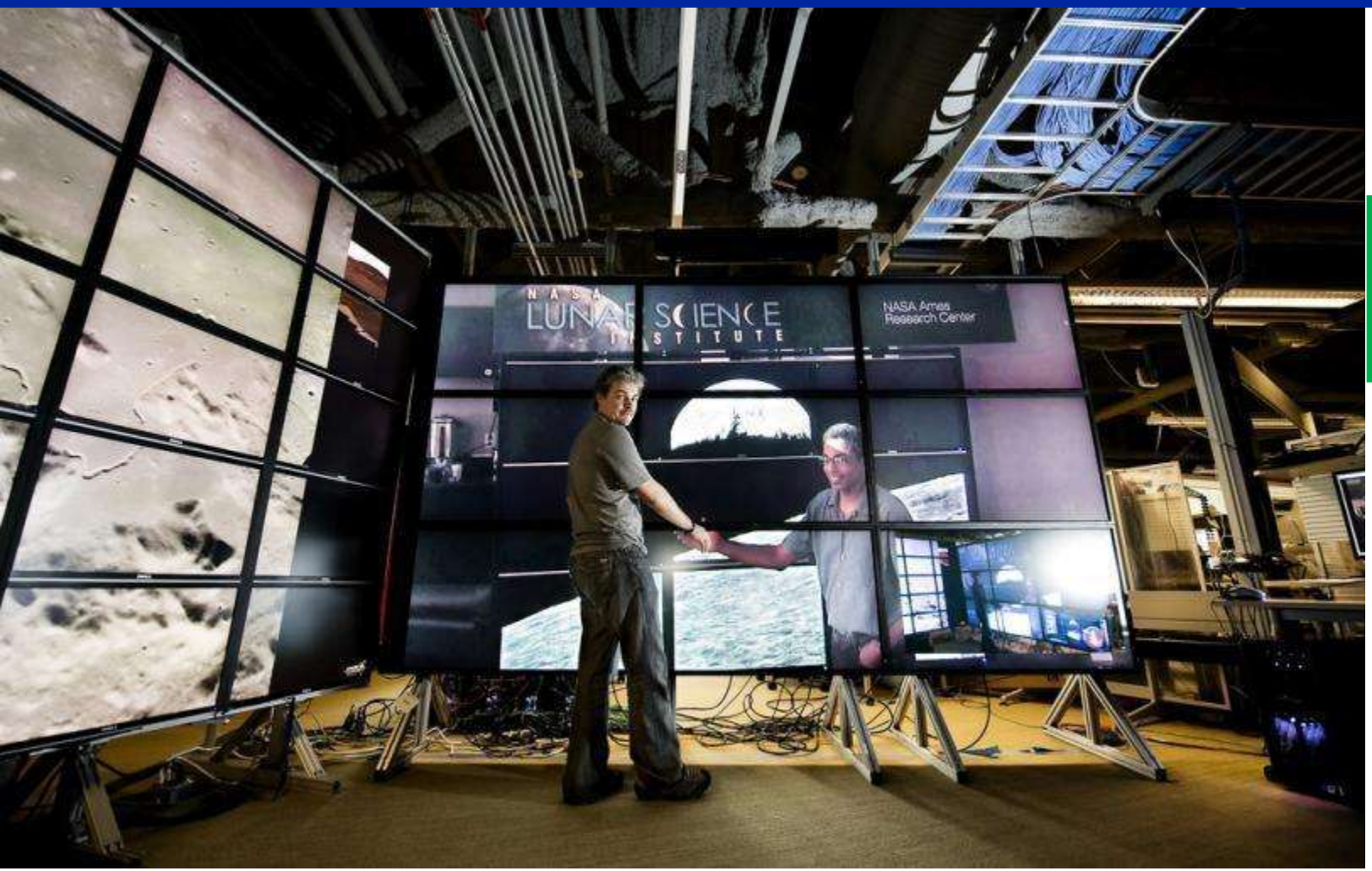


Array of JVC HDTV 3D LCD Screens
KAUST NexCAVE = 22.5MPixels

www.calit2.net/newsroom/article.php?id=1584

Source: Tom DeFanti, Calit2@UCSD

High Definition Video Connected OptIPortals: Virtual Working Spaces for Data Intensive Research



NASA Supports Two Virtual Institutes

LifeSize HD

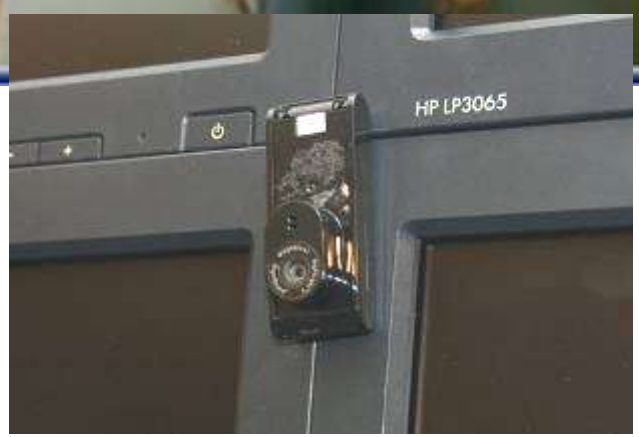
Calit2@UCSD 10Gbps Link to NASA Ames Lunar Science Institute, Mountain View, CA

Source: Falko Kuester, Kai Doerr Calit2; Michael Sims, Larry Edwards, Estelle Dodson NASA

U Michigan Virtual Space Interaction Testbed (VISIT) Instrumenting OptIPortals for Social Science Research

- **Using Cameras Embedded in the Seams of Tiled Displays and Computer Vision Techniques, we can Understand how People Interact with OptIPortals**
  - **Classify Attention, Expression, Gaze**
  - **Initial Implementation Based on Attention Interaction Design Toolkit (J. Lee, MIT)**
- **Close to Producing Usable Eye/Nose Tracking Data using OpenCV**

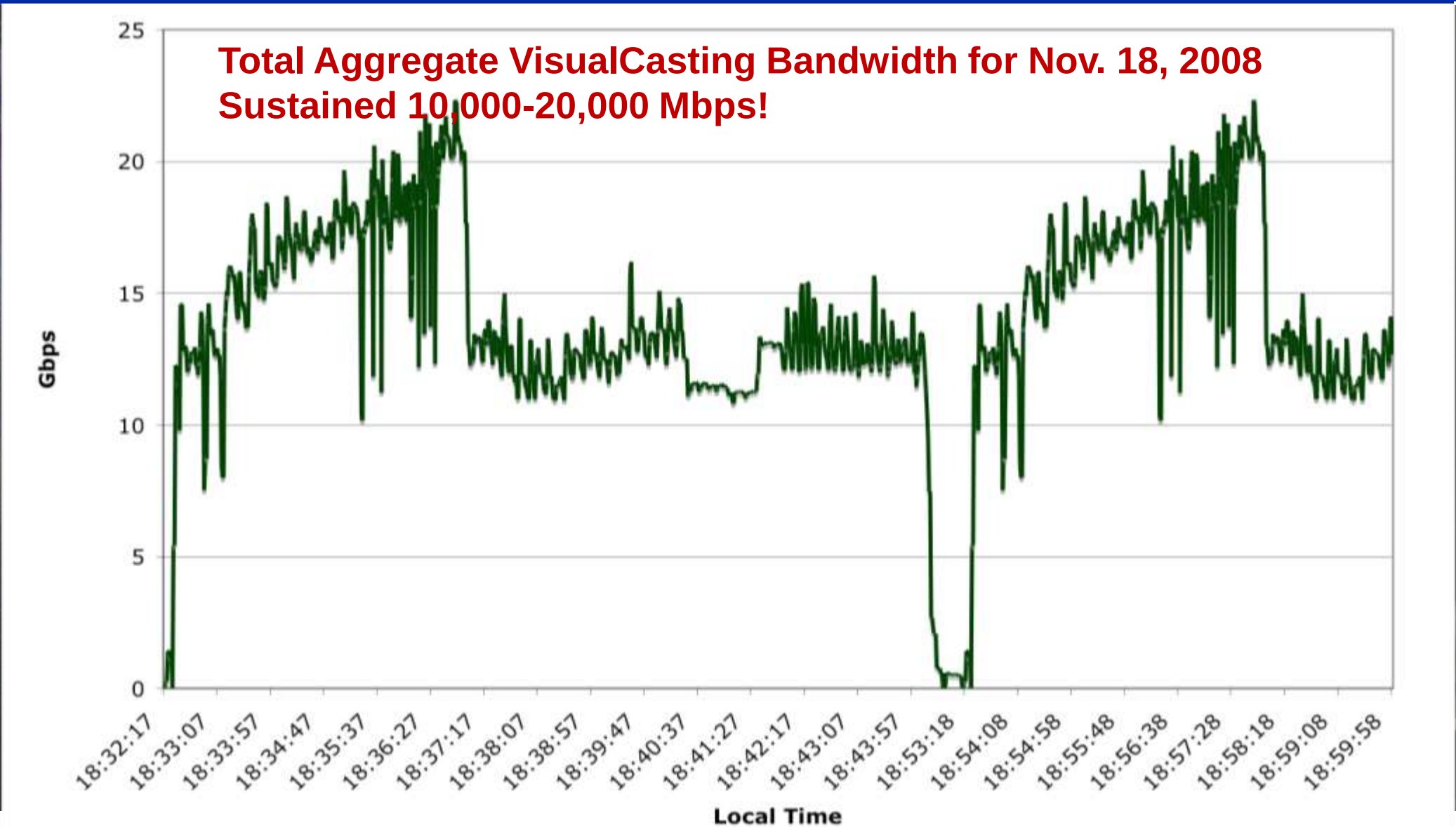




Leading U.S. Researchers on the Social Aspects of Collaboration

Source: Erik Hofer, UMich, School of Information

EVL's SAGE OptIPortal VisualCasting Multi-Site OptIPuter Collaboratory

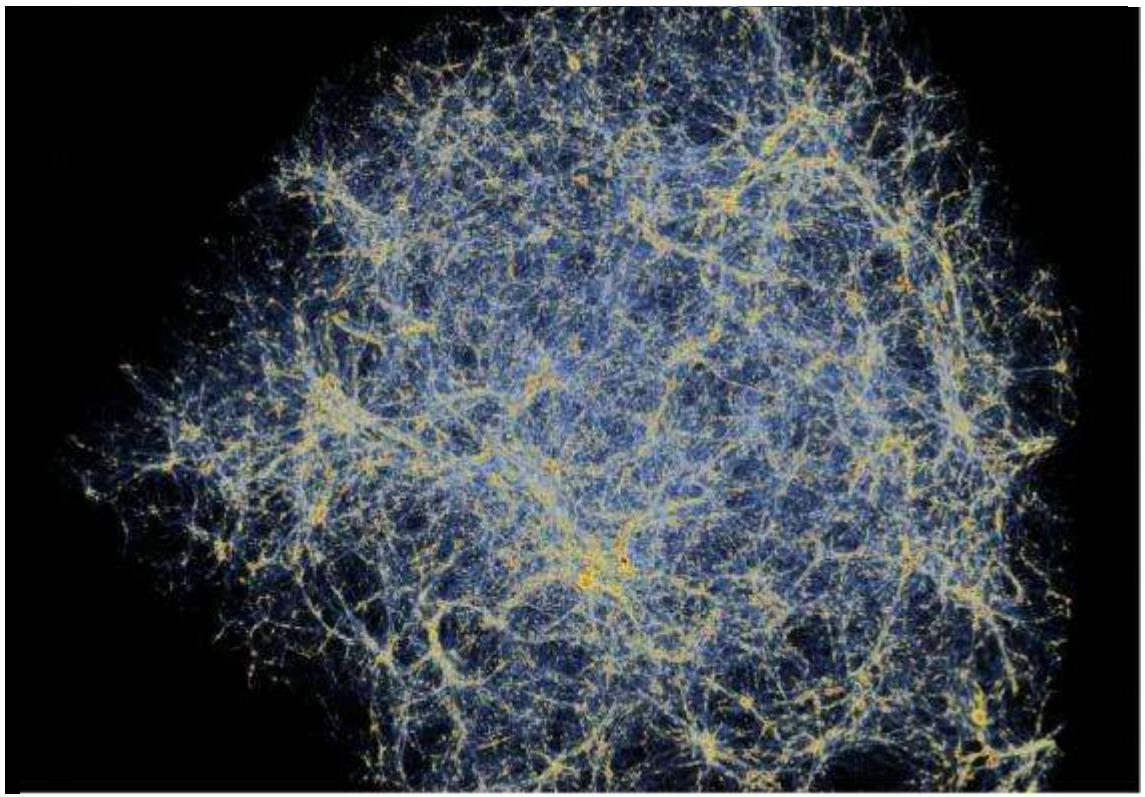


Source: Jason Leigh, Luc Renambot, EVL, UI Chicago

Exploring Cosmology With Supercomputers, Supernetworks, and Supervisualization

Source: Mike Norman, SDSC

Intergalactic Medium on 2 GLyr Scale



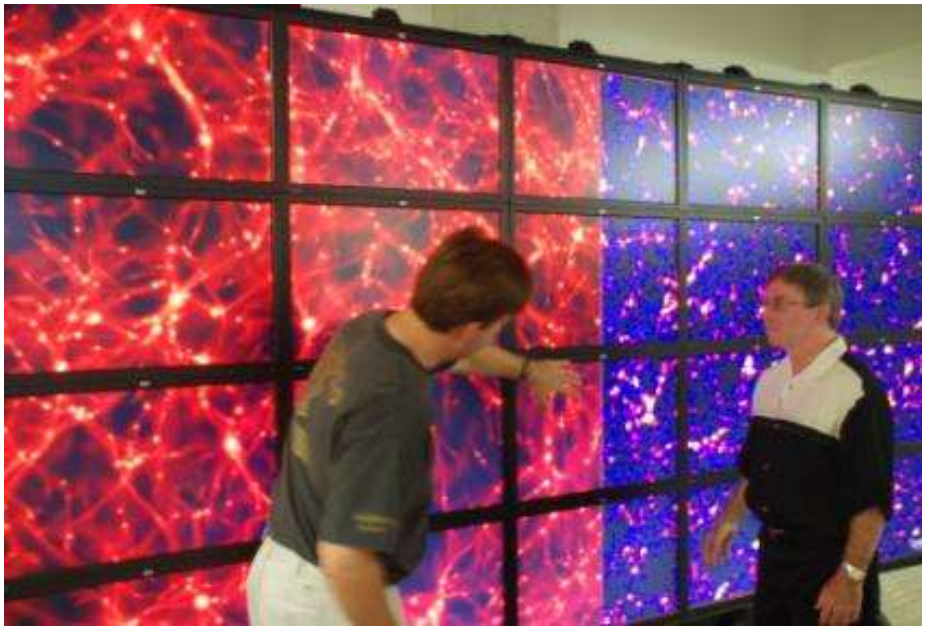
Science: Norman, Harkness,Paschos SDSC
Visualization: Insley, ANL; Wagner SDSC

- **4096^3 Particle/Cell Hydrodynamic Cosmology Simulation**
- **NICS Kraken (XT5)**
  - **16,384 cores**
- **Output**
  - **148 TB Movie Output (0.25 TB/file)**
  - **80 TB Diagnostic Dumps (8 TB/file)**

Project StarGate Goals: Combining Supercomputers and Supernetworks

- Create an "End-to-End" 10Gbps Workflow
- Explore Use of OptIPortals as Petascale Supercomputer "Scalable Workstations"
- Exploit Dynamic 10Gbps Circuits on ESnet
- Connect Hardware Resources at ORNL, ANL, SDSC
- Show that Data Need Not be Trapped by the Network "Event Horizon"

OptIPortal@SDSC



Rick Wagner Mike Norman

Source: Michael Norman, SDSC, UCSD

- ANL * Calit2 * LBNL * NICS * ORNL * SDSC

Using Supernetworks to Couple End User's OptIPortal to Remote Supercomputers and Visualization Servers

Source: Mike Norman, Rick Wagner, SDSC

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



SDSC

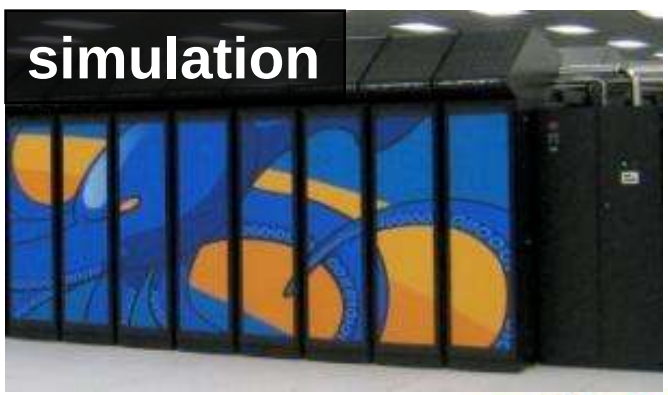


ESnet
10 Gb/s fiber optic network

NICS ORNL

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

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



SDSC

***ANL * Calit2 * LBNL * NICS * ORNL * SDSC**

it²

National-Scale Interactive Remote Rendering of Large Datasets Over 10Gbps Fiber Network

SDSC



OptIPortal (40M pixels LCDs)
10 NVIDIA FX 4600 Cards
10 Gb/s Network Throughout

ESnet

Science Data Network (SDN)
> 10 Gb/s Fiber Optic Network
Dynamic VLANs Configured
Using OSCARS

ALCF



Eureka
100 Dual Quad Core Xeon Servers
200 NVIDIA FX GPUs
3.2 TB RAM

Interactive Remote Rendering

Real-Time Volume Rendering Streamed from ANL to SDSC

Last Year

High-Resolution (4K+, 15+ FPS)—But:
- Command-Line Driven
- Fixed Color Maps, Transfer Functions
- Slow Exploration of Data

Last Week

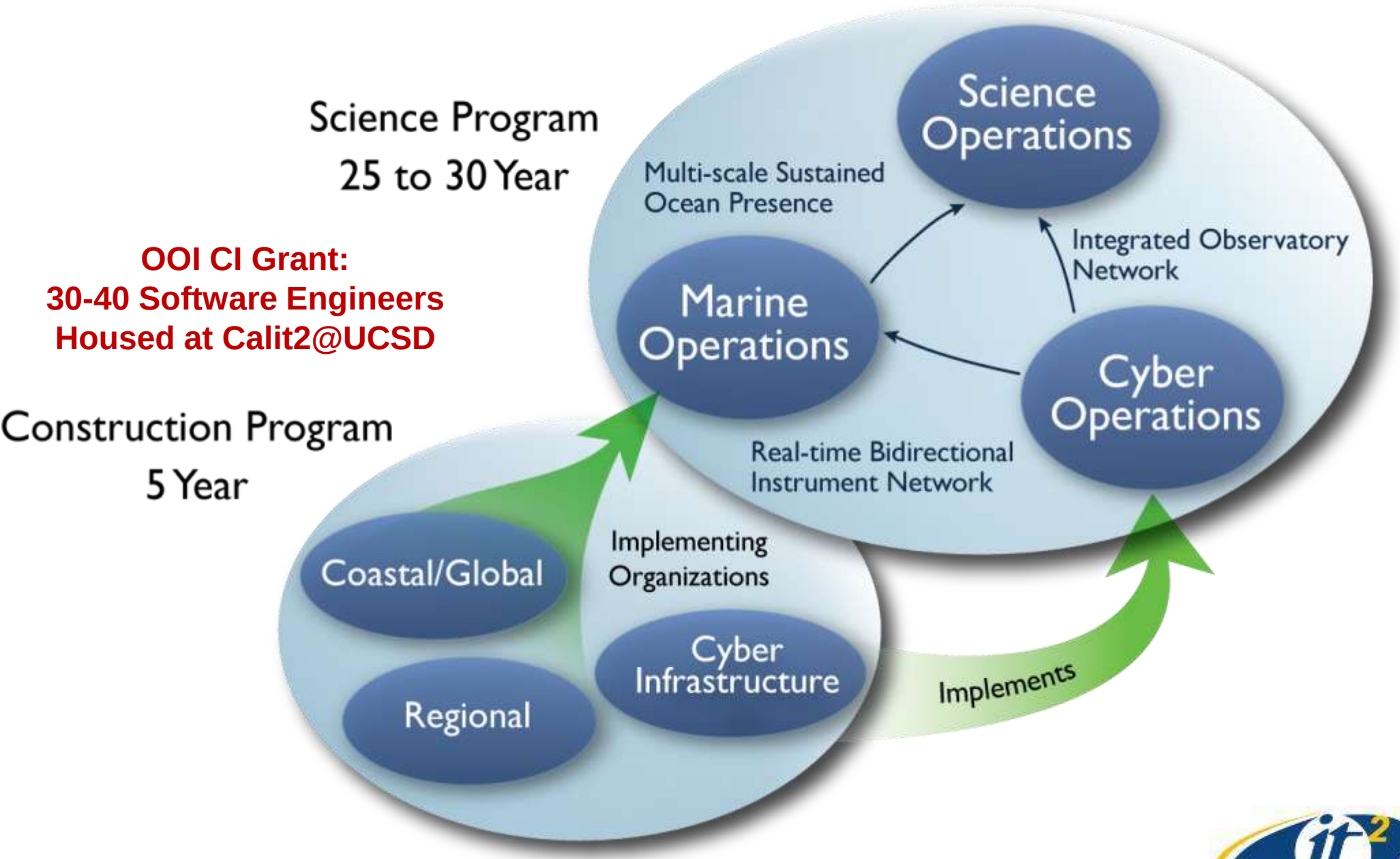
Now Driven by a Simple Web GUI
- Rotate, Pan, Zoom
- GUI Works from Most Browsers
- Manipulate Colors and Opacity
- Fast Renderer Response Time

Source: Rick Wagner, SDSC

NSF's Ocean Observatory Initiative Has the Largest Funded NSF CI Grant

OCEAN OBSERVATORIES INITIATIVE

OOI CI Grant:
30-40 Software Engineers
Housed at Calit2@UCSD

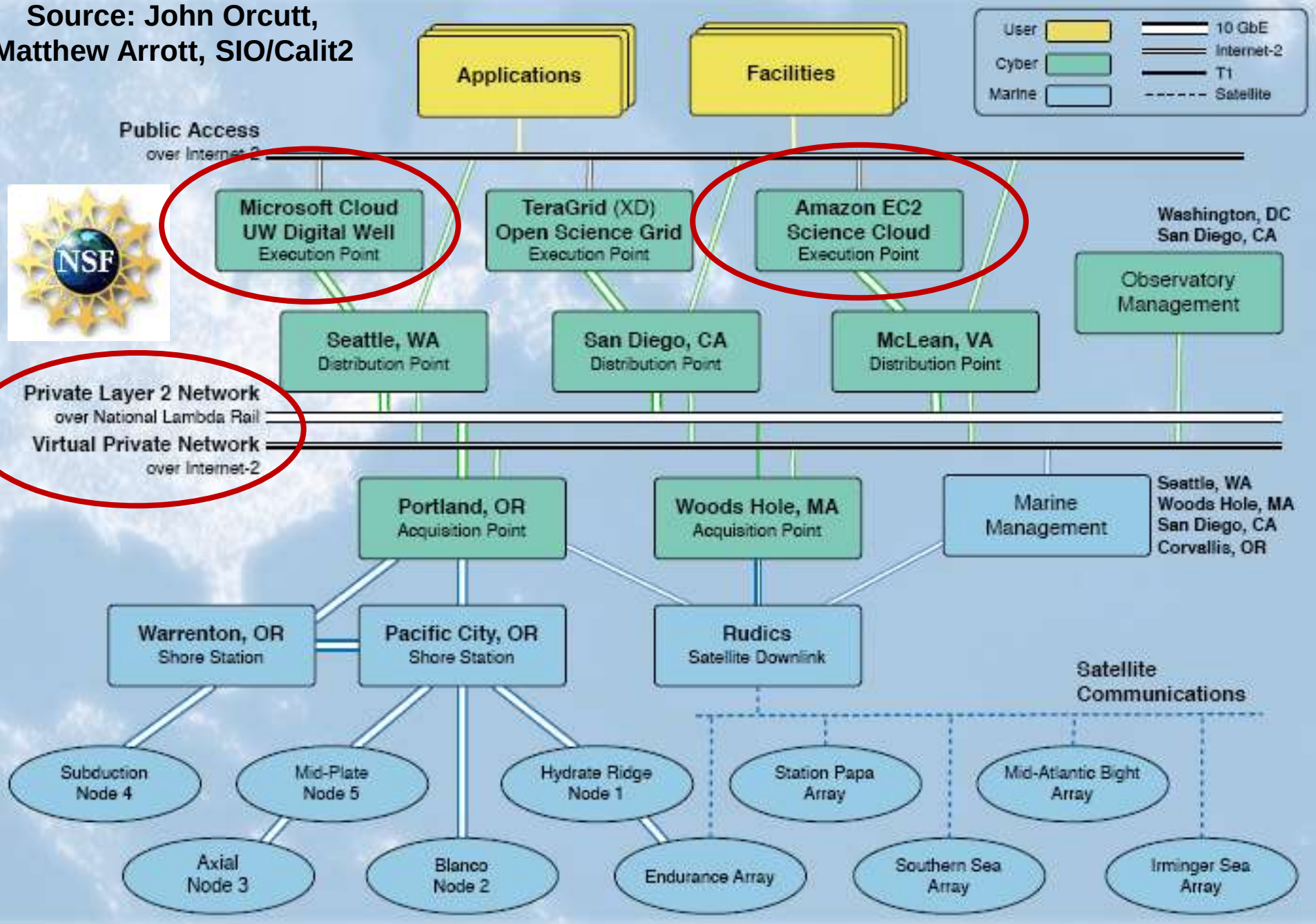


Source: Matthew Arrott, Calit2 Program Manager for OOI CI

OOI CI is Built on Dedicated Optical Infrastructure Using Clouds

OCEAN OBSERVATORIES INITIATIVE

Source: John Orcutt, Matthew Arrott, SIO/Calit2



California and Washington Universities Are Testing a 10Gbps Connected Commercial Data Cloud

- **Amazon Experiment for Big Data**
  - **Only Available Through CENIC & Pacific NW GigaPOP**
    - **Private 10Gbps Peering Paths**
  - **Includes Amazon EC2 Computing & S3 Storage Services**
- **Early Experiments Underway**
  - **Robert Grossman, Open Cloud Consortium**
  - **Phil Papadopoulos, Calit2/SDSC Rocks**

amazon web services™

cenic

PACIFIC NORTHWEST GIGAPOP

it²

Open Cloud OptIPuter Testbed--Manage and Compute Large Datasets Over 10Gbps Lambdas

Calit2

CENIC

NLR

CISCO

NLR C-Wave

JOHNS HOPKINS UNIVERSITY

Dragon



STARLIGHT SM
The Optical STAR TAP SM

MREN

UIC

- 9 Racks
- 500 Nodes
- 1000+ Cores
- 10+ Gb/s Now
- Upgrading Portions to 100 Gb/s in 2010/2011

Open Source SW

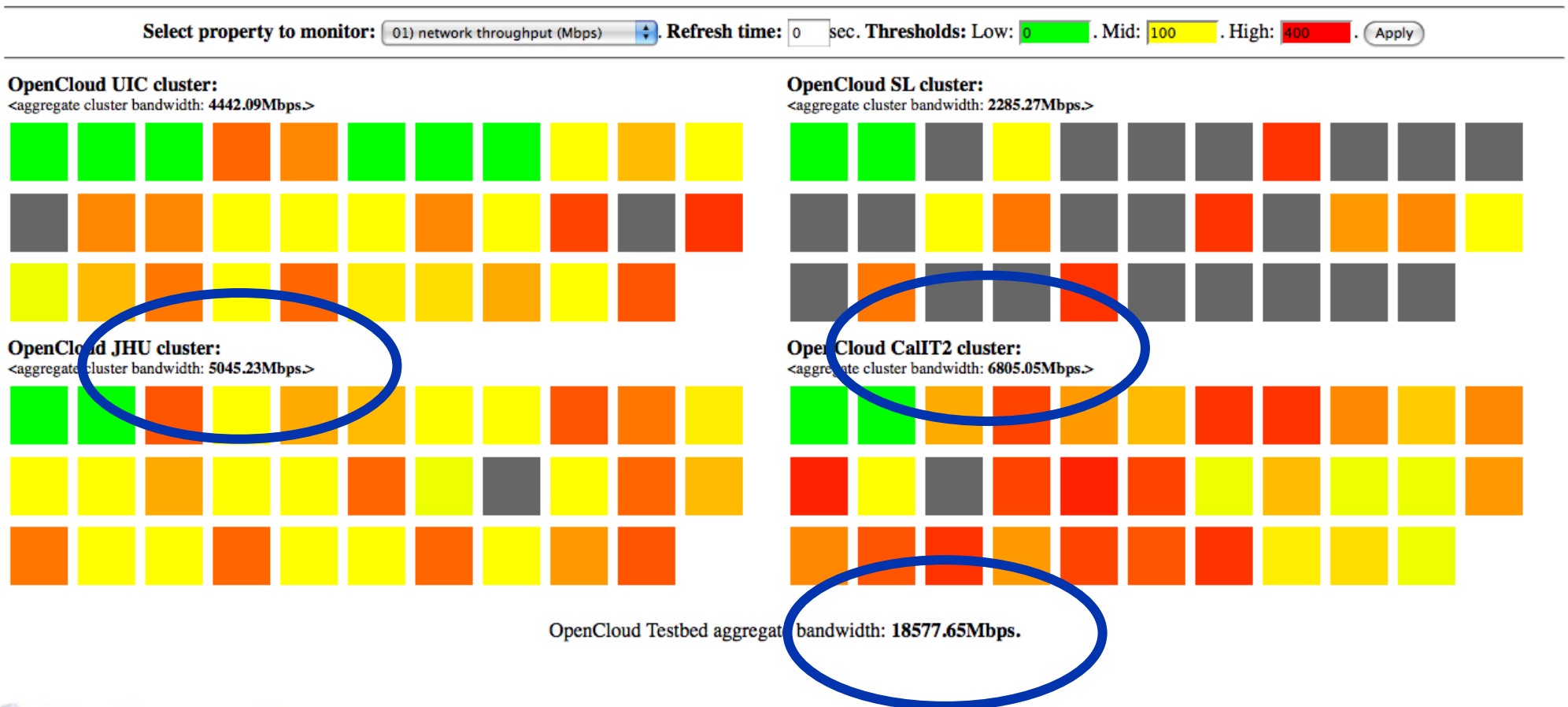
- Hadoop
- Sector/Sphere
- Nebula
- Thrift, GPB
- Eucalyptus
- Benchmarks

Open Cloud Consortium

Source: Robert Grossman, UChicago

Terasort on Open Cloud Testbed Sustains >5 Gbps--Only 5% Distance Penalty!

Sorting 10 Billion Records (1.2 TB) at 4 Sites (120 Nodes)

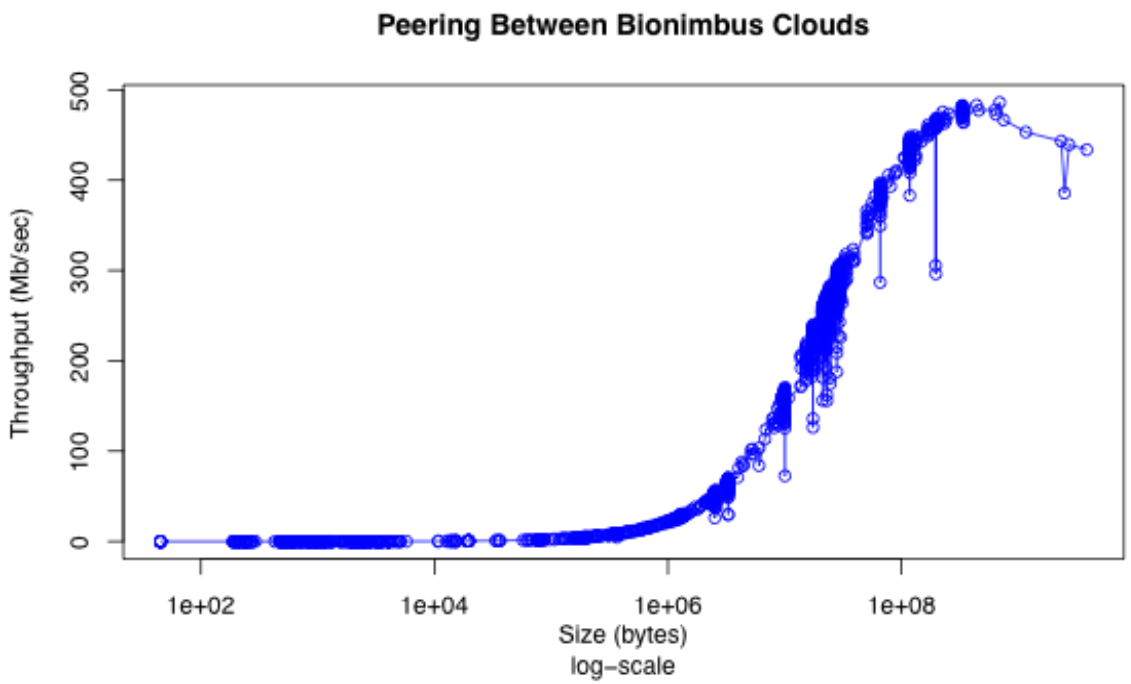


Source: Robert Grossman, UChicago

Hybrid Cloud Computing with modENCODE Data

- Computations in Bionimbus Can Span the Community Cloud & the Amazon Public Cloud to Form a Hybrid Cloud
- Sector was used to Support the Data Transfer between Two Virtual Machines
  - One VM was at UIC and One VM was an Amazon EC2 Instance
- Graph Illustrates How the Throughput between Two Virtual Machines in a Wide Area Cloud Depends upon the File Size





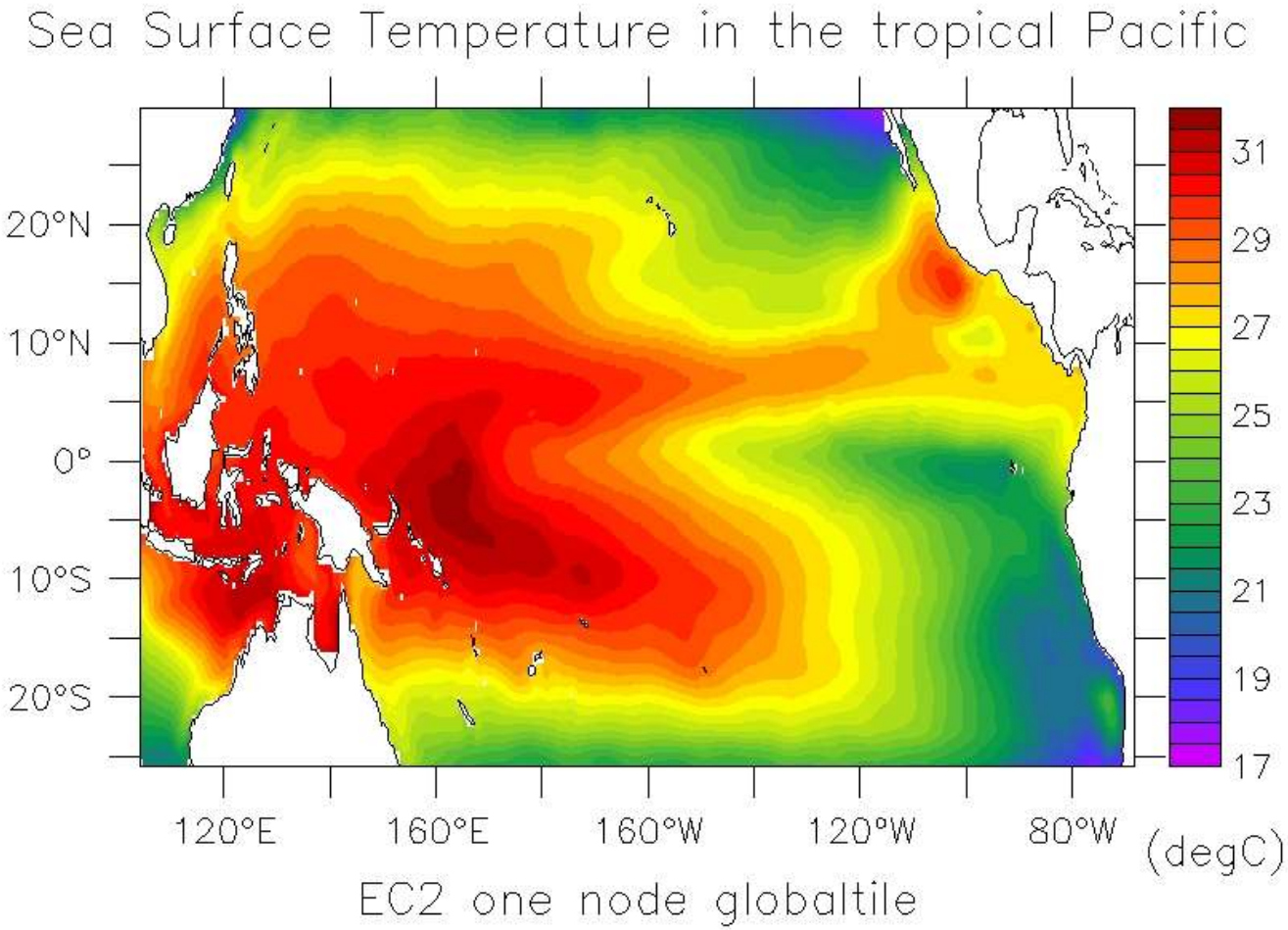


Biological data (Bionimbus)

Source: Robert Grossman, UChicago

Ocean Modeling HPC In the Cloud: Tropical Pacific SST (2 Month Ave 2002)

MIT GCM 1/3 Degree Horizontal Resolution, 51 Levels, Forced by NCEP2. Grid is 564x168x51, Model State is T,S,U,V,W and Sea Surface Height

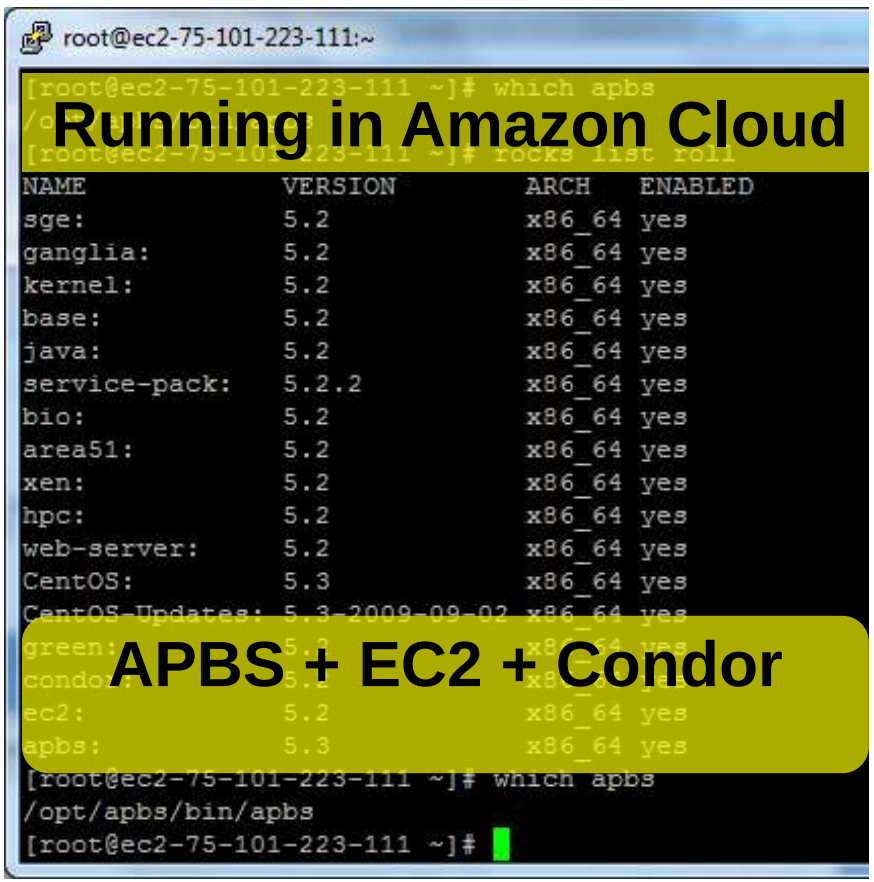


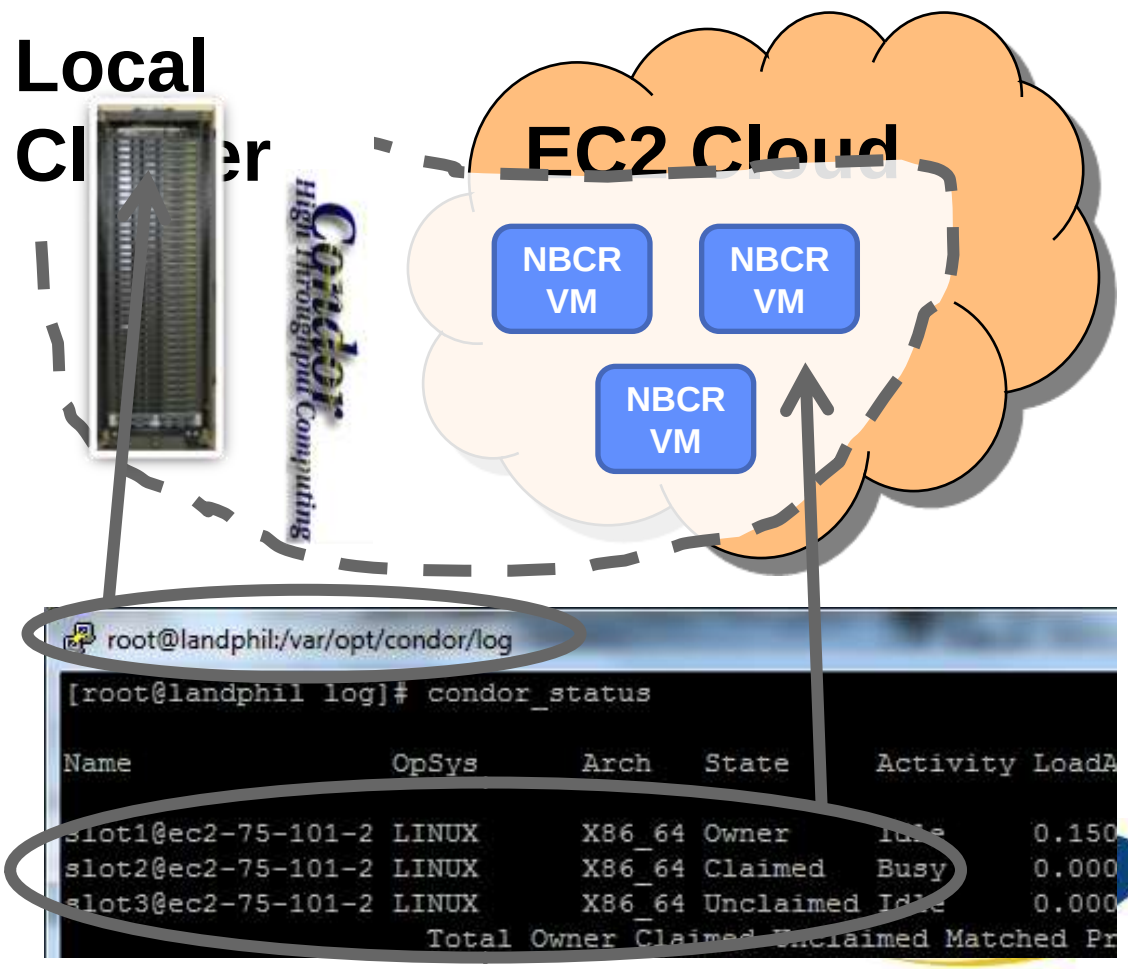
Run on EC2 HPC Instance. In Collaboration with OOI CI/Calit2

Source: B. Cornuelle, N. Martinez, C.Papadopoulos COMPAS, SIO

Using Condor and Amazon EC2 on Adaptive Poisson-Boltzmann Solver (APBS)

- APBS Rocks Roll (NBCR) + EC2 Roll + Condor Roll = Amazon VM
- Cluster extension into Amazon using Condor





SDSC

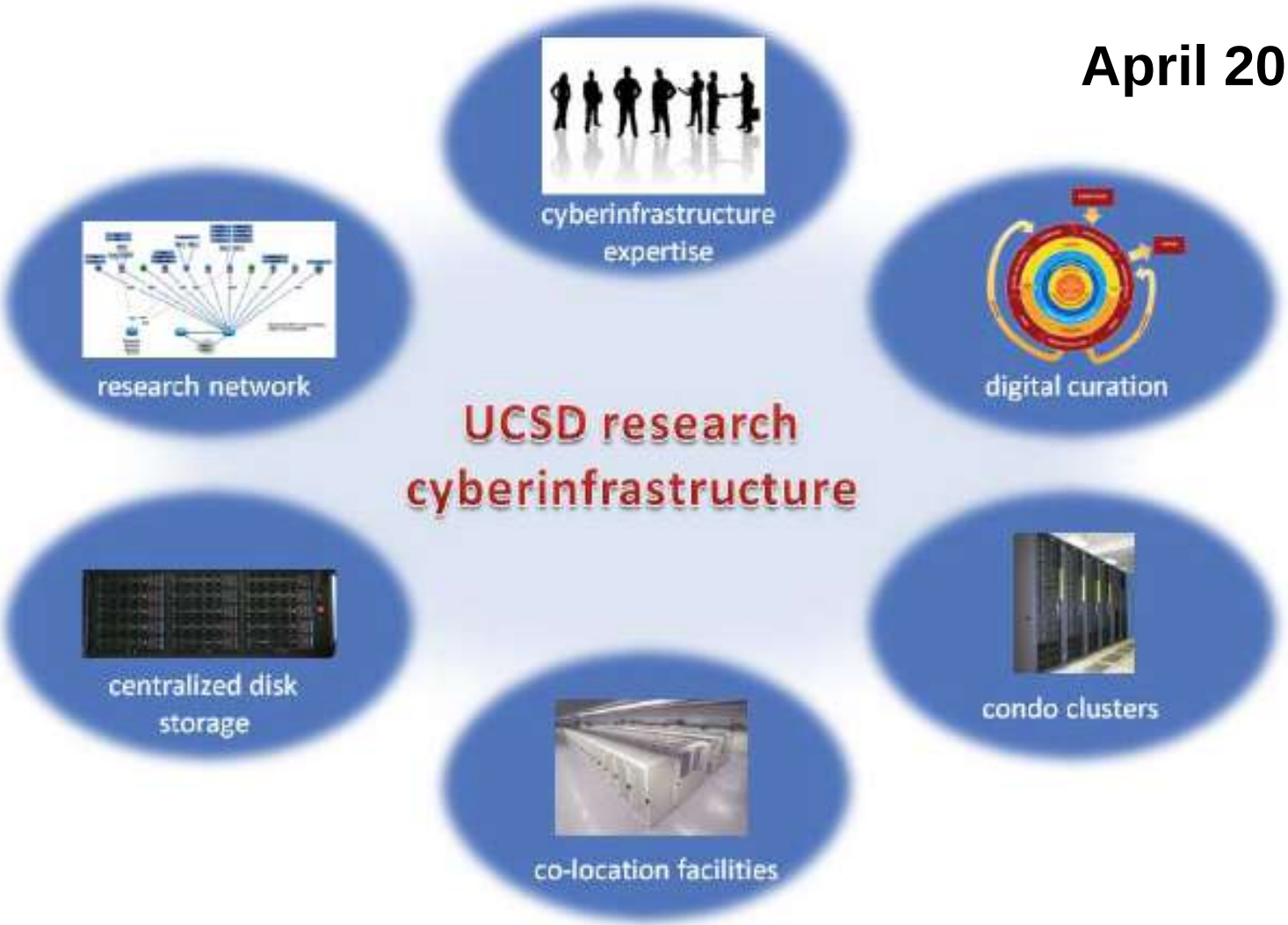
Source: Phil Papadopoulos, SDSC/Calit2

"Blueprint for the Digital University"--Report of the UCSD Research Cyberinfrastructure Design Team

- **Focus on Data-Intensive Cyberinfrastructure**

April 2009

No Data Bottlenecks --Design for Gigabit/s Data Flows



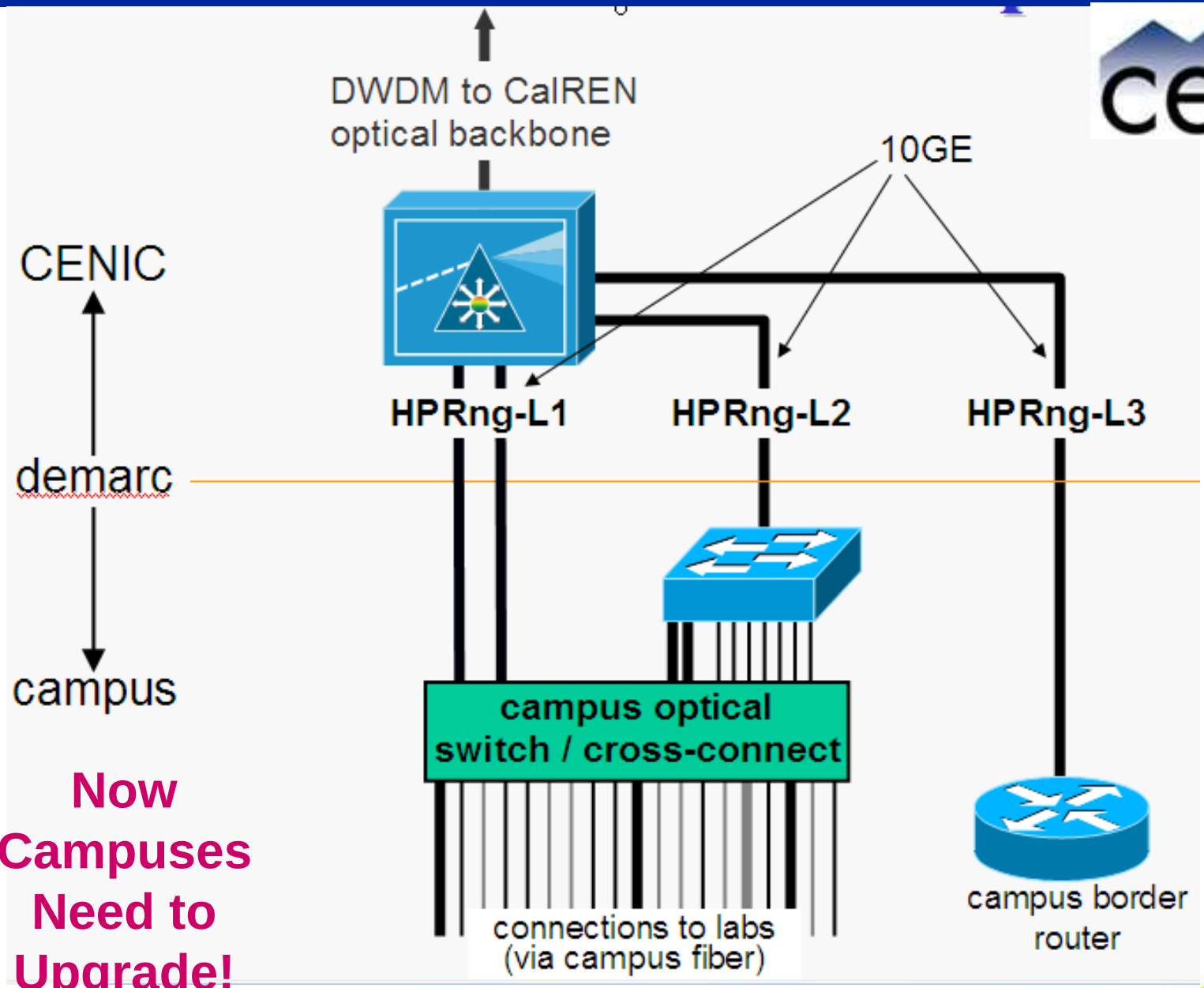
http://research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf

SDSC

it²

What do Campuses Need to Build to Utilize CENIC's Three Layer Network?

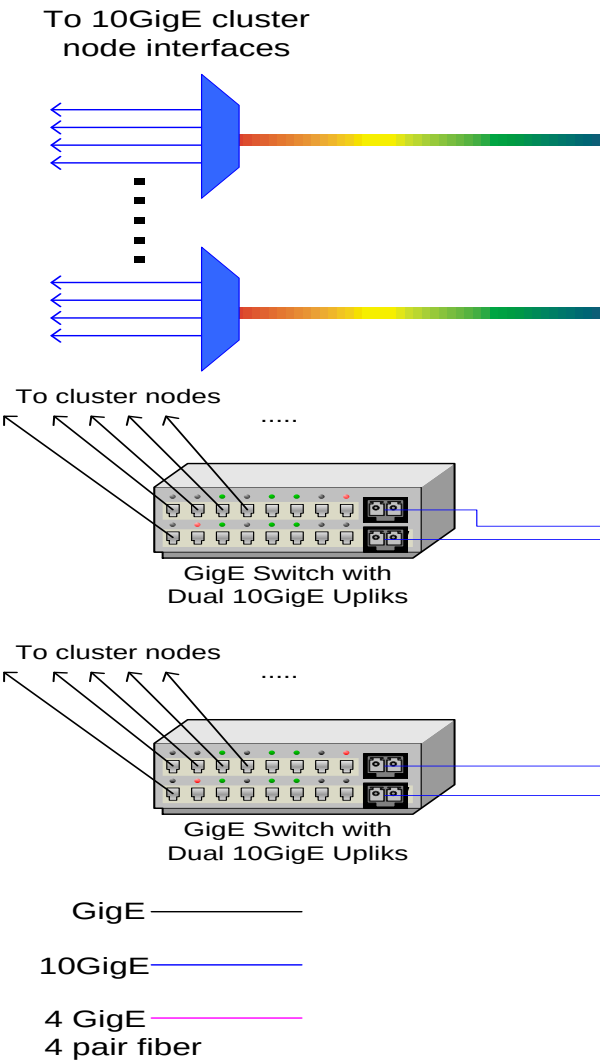
~ $14M Invested in Upgrade



Now Campuses Need to Upgrade!

Source: Jim Dolgonas, CENIC

Current UCSD Optical Core: Bridging End-Users to CENIC L1, L2, L3 Services



Enpoints:

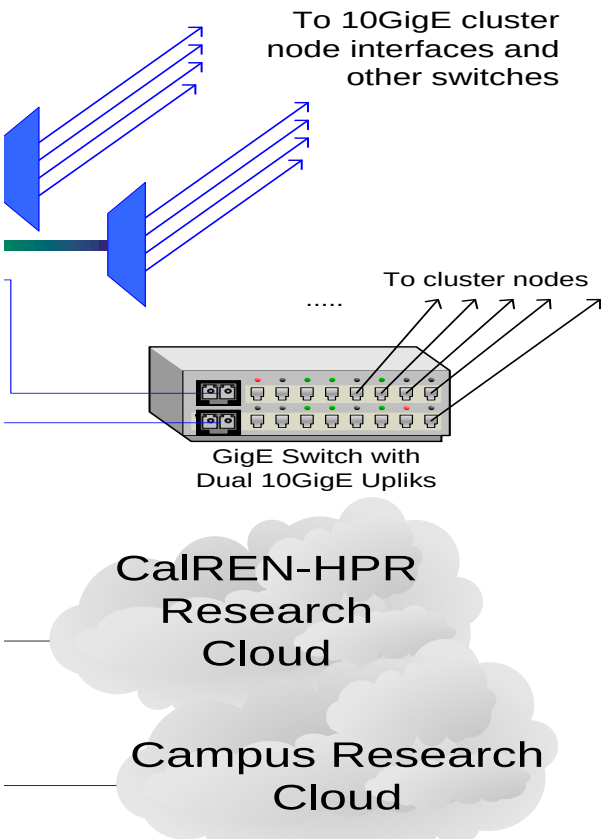
>= 60 endpoints at 10 GigE

>= 32 Packet switched

>= 32 Switched wavelengths

>= 300 Connected endpoints

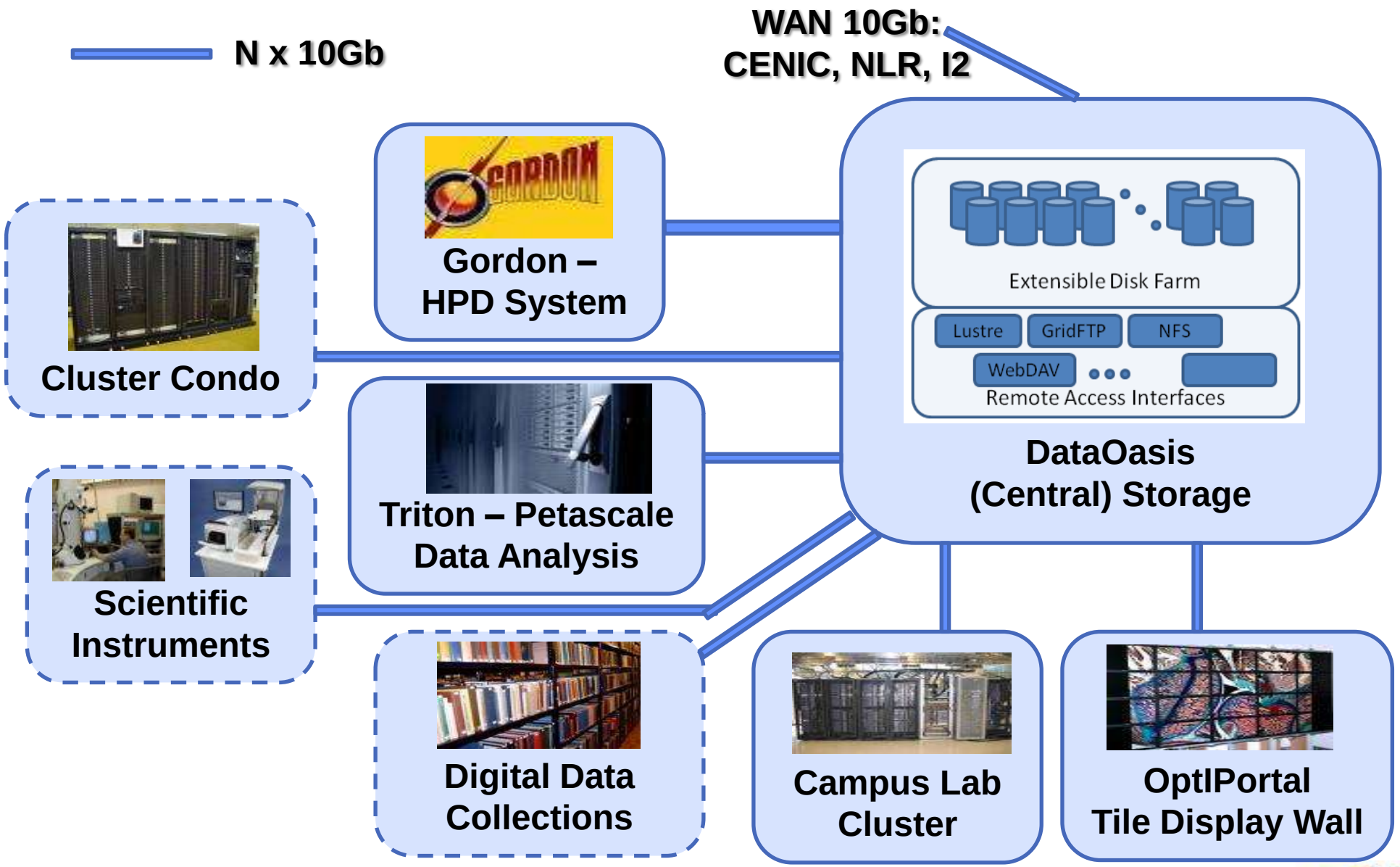
Approximately 0.5 TBit/s Arrive at the "Optical" Center of Campus. Switching is a Hybrid of: Packet, Lambda, Circuit -- OOO and Packet Switches



Source: Phil Papadopoulos, SDSC/Calit2 (Quartzite PI, OptIPuter co-PI)

Quartzite Network MRI #CNS-0421555; OptIPuter #ANI-0225642

UCSD Campus Investment in Fiber Enables Consolidation of Energy Efficient Computing & Storage



Source: Philip Papadopoulos, SDSC/Calit2

The GreenLight Project: Instrumenting the Energy Cost of Computational Science

- **Focus on 5 Communities with At-Scale Computing Needs:**
  - **Metagenomics**
  - **Ocean Observing**
  - **Microscopy**
  - **Bioinformatics**
  - **Digital Media**





- **Measure, Monitor, & Web Publish Real-Time Sensor Outputs**
  - **Via Service-oriented Architectures**
  - **Allow Researchers Anywhere To Study Computing Energy Cost**
  - **Enable Scientists To Explore Tactics For Maximizing Work/Watt**
- **Develop Middleware that Automates Optimal Choice of Compute/RAM Power Strategies for Desired Greenness**
- **Partnering With Minority-Serving Institutions Cyberinfrastructure Empowerment Coalition**

NSF

it²

Source: Tom DeFanti, Calit2; GreenLight PI

UCSD Biomed Centers Drive High Performance CI





National Center for Research Resources
National Institutes of Health



National Centers for Biomedical Computing

National Resource for Network Biology

iDASH: Integrating Data for Analysis, Anonymization, and Sharing





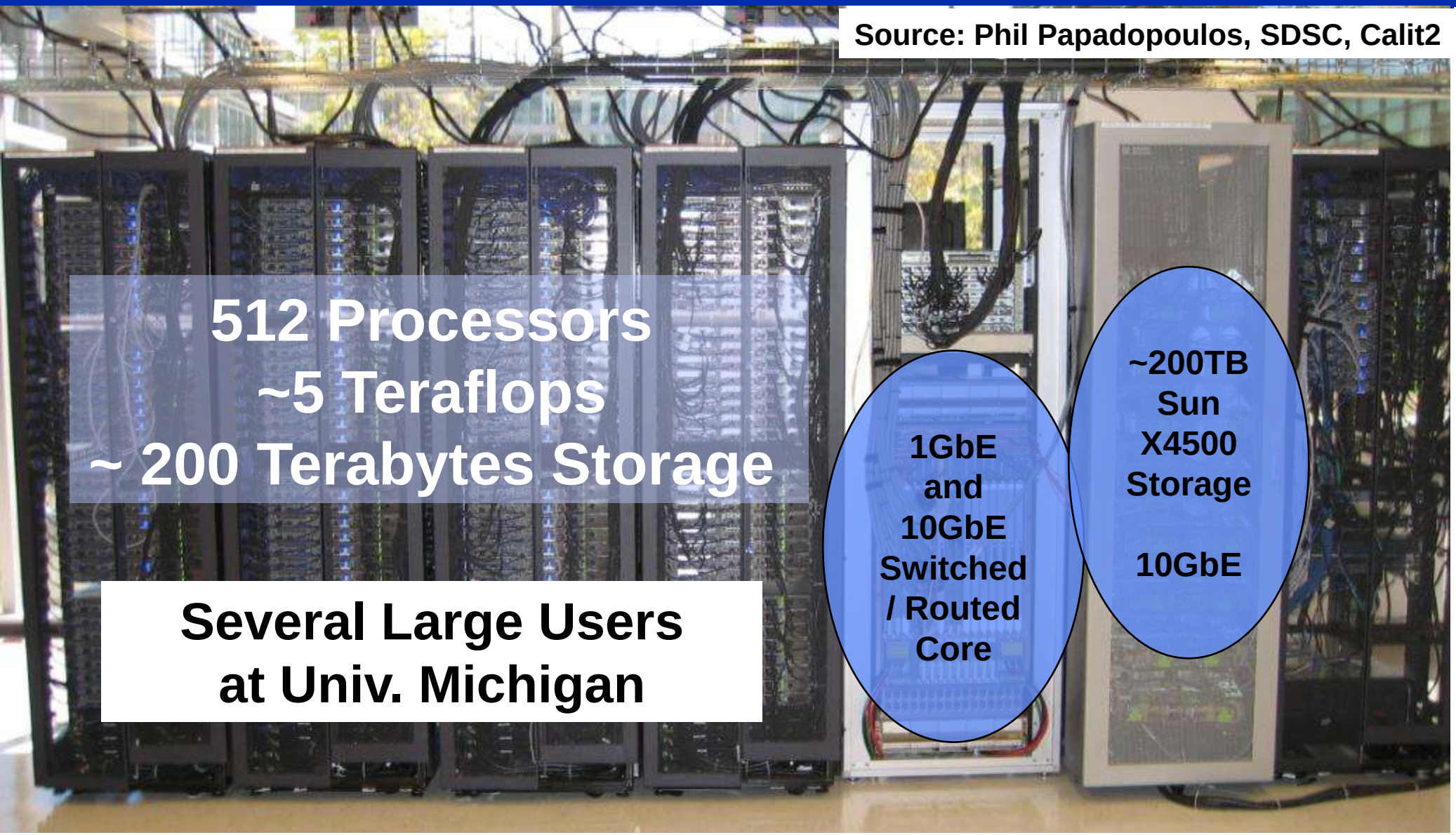








Calit2 Microbial Metagenomics Cluster- Next Generation Optically Linked Science Data Server



4000 Users
From 90 Countries

Gordon and Betty MOORE FOUNDATION

ROCKS

SDSC

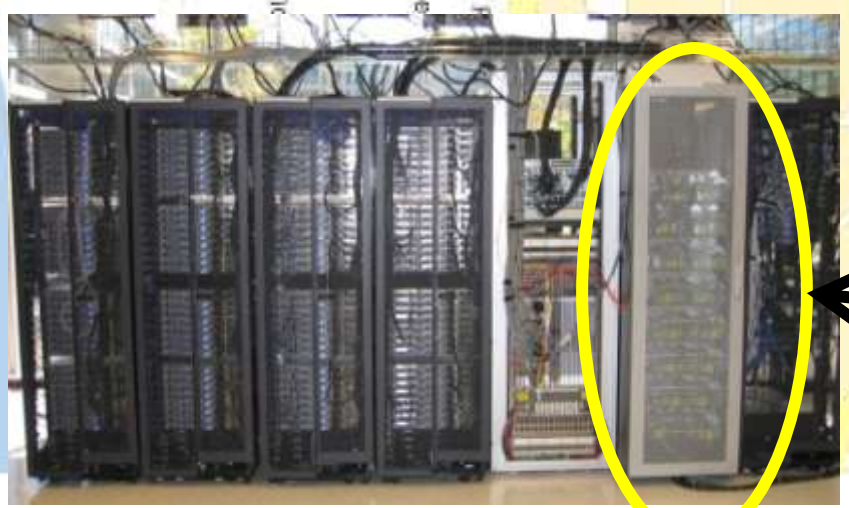
J. Craig Venter INSTITUTE

it²

Calit2 CAMERA Automatic Overflows into SDSC Triton



@ CALIT2



@ SDSC

Triton Resource

CAMERA - Managed Job Submit Portal (VM)

Transparently Sends Jobs to Submit Portal on Triton

10Gbps

CAMERA DATA

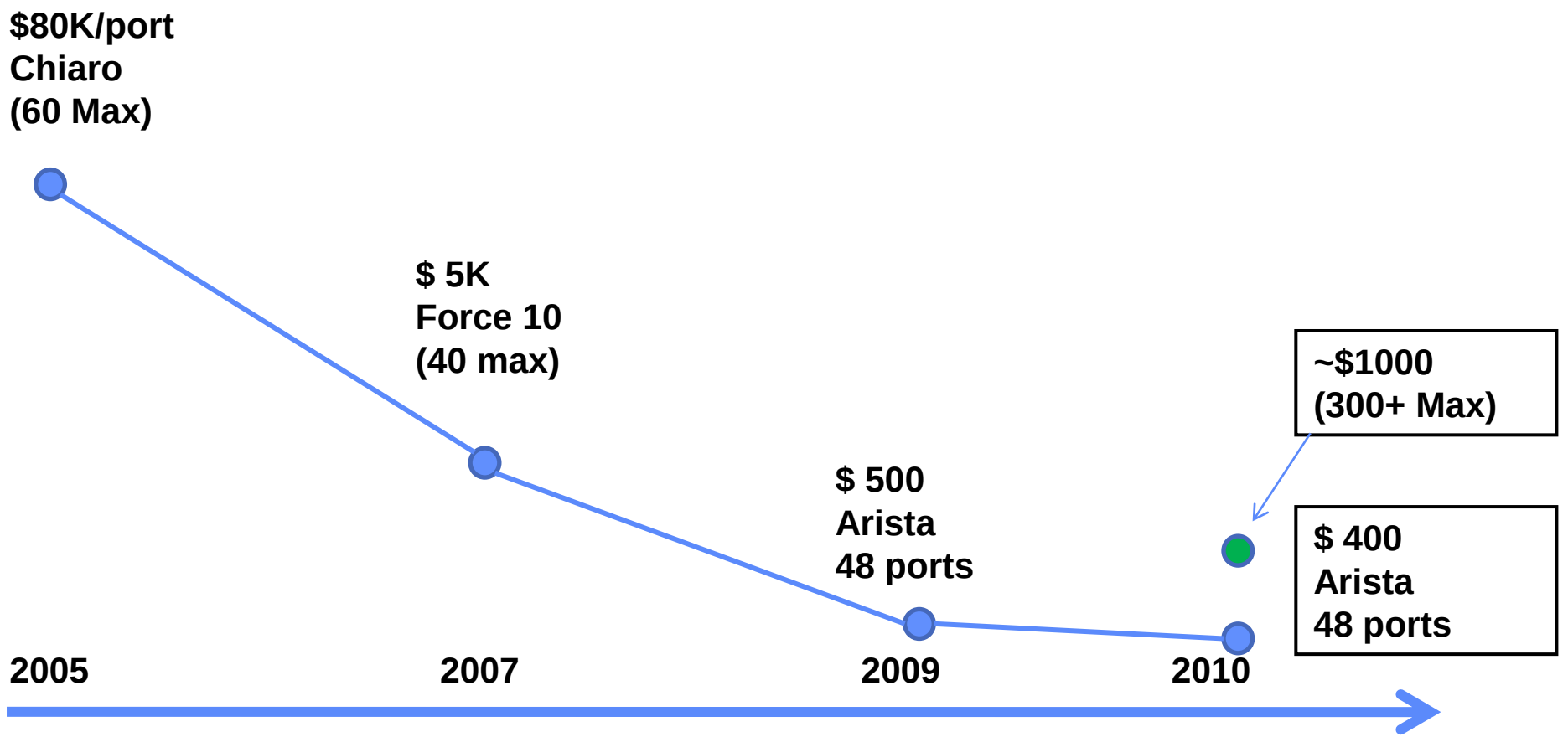
Direct Mount == No Data Staging

SDSC

ROCKS

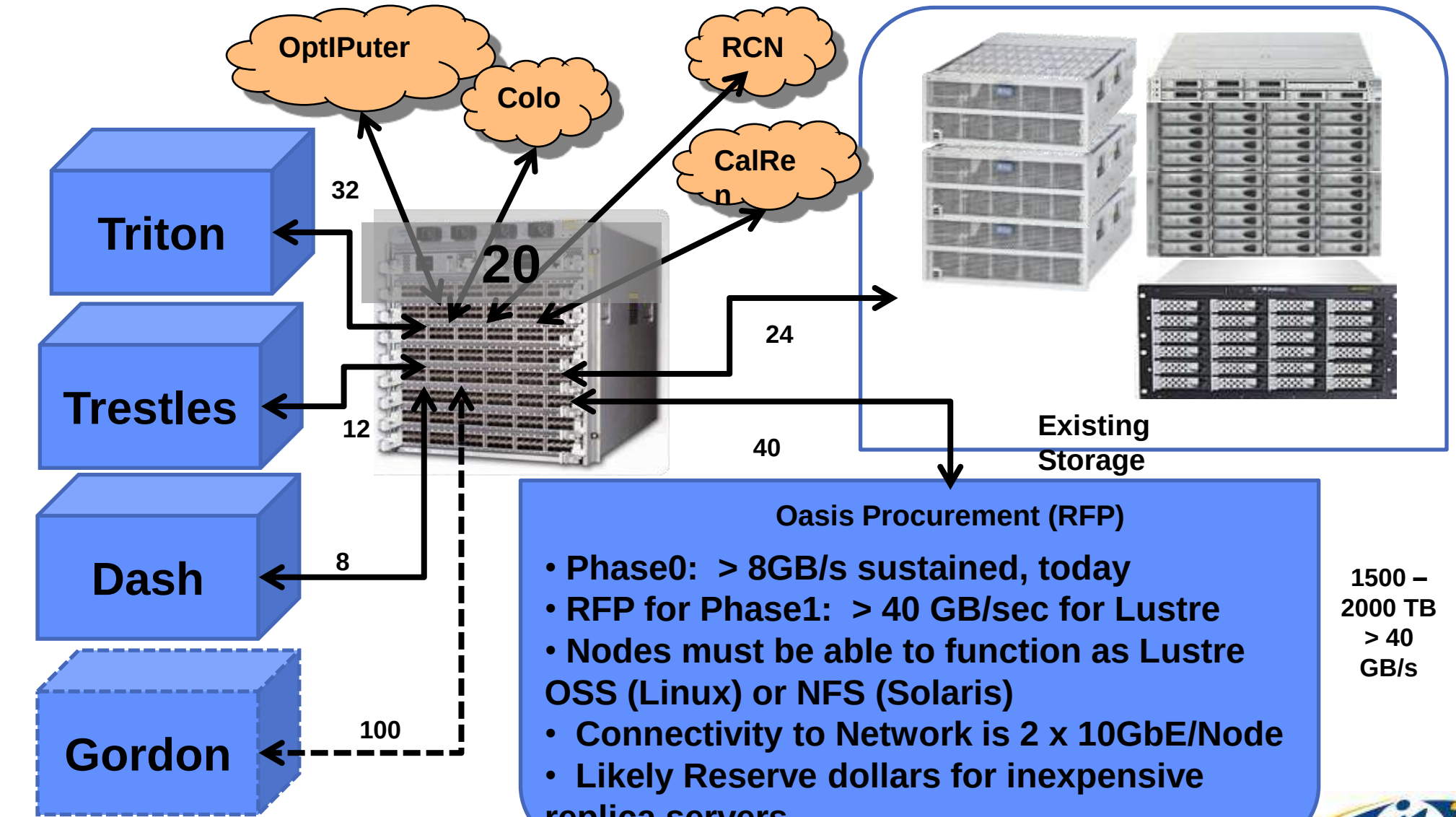
Rapid Evolution of 10GbE Port Prices Makes Campus-Scale 10Gbps CI Affordable

- **Port Pricing is Falling**
- **Density is Rising – Dramatically**
- **Cost of 10GbE Approaching Cluster HPC Interconnects**



Source: Philip Papadopoulos, SDSC/Calit2

10G Switched Data Analysis Resource: SDSC's Data Oasis



Source: Philip Papadopoulos, SDSC/Calit2

NSF Funds a Data-Intensive Track 2 Supercomputer: SDSC's Gordon-Coming Summer 2011

- **Data-Intensive Supercomputer Based on SSD Flash Memory and Virtual Shared Memory SW**
  - ***Emphasizes MEM and IOPS over FLOPS***
  - ***Supernode has Virtual Shared Memory:***
    - ***2 TB RAM Aggregate***
    - ***8 TB SSD Aggregate***
    - ***Total Machine = 32 Supernodes***
    - ***4 PB Disk Parallel File System >100 GB/s I/O***



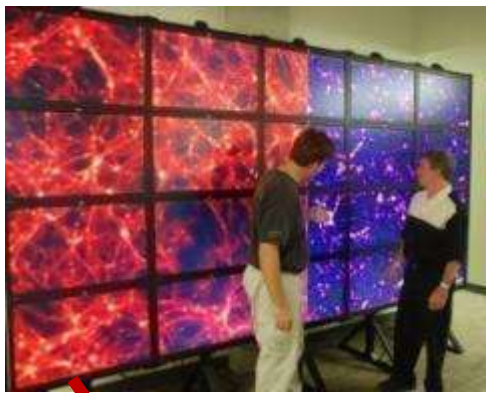
- **System Designed to Accelerate Access to Massive Data Bases being Generated in all Fields of Science, Engineering, Medicine, and Social Science**

SDSC

Source: Mike Norman, Allan Snavely SDSC

it²

Academic Research "OptIPlatform" Cyberinfrastructure: A 10Gbps "End-to-End" Lightpath Cloud



End User OptIPortal



Instruments



HD/4k Telepresence



HD/4k Video Cams

HPC



10G Lightpaths

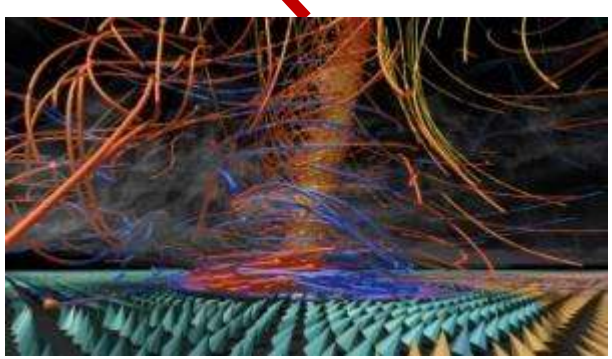


TeraGrid

Campus Optical Switch



Data Repositories & Clusters



HD/4k Video Images

You Can Download This Presentation at lsmarr.calit2.net



ABOUT LARRY SMARR

Larry Smarr is the founding Director of the California Institute for Telecommunications and Information Technology (Calit2), a UC San Diego/UC Irvine partnership, and holds the Harry E. Gruber professorship in Computer Science and Engineering (CSE) at UCSD's Jacobs School. At Calit2, Smarr has

WHAT I'M READING

Complete List of Larry's Recommendations

Homo Britannicus: The Incredible Story of Human Life in Britain (2006)

American College and University

RECENT TWITTER POSTS

Twitter? Click To Follow Me

Ismarr: SDSC Founding Director Sid Karin and I are making a few comments on the history of the NSF Supercomputing Centers at