“Peering The Pacific Research Platform With The Great Plains Network”

Keynote
Great Plains Network 2018 Annual Meeting
Kansas City, MO
May 31, 2018

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
http://lsmarr.calit2.net

PRP PACIFIC RESEARCH PLATFORM

it²

I Was Born and Raised in the Midwest: Columbia, Missouri



My Mother and Me
On My First Birthday



My Grandfather, Father, and Me
At My MU Graduation

PRP
PACIFIC RESEARCH PLATFORM

I Earned Three of the Sixteen University of Missouri Degrees in My Family



Framing
By Brother
David Smarr

PRP
PACIFIC RESEARCH PLATFORM

it²

30 Years Ago NSF Brought to University Researchers a DOE HPC Center Model

1985/6







NCSA Was Modeled on LLNL

SDSC Was Modeled on MFEnet

PRP PACIFIC RESEARCH PLATFORM

NSF

it²

Thirty Years After NSF Adopts DOE Supercomputer Center Model
NSF Adopts DOE ESnet's Science DMZ for High Performance Applications

- **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 as data transfer nodes (DTNs)**
  - **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**

Science DMZ Coined 2010

The DOE ESnet Science DMZ and the NSF "Campus Bridging" Taskforce Report Formed the Basis for the NSF Campus Cyberinfrastructure Network Infrastructure and Engineering (CC-NIE) Program

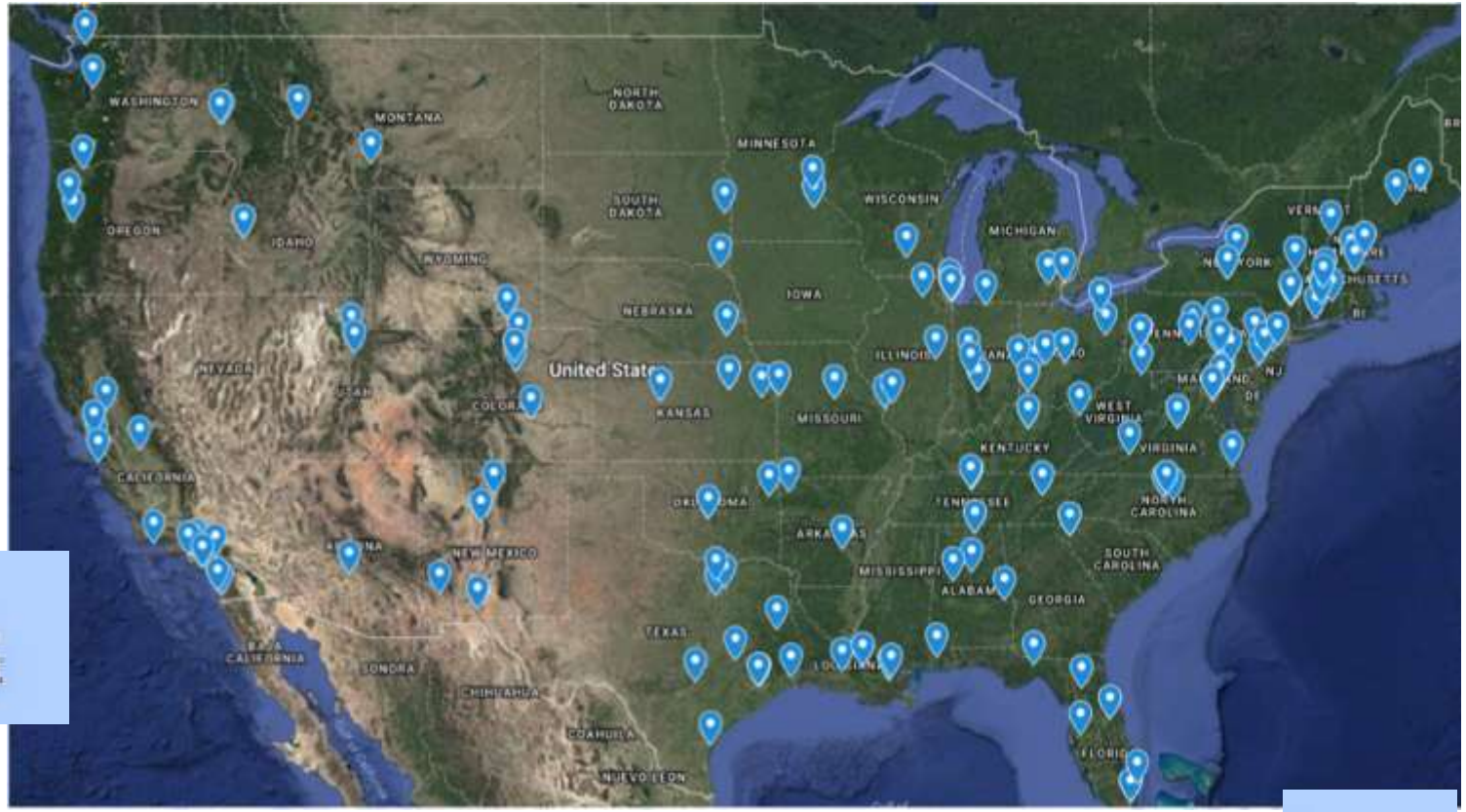


PRP PACIFIC RESEARCH PLATFORM

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



Based on Community Input and on ESnet's Science DMZ Concept, NSF Has Made Over 200 Campus-Level Awards in 44 States







2012-2017 CC-NIE / CC*IIE / CC*DNI PROGRAMS

Source: Kevin Thompson, NSF

PRP
PACIFIC RESEARCH PLATFORM

Logical Next Step: The Pacific Research Platform Networks Campus DMZs to Create a Regional End-to-End Science-Driven "Big Data Superhighway" System



PACIFIC RESEARCH PLATFORM

Source: John Hess, CENIC

NSF CC*DNI Grant
$5M 10/2015-10/2020

PI: Larry Smarr, UC San Diego Calit2

Co-PIs:

- **Camille Crittenden, UC Berkeley CITRIS,**
- **Tom DeFanti, UC San Diego Calit2/QI,**
- **Philip Papadopoulos, UCSD SDSC,**
- **Frank Wuerthwein, UCSD Physics and SDSC**

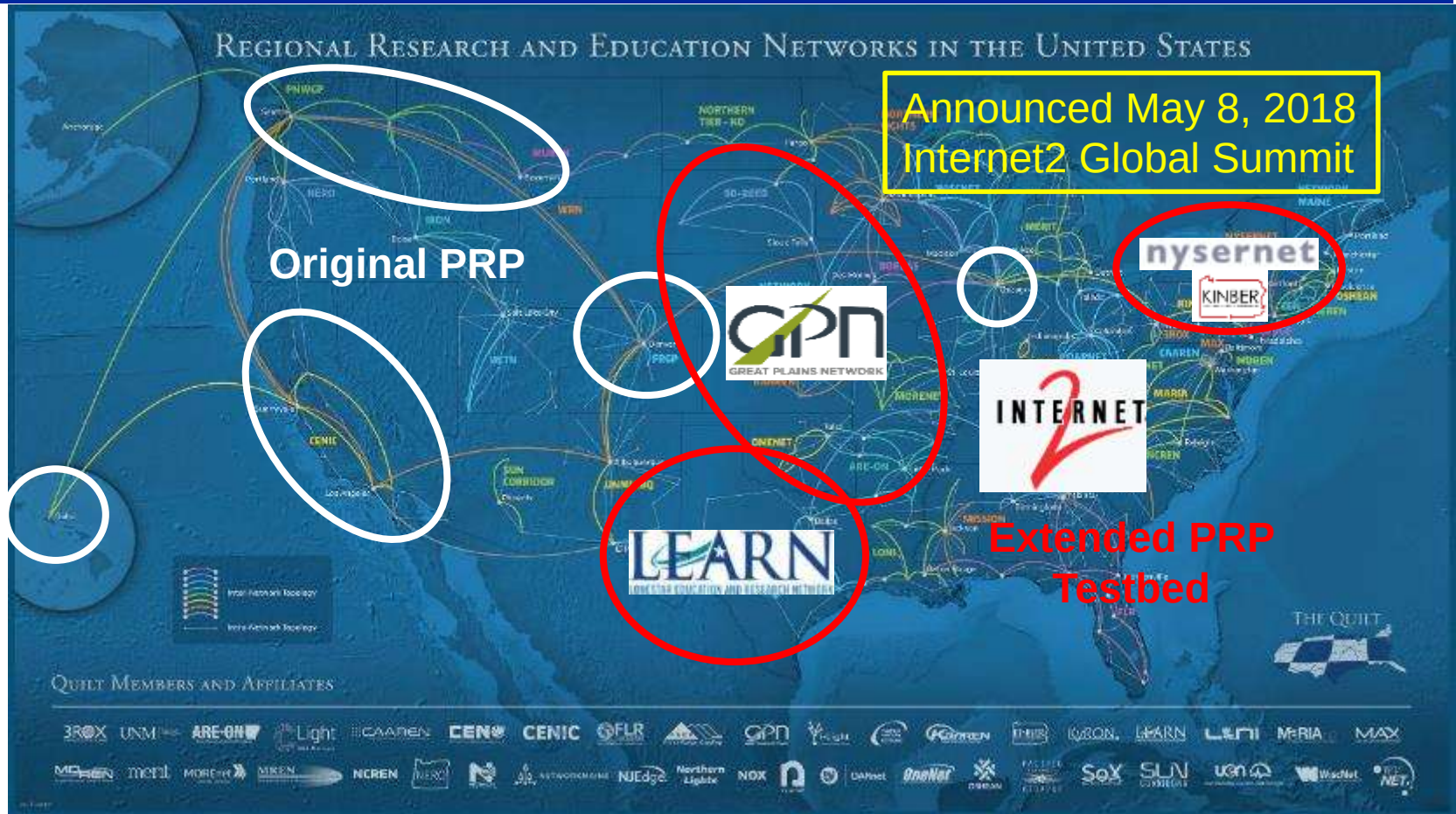
Letters of Commitment from:

- 50 Researchers from 15 Campuses
- 32 IT/Network Organization Leaders

NSF Program Officer: Amy Walton

SDSC CITRIS BANATAO INSTITUTE it²

PRP National-Scale Experimental Distributed Testbed: Using Internet2 to Connect Early-Adopter Quilt Regional R&E Networks



NSF

ESnet

CENIC

SDSC

CITRIS BANATAO INSTITUTE

it²

PRP PACIFIC RESEARCH PLATFORM

PRP Science DMZ Data Transfer Nodes (DTNs) - Flash I/O Network Appliances (FIONAs)

UCSD Designed FIONAs
To Solve the Disk-to-Disk
Data Transfer Problem
at Full Speed
on 10G, 40G and 100G Networks

FIONette—1G, $250





Five Racked FIONAs at Calit2

- Each Contains:
  - Dual 12-Core CPUs
  - 96GB RAM
  - 1TB SSD
  - 2 10GbE interfaces
    - Total ~$10,500
  - With 8 GPUs
    - total ~$18,500

PACIFIC RESEARCH PLATFORM

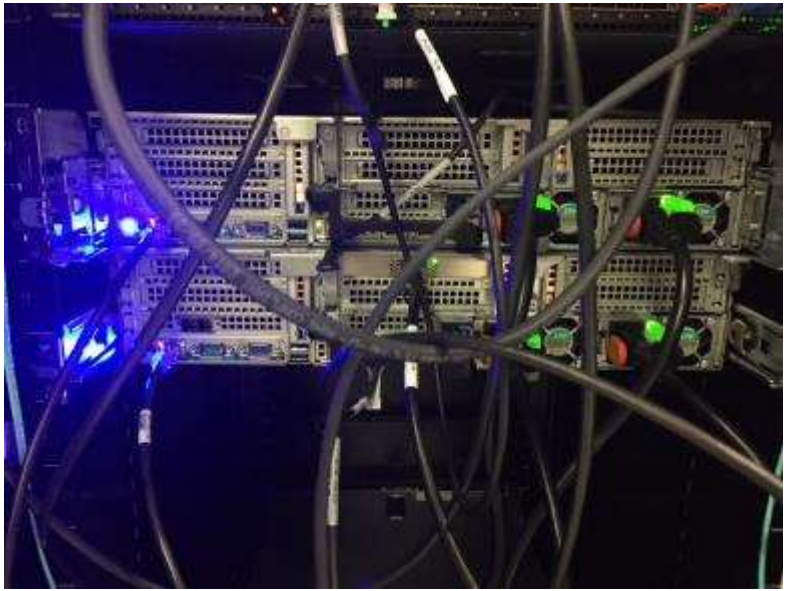
SDSC

Phil Papadopoulos, SDSC &
Tom DeFanti, Joe Keefe & John Graham, Calit2

GPN Becomes the First Multi-State Regional Network to Peer with the PRP

Seeing 5Gbs Between
the PRP-Contributed PWave DTN in Los Angeles
To GPN FIONA in UMC





Source: John Hess, CENIC and George Rob III, UMissouri

Game Changer: Using Kubernetes to Manage Containers Across the PRP

CADE METZ BUSINESS 06.10.14 01:15 PM

GOOGLE OPEN SOURCES ITS SECRET WEAPON IN CLOUD COMPUTING

WIRED

"Kubernetes is a way of stitching together
a collection of machines into, basically, a big computer,"
--Craig Mcluckie, Google
and now CEO and Founder of Heptio

"Everything at Google runs in a container."
--Joe Beda,Google



"Kubernetes has emerged as
the container orchestration engine of choice
for many cloud providers including
Google, AWS, Rackspace, and Microsoft,
and is now being used in HPC and Science DMZs.
--John Graham, Calit2/QI UC San Diego

PRP
PACIFIC RESEARCH PLATFORM

CLOUD NATIVE COMPUTING FOUNDATION

Rook is Ceph Cloud-Native Object Storage ‘Inside’ Kubernetes



Battle-tested, production storage

Rook is based on an embedded version of Ceph, which has 10+ years of production deployments and runs some of the worlds largest clusters.

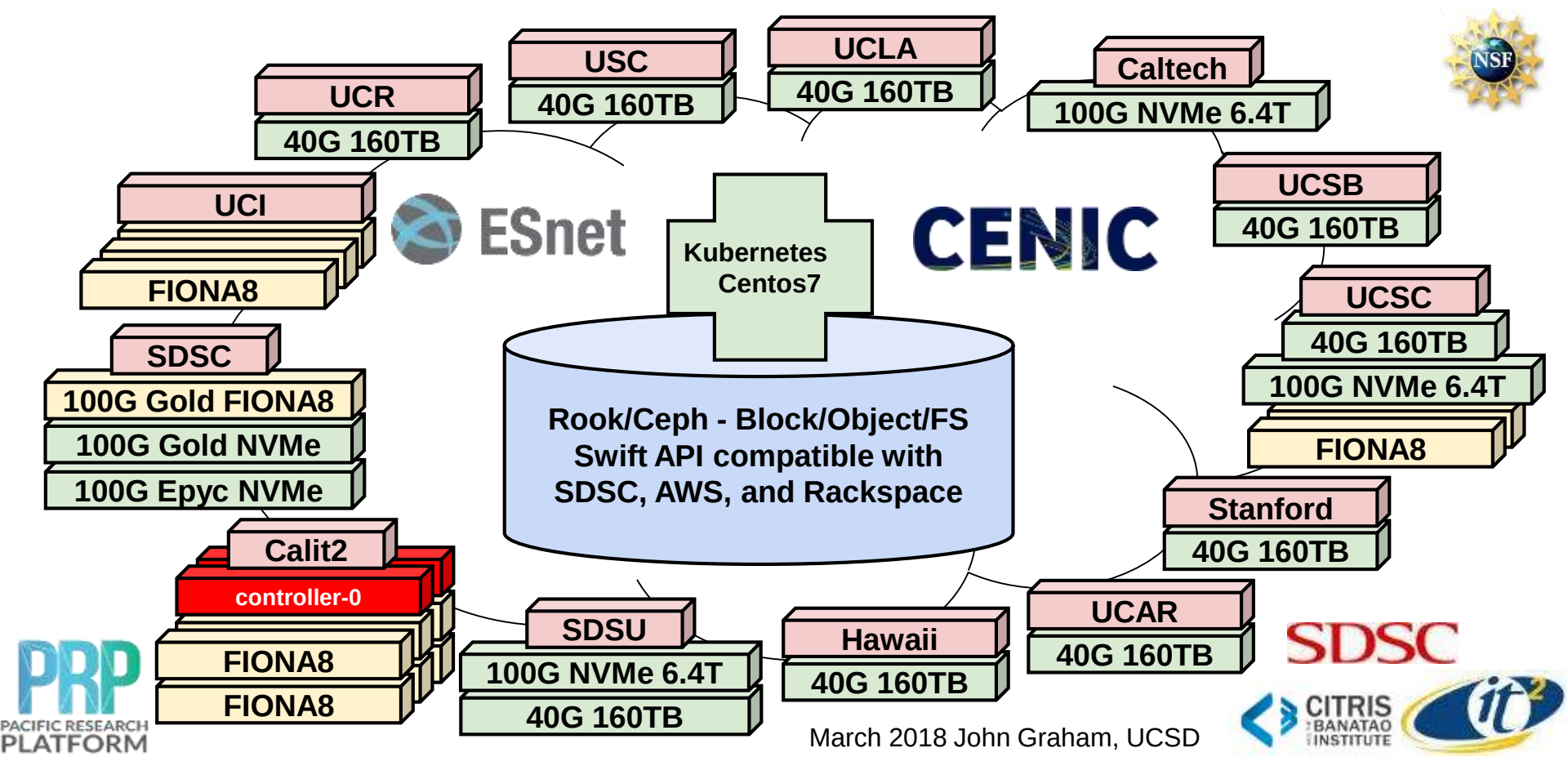
Cloud-native environment integration

Rook runs as a cloud-native service for optimal integration with applications in need of block, object, or file storage.

PACIFIC RESEARCH PLATFORM

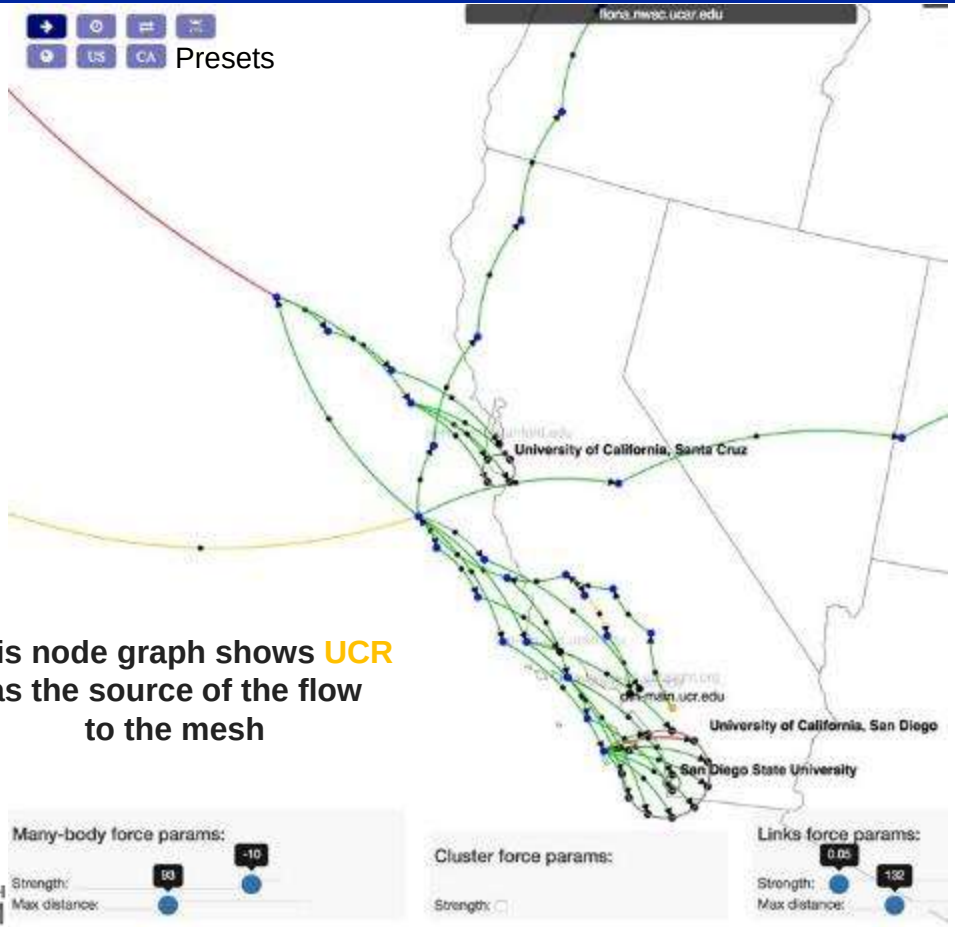
Source: John Graham, Calit2/QI

Running Kubernetes/Rook/Ceph On PRP
Allows Us to Deploy a Distributed PB+ of Storage for Posting Science Data



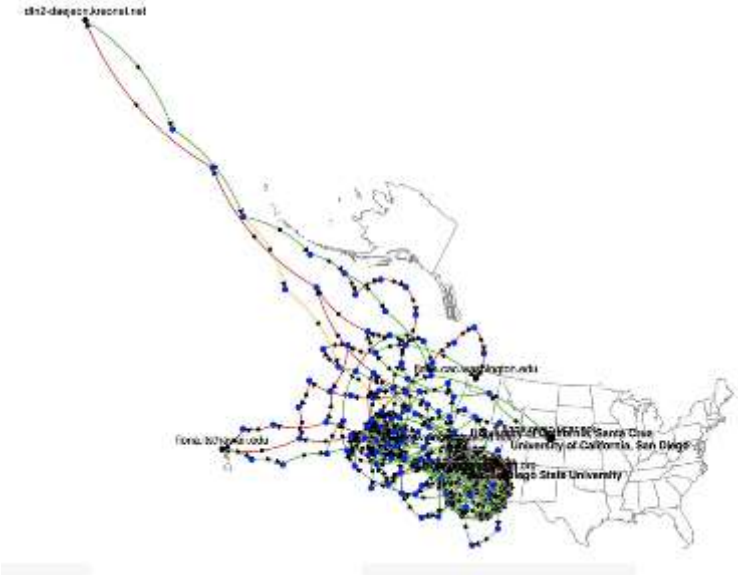
March 2018 John Graham, UCSD

Operational Metrics: Containerized Trace Route Tool Allows Realtime Visualization of Status of Network Links



This node graph shows UCR as the source of the flow to the mesh

Source: Dmitry Mishin(SDSC), John Graham (Calit2)



All Kubernetes Nodes on PRP

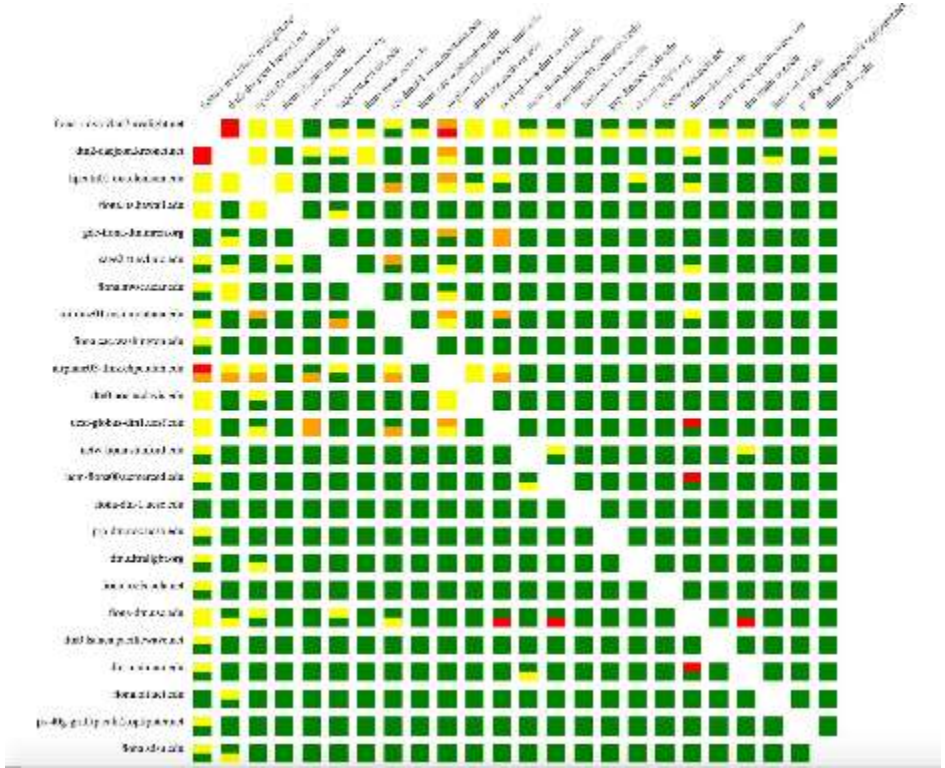
PACIFIC RESEARCH PLATFORM

We Measure Disk-to-Disk Throughput with 10GB File Transfer Using Globus GridFTP 4 Times Per Day in Both Directions for All PRP Sites

PRPGridFTP

Throughput >= 5000Mbps | Throughput < 5000Mbps | Throughput <= 1000Mbps | Unable to retrieve data

April 24, 2017



Source: John Graham, Calit2

PRP PACIFIC RESEARCH PLATFORM

ESnet CENIC CITRIS BANATAO INSTITUTE SDSC it²

PRP's First 2.5 Years:
Connecting Multi-Campus Application Teams and Devices



CMS

Particle Physics

Biomedical 'omics

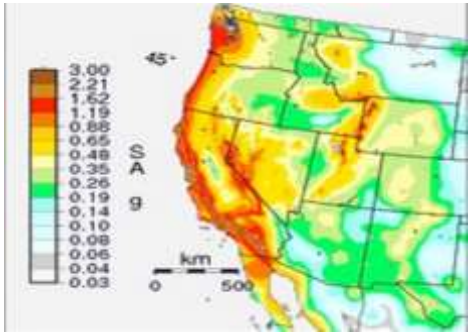


NSF



Telescope Surveys

Earth Sciences



GPN Is Beginning to Define Its Application Drivers

Visualization, Virtual Reality, Collaboration



CITRIS THE BANATAO INSTITUTE

SDSC it²

PRP Provides High Performance Access to Distributed Data Analysis

PACIFIC RESEARCH PLATFORM

Distributed LHC Data Analysis Running Over PRP

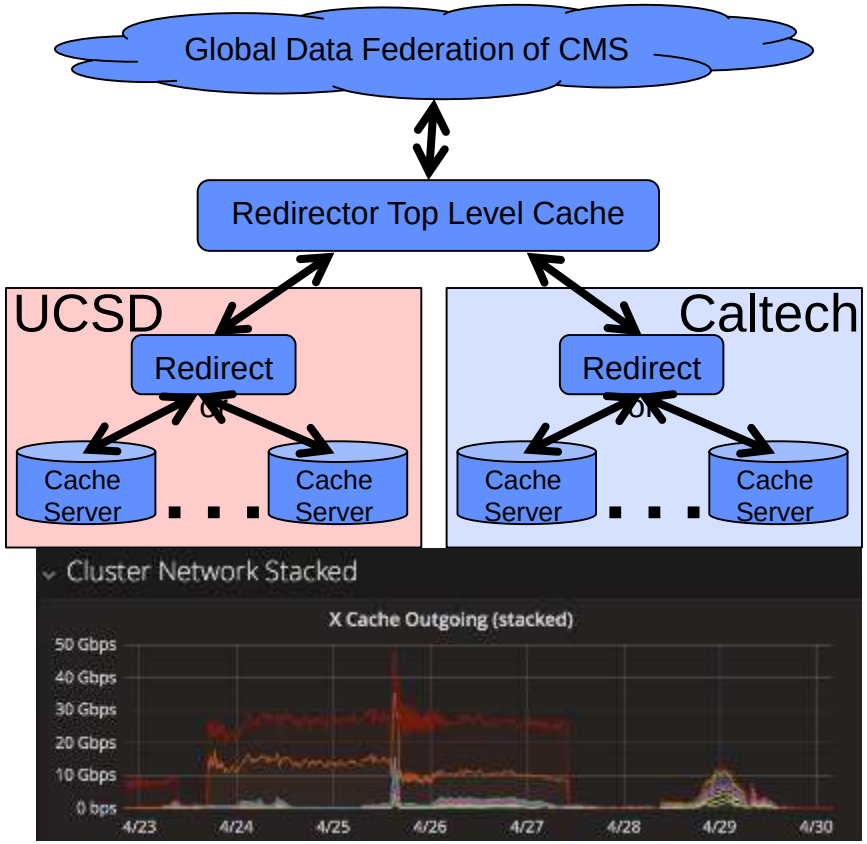
Source: Frank Würthwein, OSG, UCSD/SDSC, PRP



GPN Can Connect Campus LHC Atlas and CMS Data Analysis

PACIFIC RESEARCH PLATFORM

PRP Distributed Tier-2 Cache
Across Caltech & UCSD-Thousands of Flows Sustaining >10Gbps!



Provisioned pilot systems:

PRP UCSD: 9 x 12 SATA Disk of 2TB
@ 10Gbps for Each System

PRP Caltech: 2 x 30 SATA Disk of 6TB
@ 40Gbps for Each System

Source: Frank Würthwein, OSG, UCSD/SDSC, PRP

Collaboration Opportunity with OSG & PRP on Distributed Storage



Open Science Grid

Total data volume pulled last year is dominated by 4 caches.

OSG Is Operating a Distributed Caching CI. At Present, 4 Caches Provide Significant Use

PRP Kubernetes Infrastructure Could Either Grow Existing Caches by Adding Servers, or by Adding Additional Locations

PRP PACIFIC RESEARCH PLATFORM

Source: Frank Würthwein, OSG, UCSD/SDSC, PRP

StashCache Users include:

LIGO





DES

it²

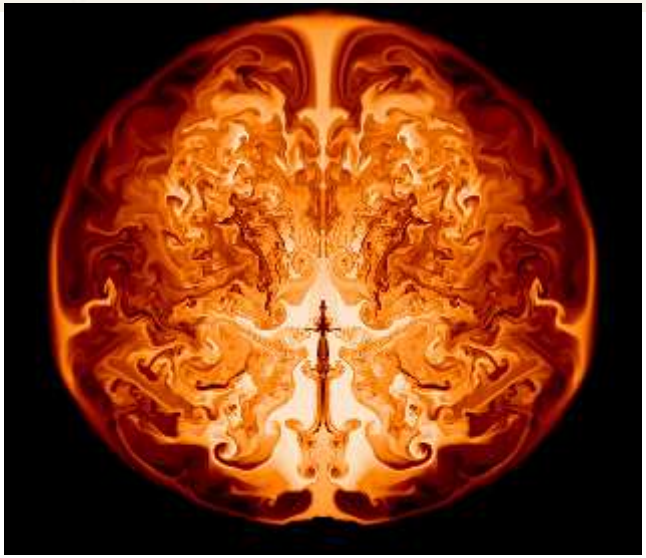
PRP Provides High Performance Access to Remote Supercomputers

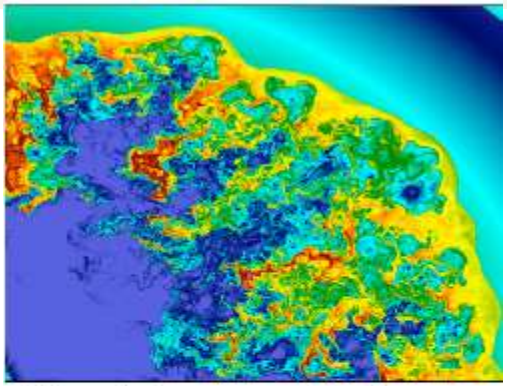
PRP
PACIFIC RESEARCH PLATFORM

it²

100 Gbps PRP Over CENIC
Couples UC Santa Cruz Astrophysics Cluster to LBNL NERSC Supercomputer

CENIC recognizes UCSC's Hyades supercomputer cluster connection





A 2D superluminous supernova simulation generated with the CASTRO code. (Image: Ken Chen, National Astronomical Observatory of Japan)

K. Chen, S. Woolsey, T. Sukhbold, The Astrophysical Journal, 832, 1, Nov. 2016

CENIC 2018 Innovations in Networking Award for Research Applications

PRP PACIFIC RESEARCH PLATFORM

BERKELEY LAB Bringing Science Solutions to the World

NERSC

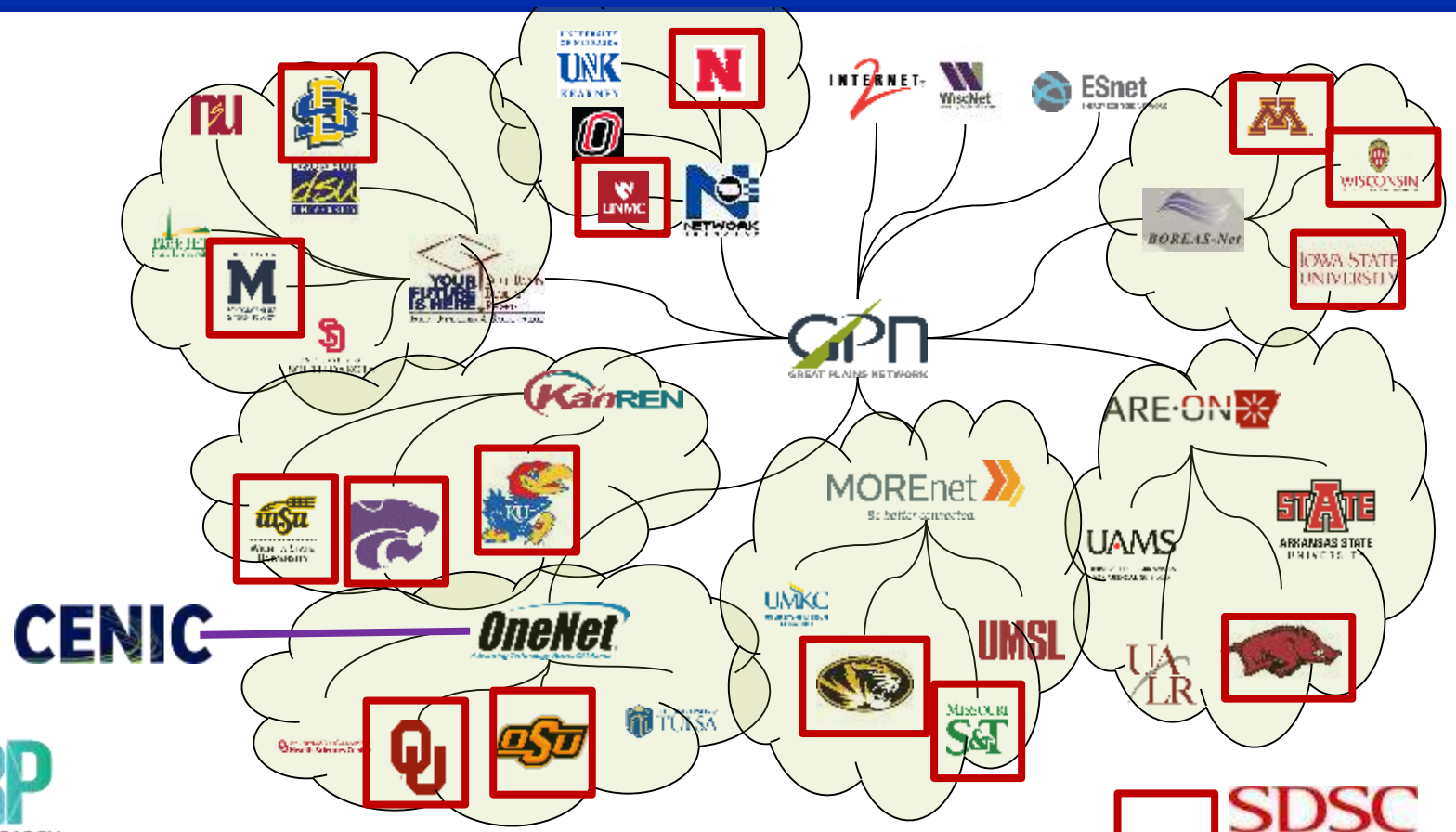
NERSC Project PI:
S. Woolsey, UC Santa Cruz

SDSC

CITRIS THE BANATAO INSTITUTE

it²

The Great Plains Network
Has Many Campuses With Active Projects at SDSC



PRP
PACIFIC RESEARCH PLATFORM

GPN Map Source: James Deaton, GPN

SDSC
Shawn Strande, SDSC

it²

Planning for climate change in California

substantial shifts on top of already high climate variability

SIO Campus Climate Researchers Need to Download Results from NCAR 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

NCAR Upgrading to 100Gbps Link from Wyoming and Boulder to CENIC/PRP

GPN Can Connect Campus NCAR Users

Sponsors:

California Energy Commission

NOAA RISA program

California DWR, DOE, NSF

PACIFIC RESEARCH PLATFORM

PRP Provides High Performance Access to Multi-Campus Big Data Collaborative Teams

PACIFIC RESEARCH PLATFORM

PRP Will Link the Laboratories of the Pacific Earthquake Engineering Research Center

PEER

PACIFIC EARTHQUAKE ENGINEERING RESEARCH CENTER

http://peer.berkeley.edu/

HOME | ABOUT PEER | NEWS | EVENTS | RESEARCH | PRODUCTS | LABORATORIES | PUBLICATIONS | NISEE | BIP MEMBERS | EDUCATION | FAQs | LINKS

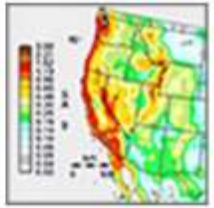
PEER Mega Research Programs

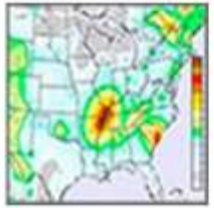

Tall Buildings Initiative


Transportation Systems


Lifelines & NGL


Tsunami


NGA-West 2


NGA-East


Global GMPE



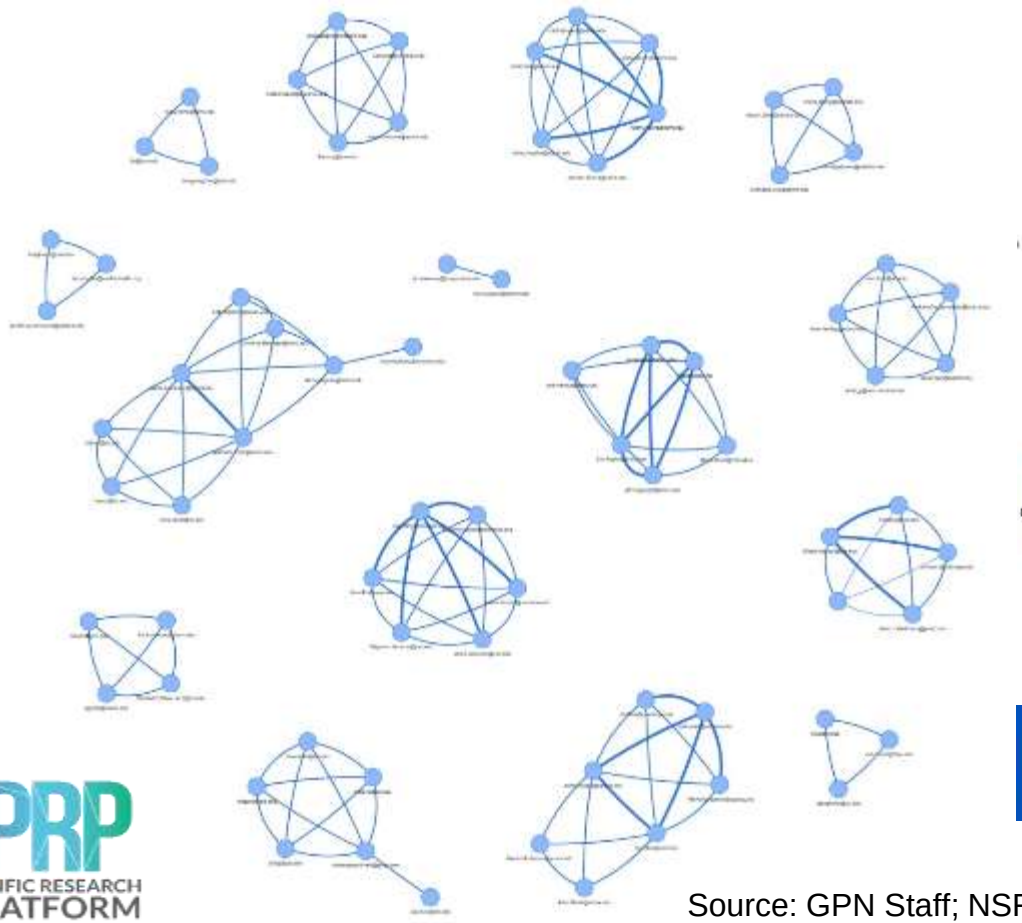
John Graham Installing FIONette at PEER
Feb 10, 2017

PEER Labs: UC Berkeley, Caltech, Stanford, UC Davis, UC San Diego, and UC Los Angeles

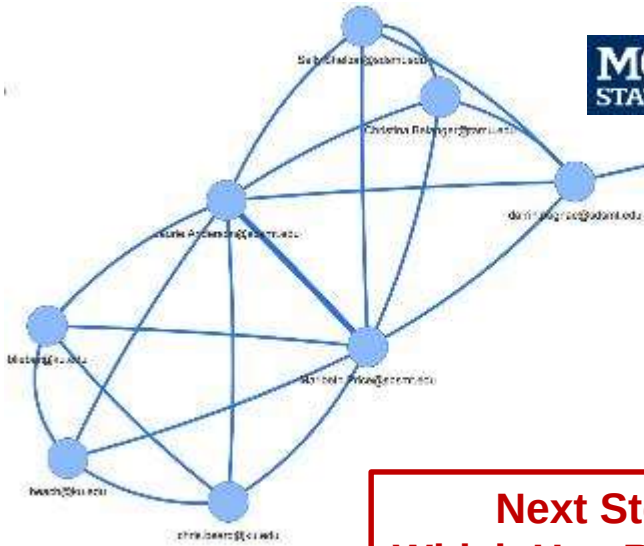
PRP
PACIFIC RESEARCH PLATFORM

it²

Identifying NSF Multi-Institutional Grants Across and Beyond the PRP



SD MINES



MONTANA STATE UNIVERSITY

KU

Next Step:
Which Use Big Data?

PRP
PACIFIC RESEARCH PLATFORM

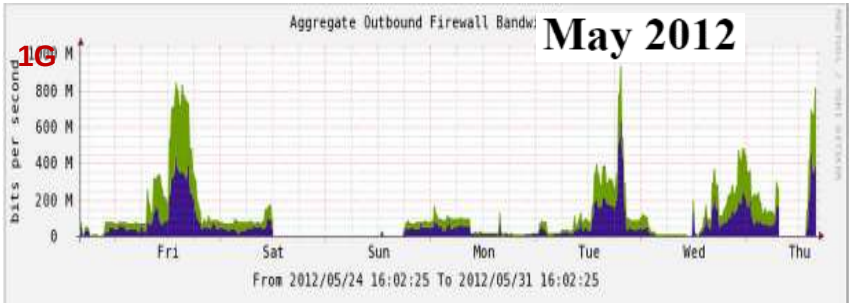
Source: GPN Staff; NSF Public Data

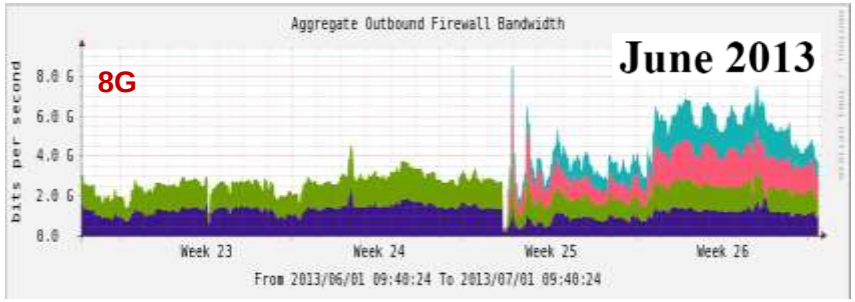
it²

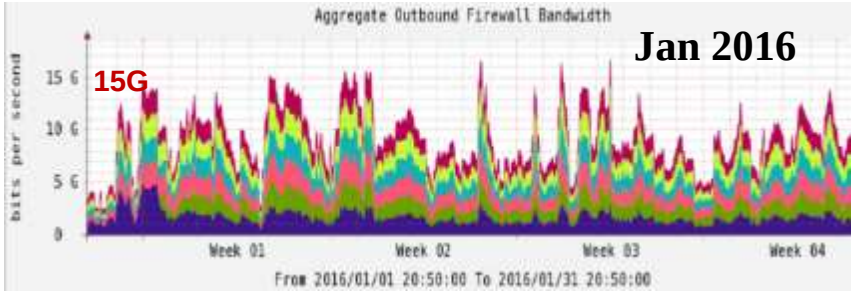
PRP Provides High Performance Access to Large Community Data Repositories

PACIFIC RESEARCH PLATFORM

Cancer Genomics Hub (UCSC) Was Housed in SDSC, But NIH Moved Dataset From SDSC to UChicago - So the PRP Deployed a FIONA to Chicago's MREN









NIH NATIONAL CANCER INSTITUTE Genomic Data Commons heralds new era of data sharing for cancer research

June 6, 2016

UChicagoNews

Data Source: David Haussler, Brad Smith, UCSC

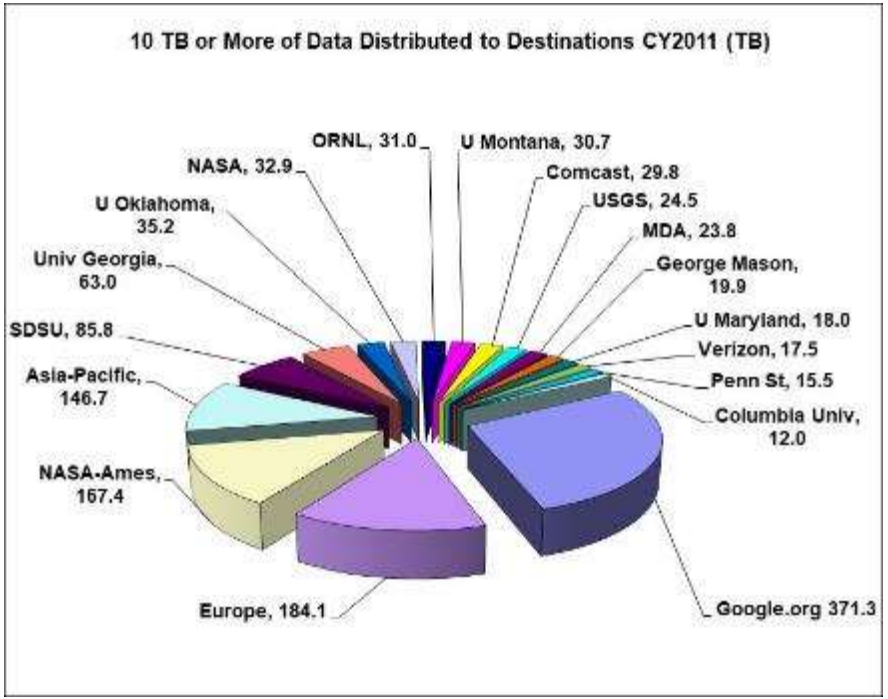
PRP PACIFIC RESEARCH PLATFORM

it²

USGS Earth Resources Observation and Science (EROS) Center Is a Natural GPN/PRP Big Data Repository



EROS is located ~15 miles north of Sioux Falls, South Dakota



In 2011 EROS sent the equivalent of the entire library of congress every 9 days
In 2011 SDSU was the 3rd largest user downloading data (GIS)
In 2016 EROS sent the equivalent of the entire library of congress every 6 hours

Source: Claude Garelik

PRP Provides High Performance Access to Large Scientific Instruments

PRP
PACIFIC RESEARCH PLATFORM

it²

100 Gbps FIONA at UCSC Allows for Downloads to the UCSC Hyades Cluster from the LBNL NERSC Supercomputer for DESI Science Analysis



INTERMEDIATE PALOMAR TRANSIENT FACTORY

300 images per night.
100MB per raw image

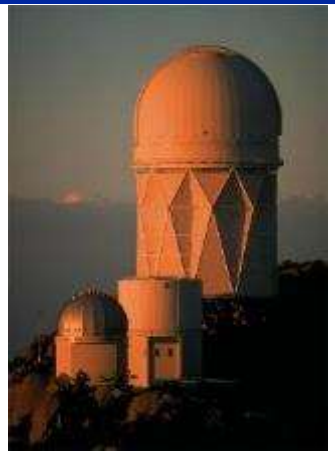
120GB per night

Source: Peter Nugent, LBNL
Professor of Astronomy, UC Berkeley

Precursors to LSST and NCSA



NSF-Funded Cyberengineer
Shaw Dong @UCSC
Receiving FIONA
Feb 7, 2017



Dark Energy Spectroscopic Instrument

250 images per night.
530MB per raw image

800GB per night

BERKELEY LAB
Bringing Science Solutions to the World

NERSC

PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

ESnet

SDSC

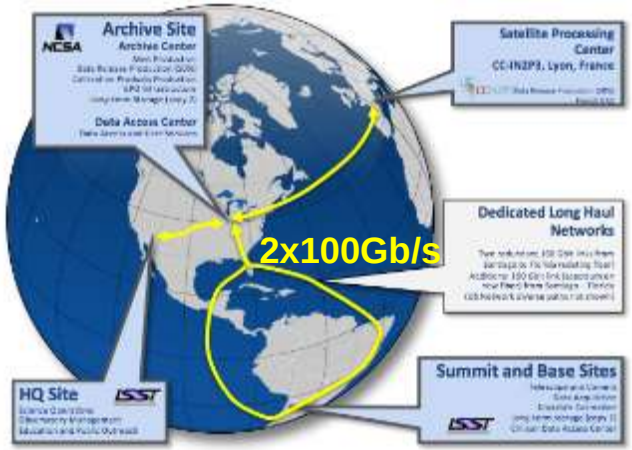
CITRIS BANATAO INSTITUTE

it²

Global Scientific Instruments Will Produce Ultralarge Datasets Continuously Requiring Dedicated Optic Fiber and Supercomputers

Large Synoptic Survey Telescope

3.2 Gpixel Camera
Tracks ~40B Objects,
Creates 1-10M Alerts/Night
Within 1 Minute of Observing
1000 Supernovas Discovered/Night



“First Light”
In 2019



PACIFIC RESEARCH PLATFORM

it²

The Prototype PRP Has Attracted New Application Drivers



Frank Vernon, Graham Kent, & Ilkay Altintas, Wildfires

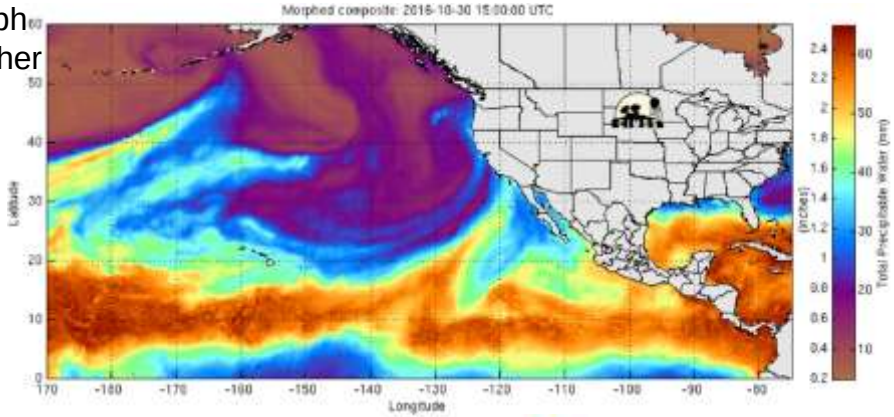


Jules Jaffe – Undersea Microscope



Tom Levy At-Risk Cultural Heritage

Scott Sellars, Marty Ralph
Center for Western Weather and Water Extremes



jupyterhub

PRP PACIFIC RESEARCH PLATFORM

NSF

CENIC

ESnet

SDSC

CITRIS BANATAO INSTITUTE

it²

PRP UC-JupyterHub Backbone Connects FIONAs At UC Berkeley and UC San Diego

Source: John Graham, Calit2

jupyterhub

A multi-user version of the notebook designed for companies, classrooms and research labs

Goal: Jupyter Everywhere



PRP PACIFIC RESEARCH PLATFORM

PRP Provides High Performance Access to SensorNets Coupled to Realtime Computing

PRP
PACIFIC RESEARCH PLATFORM

it²

New PRP Application:
Coupling Wireless Wildfire Sensors to Computing

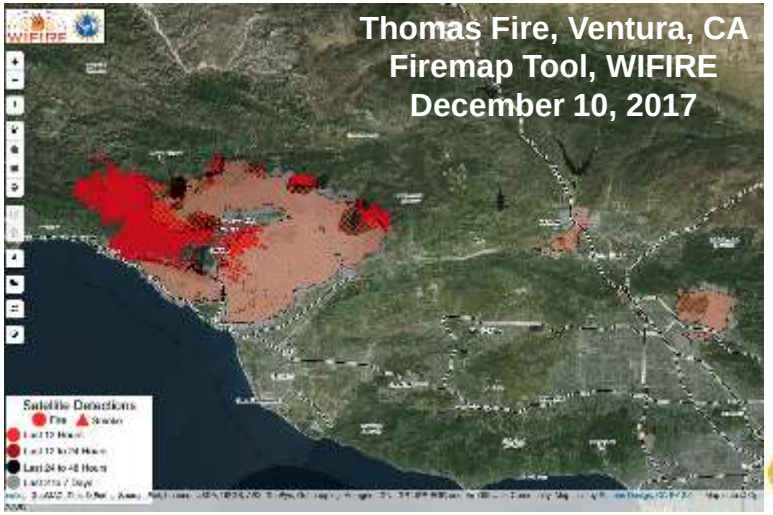
Technology Projects to Combat California Wildfires Are Recognized with CENIC Innovation Award

WIFIRE · HPWREN · Nevada Seismological Laboratory, Mackay School of Earth Sciences & Engineering

CENIC 2018
Innovations in Networking Award for Experimental Applications



Church Fire, San Diego CA
Alert SD&ECameras/HPWREN
October 21, 2017



Thomas Fire, Ventura, CA
Firemap Tool, WIFIRE
December 10, 2017

PRP PACIFIC RESEARCH PLATFORM

it²

Once a Wildfire is Spotted, PRP Brings High-Resolution Weather Data to Fire Modeling Workflows in WIFIRE



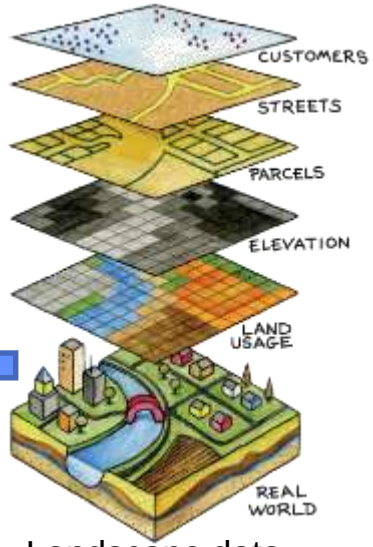
Real-Time Meteorological Sensors



Source: Ilkay Altintas, SDSC

WIFIRE

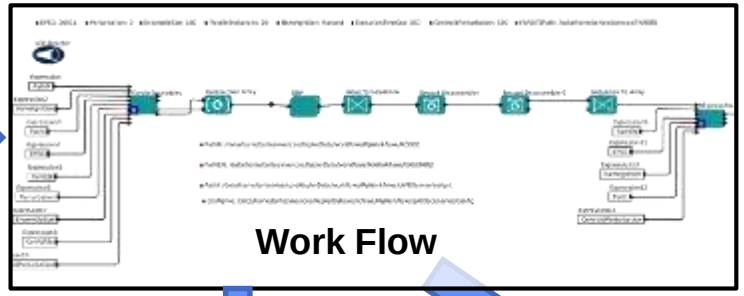




Landscape data

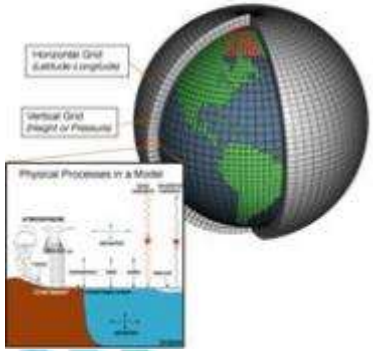
PRP





Work Flow

Weather Forecast



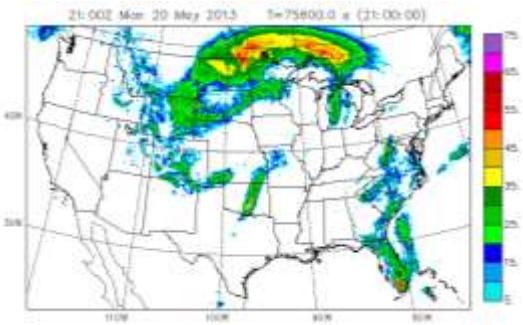






PRP
PACIFIC RESEARCH PLATFORM

High Resolution Ensemble Weather Forecasts at The Center for Analysis and Prediction of Storms, University of Oklahoma



Full CONUS Data Volumes	2014	2015-2017
Grid Spacing	4 km	3 km
Domain Size	1163x723x53	1683x1155x53
One Output Time	4.2 GB	9.7 GB
Sub-Hourly Interval	10 min	6 min
Complete Forecast Size Hourly + Sub-hourly 18h-30h	508 GB	1639 GB
For 10 members per day	5.08 TB	16.4 TB
For approx 30 days per season	152 TB	492 TB

Prime Target for GPN/OneNet

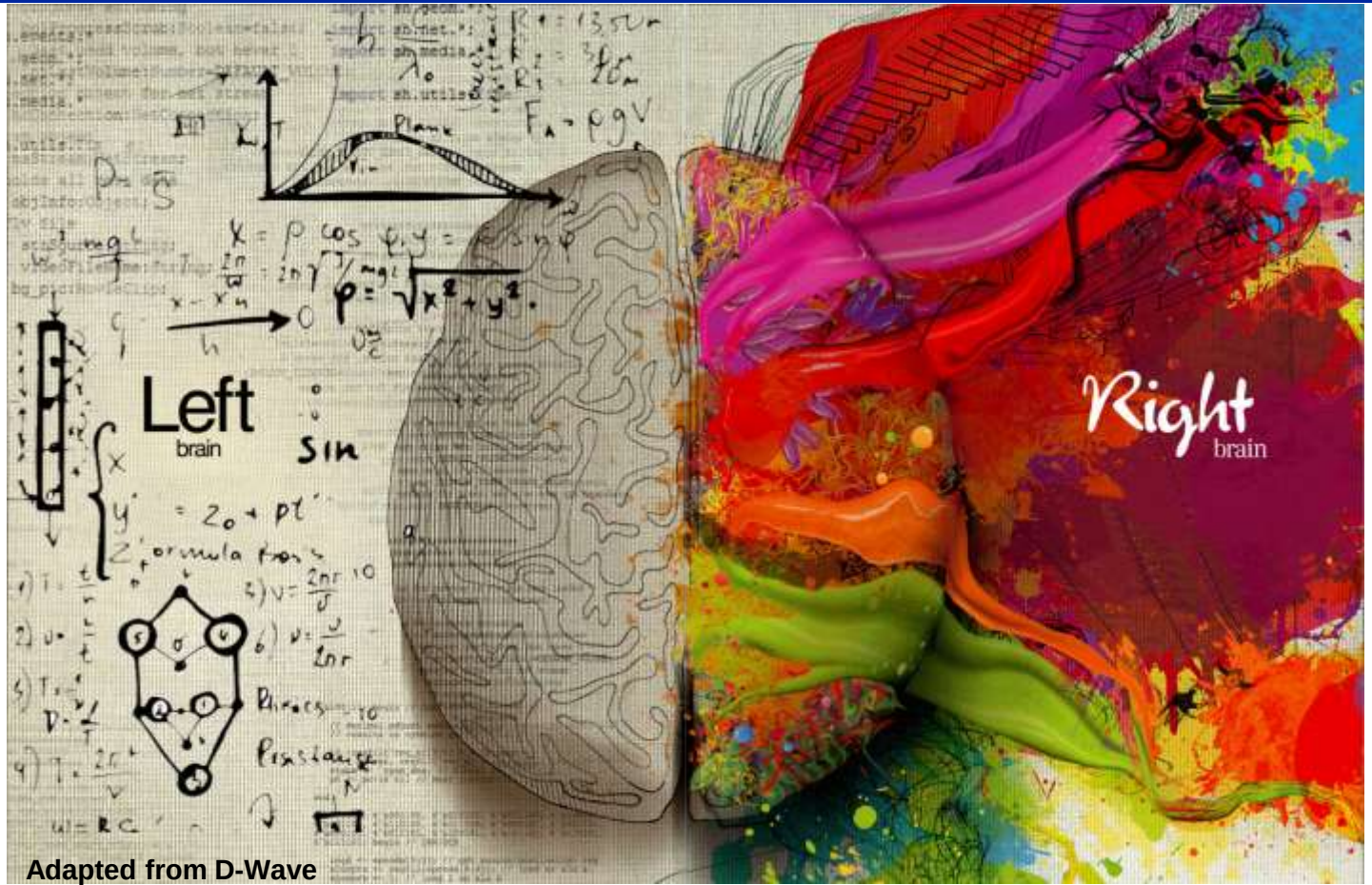
Hazardous Weather Testbed

- **In 2017, CAPS started testing the next-generation forecasting model FV3 for convective-scale forecasting.**
- **For 2018 HWT CLUE, CAPS is producing 5 ensembles using WRF and FV3 with a total of 52 forecasts of up to 84 hours.**

Dec, 2013: CAPS has >1 PB of in-house storage capacity

Source: Ming Xue and Keith Brewster, CAPS

The Rise of Brain-Inspired Computers:
Left & Right Brain Computing: Arithmetic vs. Pattern Recognition



Adapted from D-Wave

PRP
PACIFIC RESEARCH PLATFORM

it²

UC San Diego Jaffe Lab (SIO) Scripps Plankton Camera Off the SIO Pier with Fiber Optic Network

SCRIPPS INSTITUTION OF OCEANOGRAPHY

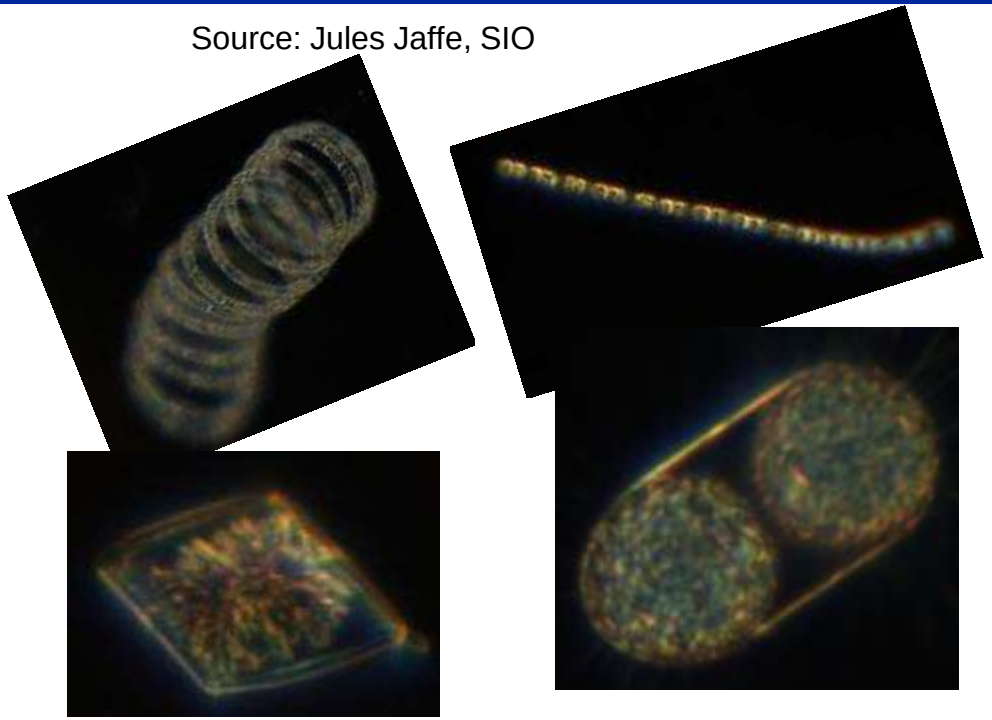


PRP PACIFIC RESEARCH PLATFORM

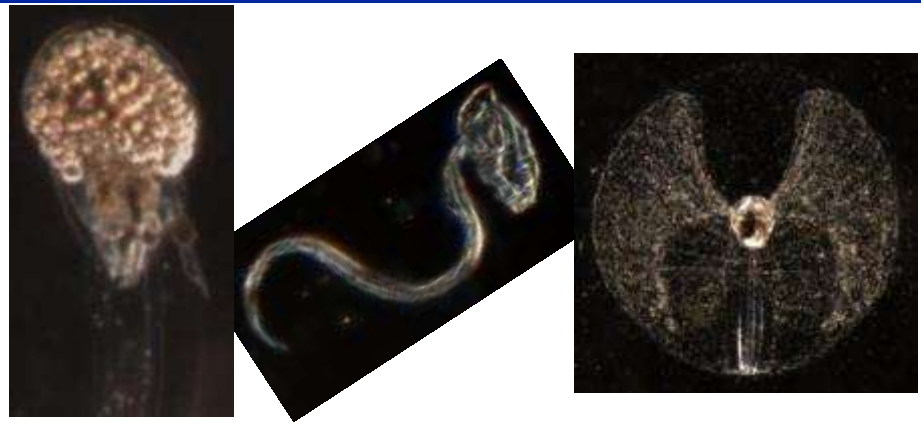
it²

Over 1 Billion Images So Far!
Requires Machine Learning for Automated Image Analysis and Classification

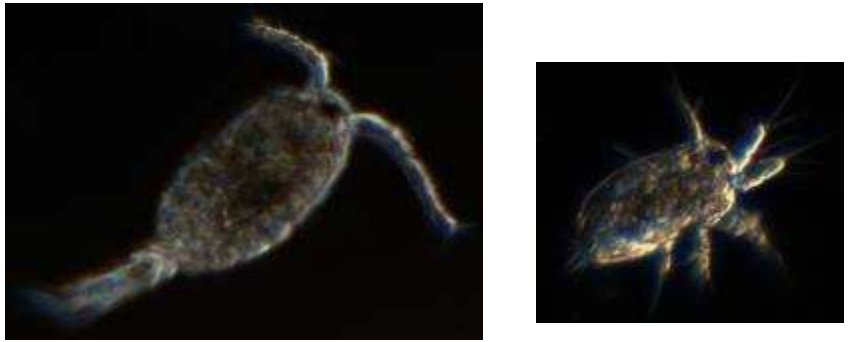
Source: Jules Jaffe, SIO



Phytoplankton: Diatoms



Zooplankton: Larvaceans



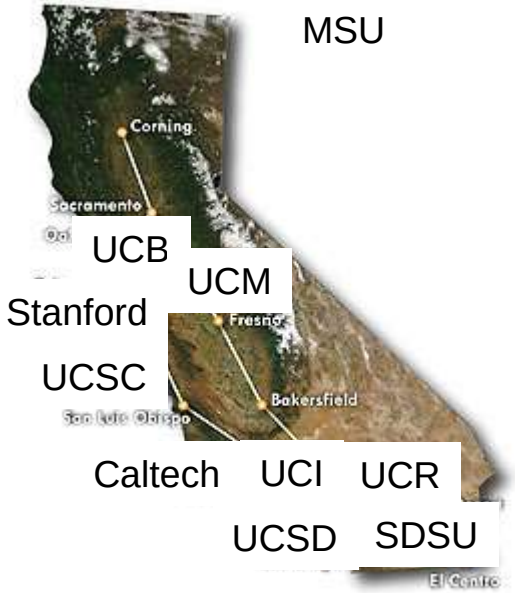
Zooplankton: Copepods

PACIFIC RESEARCH PLATFORM

”We are using the FIONAs for image processing... this includes doing Particle Tracking Velocimetry that is very computationally intense.”-Jules Jaffe

it²

New NSF CHASE-CI Grant Creates a Community Cyberinfrastructure: Adding a Machine Learning Layer Built on Top of the Pacific Research Platform



CI-New: Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI)

For the Period September 1, 2017 – August 31, 2020

SUBMITTED – January 18, 2017

PI: Larry Smarr, Professor of Computer Science and Engineering, Director Calit2, UCSD
Co-PI: Tajana Rosing, Professor of Computer Science and Engineering, UCSD
Co-PI: Ken Kreutz-Delgado, Professor of Electrical and Computer Engineering, UCSD
Co-PI: Ilkay Altintas, Chief Data Science Officer, San Diego Supercomputer Center, UCSD
Co-PI: Tom DeFanti, Research Scientist, Calit2, UCSD

NSF Program Officer: Mimi McClure

NSF Grant for High Speed "Cloud" of 256 GPUs
For 30 ML Faculty & Their Students at 10 Campuses
for Training AI Algorithms on Big Data

PACIFIC RESEARCH PLATFORM

JACOBS SCHOOL OF ENGINEERING

SDSC UC San Diego

FIONA8: Adding GPUs to FIONAs
Supports Data Science Machine Learning



Multi-Tenant Containerized GPU JupyterHub
Running Kubernetes / CoreOS
Eight Nvidia GTX-1080 Ti GPUs

~$13K

32GB RAM, 3TB SSD, 40G & Dual 10G ports

Source: John Graham, Calit2

PACIFIC RESEARCH PLATFORM

UCSD Adding >350 Game GPUs to Data Sciences Cyberinfrastructure - Devoted to Data Analytics and Machine Learning

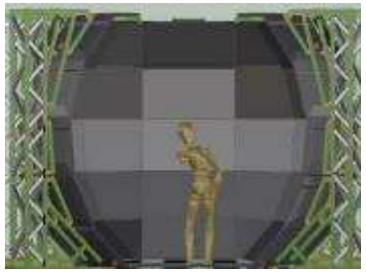
UC San Diego

IT SERVICES

95 GPUs
for Students

Open Science Grid

48 GPUs for
OSG Applications



SunCAVE 70 GPUs
WAVE + Vroom 48 GPUs

UCSD Cognitive Science

JACOBS SCHOOL OF ENGINEERING

SDSC it²



Plus 288 64-bit GPUs
On SDSC's Comet

QUALCOMM INSTITUTE



FIONA with
8-Game GPUs

CHASE-CI Grant Provides
96 GPUs at UCSD
for Training AI Algorithms on Big Data

PRP
PACIFIC RESEARCH PLATFORM

it²

Next Step: Surrounding the PRP Machine Learning Platform With Clouds of GPUs and Non-Von Neumann Processors

IBM



64-TrueNorth Cluster

IBM Cloud

IBM Watson

64-bit GPUs

XSEDE

Extreme Science and Engineering Discovery Environment

NSF





Microsoft Installs Altera FPGAs into Bing Servers & 384 into TACC for Academic Access

Project Catapult

Established: February 2, 2015

Microsoft

PATTERN COMPUTER.



CHASE-CI

PRP

PACIFIC RESEARCH PLATFORM

GPN Next Step: Add GPUs to FIONAs

Google's First Machine Learning Chip (TPU) Is 30x Faster Than CPUs And GPUs

BUILDING AN AI CHIP SAVED GOOGLE FROM BUILDING A DOZEN NEW DATA CENTERS

WIRED



SEP 30, 2016 @ 07:39 PM 3,606

Forbes / Tech

Amazon Boosts Cloud-Computing Performance With New, GPU-Accelerated AWS Instances

amazon web services

Japan plans to build the fastest deep learning supercomputer

AIST ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

4352x NVIDIA Tesla V100 GPUs

it²

The Second National Research Platform Workshop
Bozeman, MT August 6-7, 2018

Local Host: Jerry Sheehan, MSU

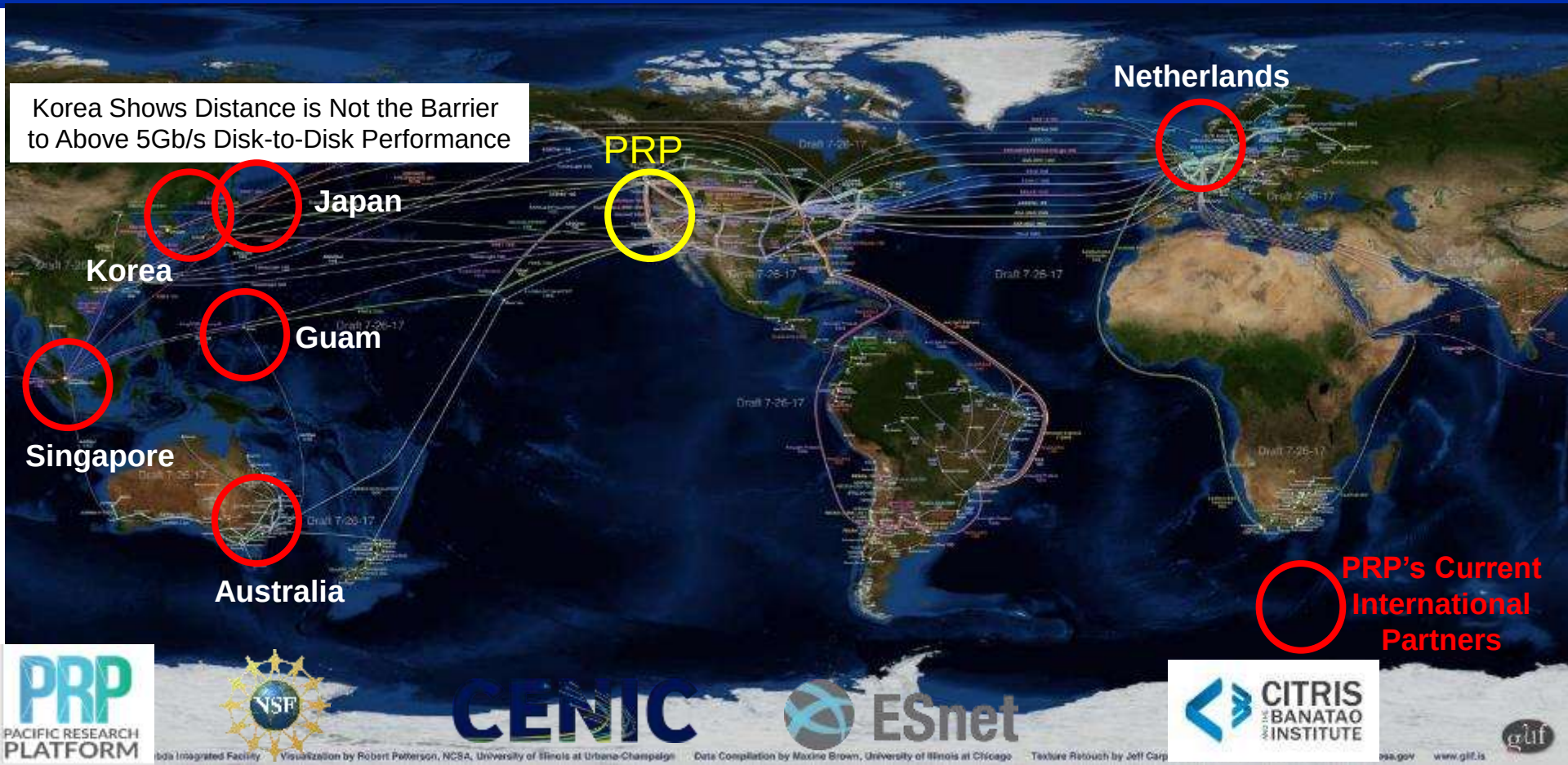




Announced in I2 Closing Keynote:
Larry Smarr "Toward a National Big Data Superhighway"
on Wednesday, April 26, 2017

Co-Chairs:
Larry Smarr, Calit2
Inder Monga, ESnet
Ana Hunsinger, Internet2

Expanding to the Global Research Platform
Via CENIC/Pacific Wave, Internet2, and International Links



Our Support:

- US National Science Foundation (NSF) awards
  - CNS 0821155, CNS-1338192, CNS-1456638, CNS-1730158, ACI-1540112, & ACI-1541349



- University of California Office of the President CIO
- UCSD Chancellor's Integrated Digital Infrastructure Program
- UCSD Next Generation Networking initiative
- Calit2 and Calit2 Qualcomm Institute
- CENIC, PacificWave and StarLight
- DOE ESnet

PRP
PACIFIC RESEARCH PLATFORM

it²