“Advanced Global-Scale Networking Supporting Data-Intensive Artificial Intelligence Applications ”

Joint Networks Summit

San Diego State University

January 30, 2020

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

PACIFIC RESEARCH PLATFORM

Creating a Tightly-Coupled, Yet Highly Distributed Cyberinfrastructure for Big Data Analysis

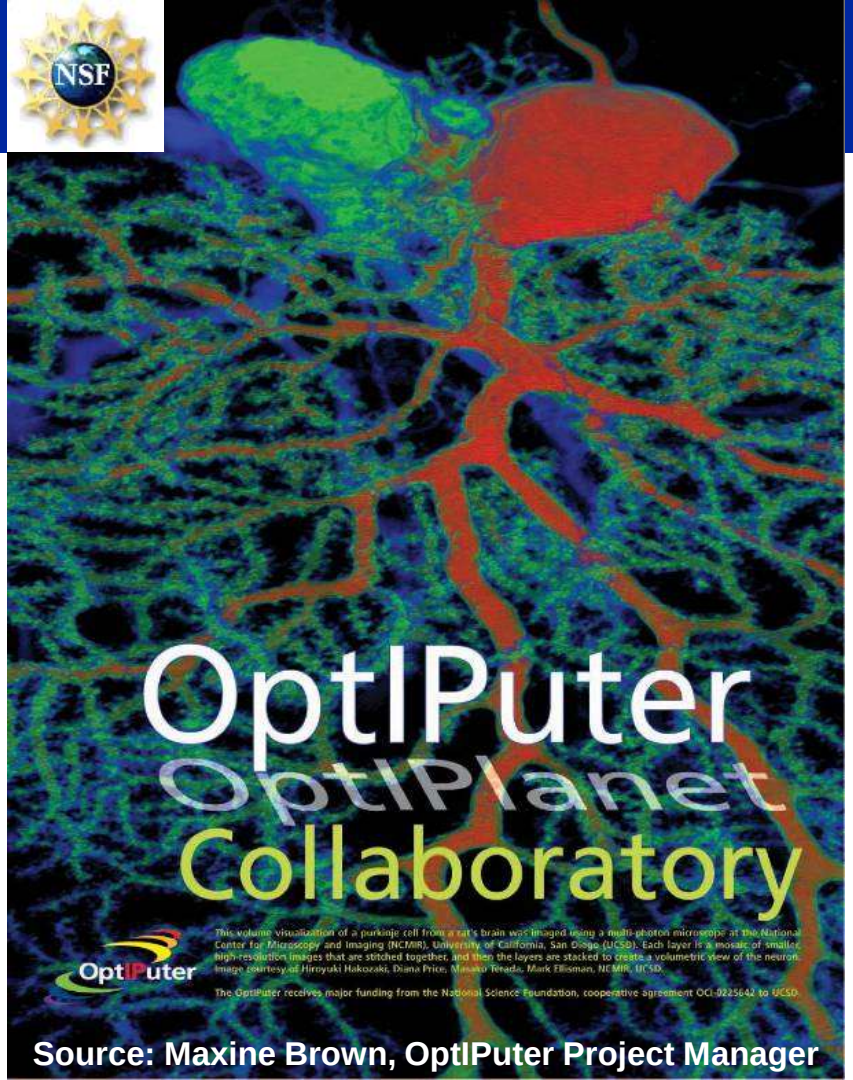
Vision:
Use Optical Fiber to Connect
Big Data Generators and Consumers,
Creating a "Big Data" Freeway System

"The Bisection Bandwidth of a Cluster Interconnect, but Deployed on a 20-Campus Scale."

PRP PACIFIC RESEARCH PLATFORM

This Vision Has Been Building for 15 Years

it²



Source: Maxine Brown, OptIPuter Project Manager

The OptIPuter
Exploits a New World
in Which
the Central Architectural Element
is Optical Networking,
Not Computers.

Distributed Cyberinfrastructure
to Support
Data-Intensive Scientific Research
and Collaboration

PI Smarr,
2002-2009

Integrated "OptIPlatform" Cyberinfrastructure System: A 10Gbps Lightpath Cloud



End User OptIPortal



Instruments



HD/4k Telepresence



HD/4k Video Cams



Campus Optical Switch

10G Lightpath



HPC

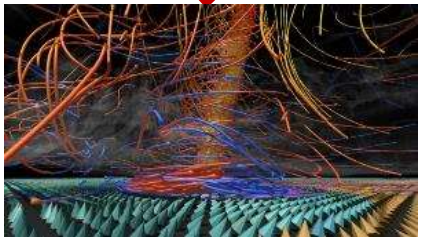
TeraGrid

OptIPuter

LS 2009 Slide



Data Repositories & Clusters



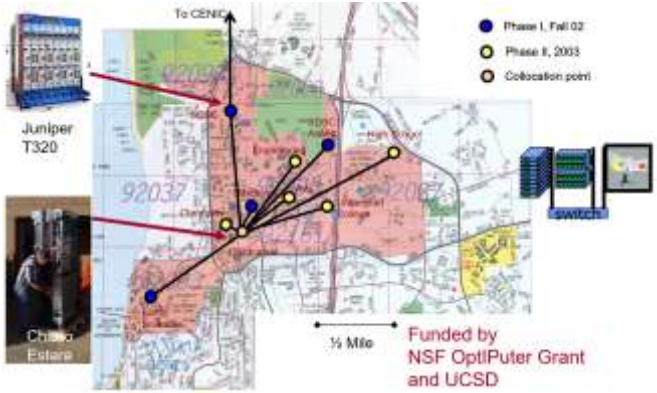
HD/4k Video Images

PRP PACIFIC RESEARCH PLATFORM

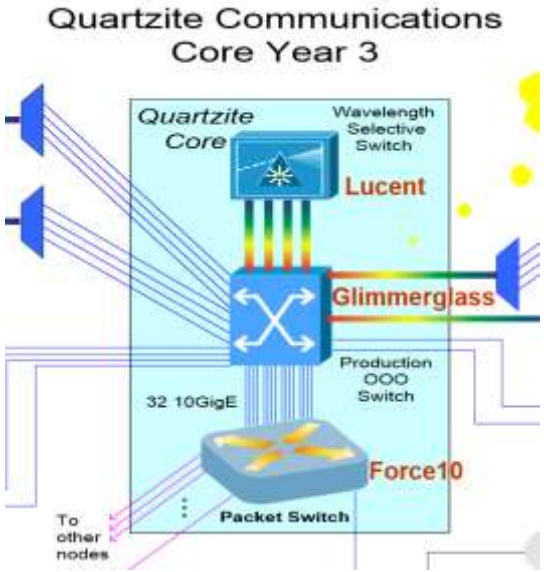
SDSC

it²

2000-2015 Using UCSD Campus as Development Prototype: NSF OptIPuter, Quartzite, Prism Awards

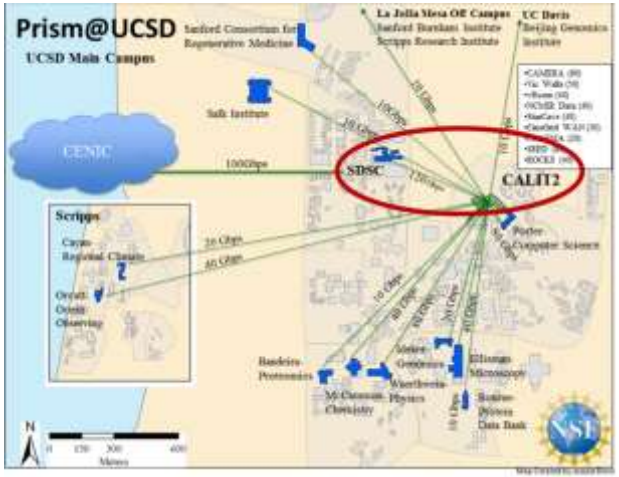


PI Smarr, 2002-2009



Goals by 2008:
>= 50 endpoints at 10 GigE
>= 32 Packet switched
>= 32 Switched wavelengths
>= 300 Connected endpoints

PI Papadopoulos, 2004-2007

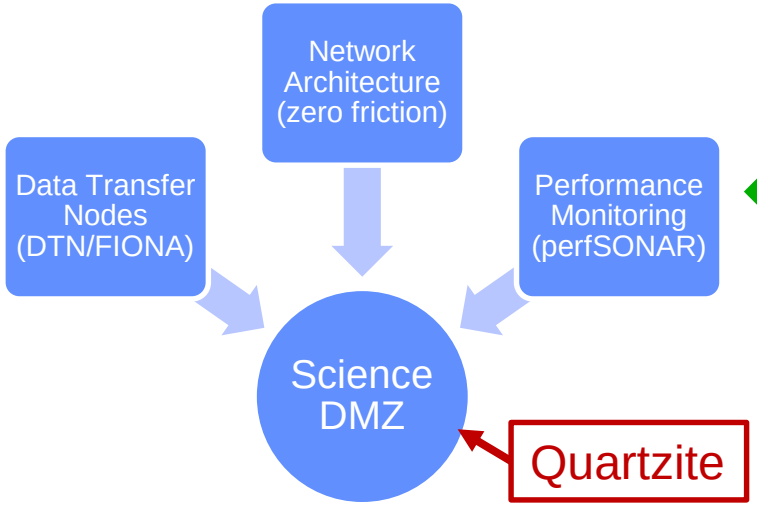


PI Papadopoulos, 2013-2015

PACIFIC RESEARCH PLATFORM

Before the PRP: ESnet's ScienceDMZ Accelerates Science Research: DOE & NSF Partnering on Science Engagement and Technology Adoption

ScienceDMZ Coined in 2010 by ESnet
Basis of PRP Architecture and Design



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

DOE

NSF

NSF Campus Cyberinfrastructure Program
Has Made Over 250 Awards



Prism

2012 2013 2014 2015 2016 2017 2018

Slide Adapted From Inder Monga, ESnet

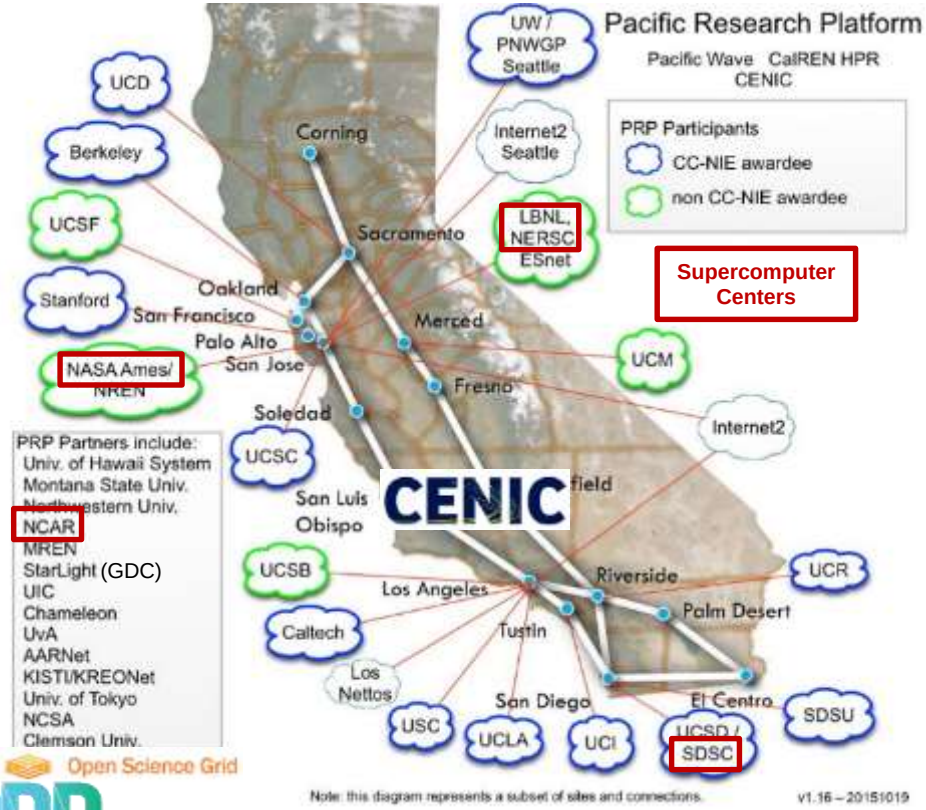
PRP PACIFIC RESEARCH PLATFORM

ESnet FASTERDATA KNOWLEDGEBASE

NSF

it²

2015 Vision: The Pacific Research Platform Will Connect Science DMZs Creating a Regional End-to-End Science-Driven Community Cyberinfrastructure



Source: John Hess, CENIC

NSF CC*DNI Grant
$6.3M 10/2015-10/2020
In Year 5 Now

PI: Larry Smarr, UC San Diego Calit2
Co-PIs:

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

Letters of Commitment from:

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

ESnet: Given Fast Networks, Need DMZs and Fast/Tuned DTNs

Terminating the Fiber Optics - Data Transfer Nodes (DTNs): Flash I/O Network Appliances (FIONAs)

UCSD-Designed FIONAs Solved the Disk-to-Disk Data Transfer Problem *at Near Full Speed* on Best-Effort 10G, 40G and 100G Networks



Two FIONA DTNs at UC Santa Cruz: 40G & 100G
Up to 192 TB Rotating Storage





Add Up to 8 Nvidia GPUs Per 2U FIONA
To Add Machine Learning Capability

PRP PACIFIC RESEARCH PLATFORM

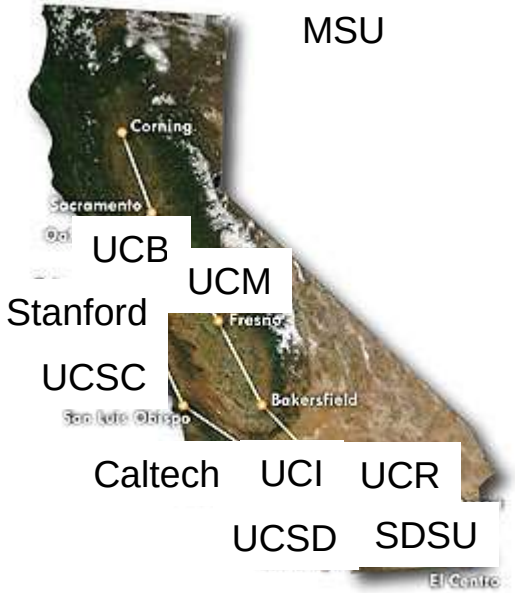
NSF

FIONAs Designed by UCSD's Phil Papadopoulos, John Graham, Joe Keefe, and Tom DeFanti

SDSC

it²

2017-2020: CHASE-CI Grant Adds 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 Grant for 256 High Speed “Cloud” GPUs
For 32 ML Faculty & Their Students at 10 Campuses
To Train AI Algorithms on Big Data

PACIFIC RESEARCH PLATFORM

SDSC UC San Diego it²

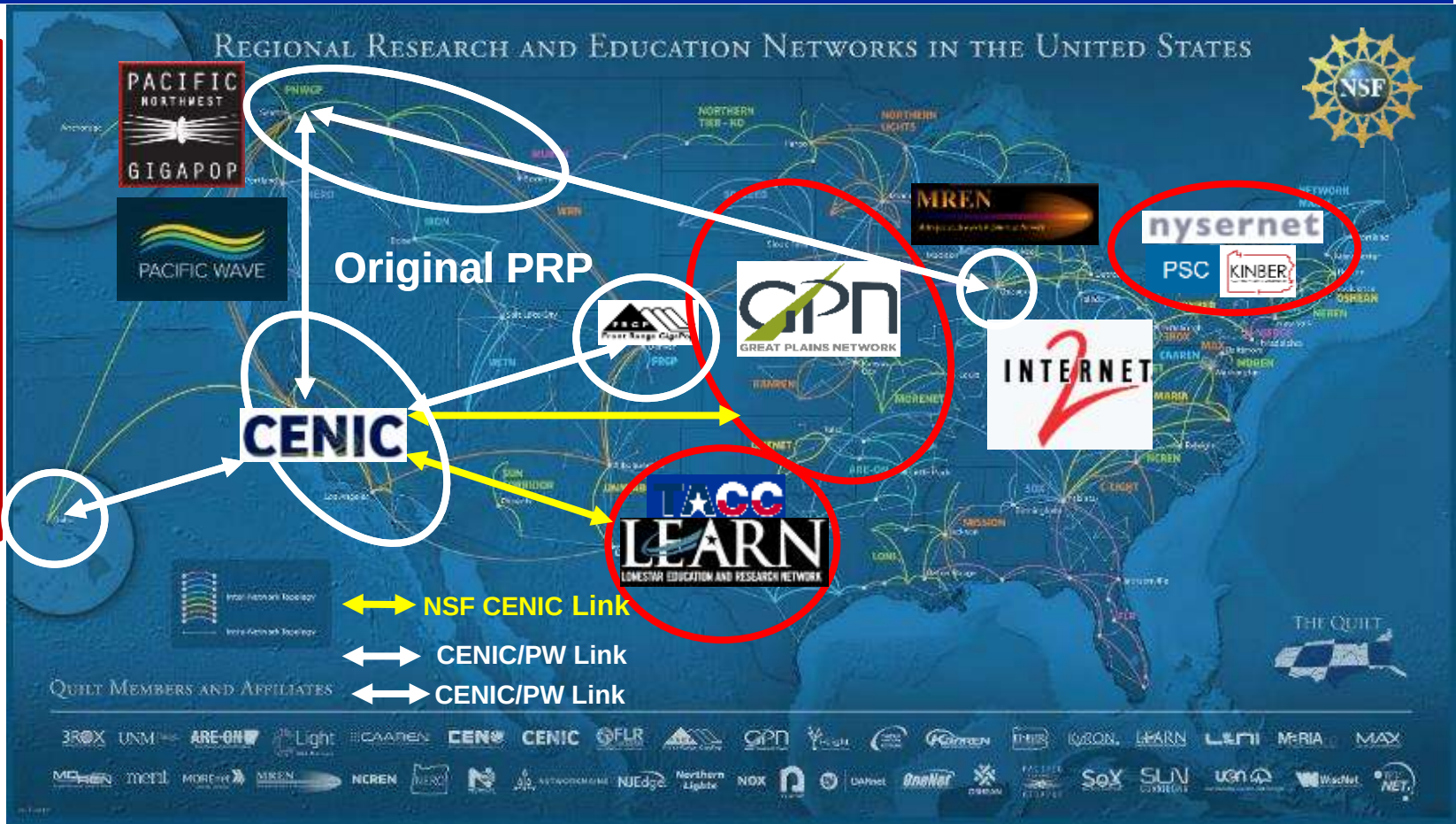
2018-2021: Toward the National Research Platform (NRP) - Using CENIC & Internet2 to Connect Quilt Regional R&E Networks

"Towards The NRP"
3-Year Grant Funded by NSF
$2.5M
October 2018

PI Smarr
Co-PIs Altintas Papadopoulos Wuerthwein Rosing DeFanti

SDSC

PRP
PACIFIC RESEARCH PLATFORM



2018/2019: PRP Game Changer!
Using Kubernetes to Orchestrate 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

News

How Kubernetes Conquered 2017 (and is Positioned for 2018)



Dan Meyer | Editor-in-Chief
December 25, 2017 4:00 AM

sdxcentral



Kubernetes, an open-source software project that started at Google, has exploded in popularity and is now used by at least 54% of the Fortune 500. BUSINESS INSIDER **Rosalie Chan** Jan. 27, 2019,

PRP PACIFIC RESEARCH PLATFORM

CLOUD NATIVE COMPUTING FOUNDATION

it²

PRP's *Nautilus Hypercluster* Adopted Kubernetes to Orchestrate Software Containers and Rook, Which Runs Inside of Kubernetes, to Manage Distributed Storage



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.

"Kubernetes with Rook/Ceph Allows Us to Manage Petabytes of Distributed Storage and GPUs for Data Science,
While We Measure and Monitor Network Use."
--John Graham, Calit2/QI UC San Diego

PRP PACIFIC RESEARCH PLATFORM

it²

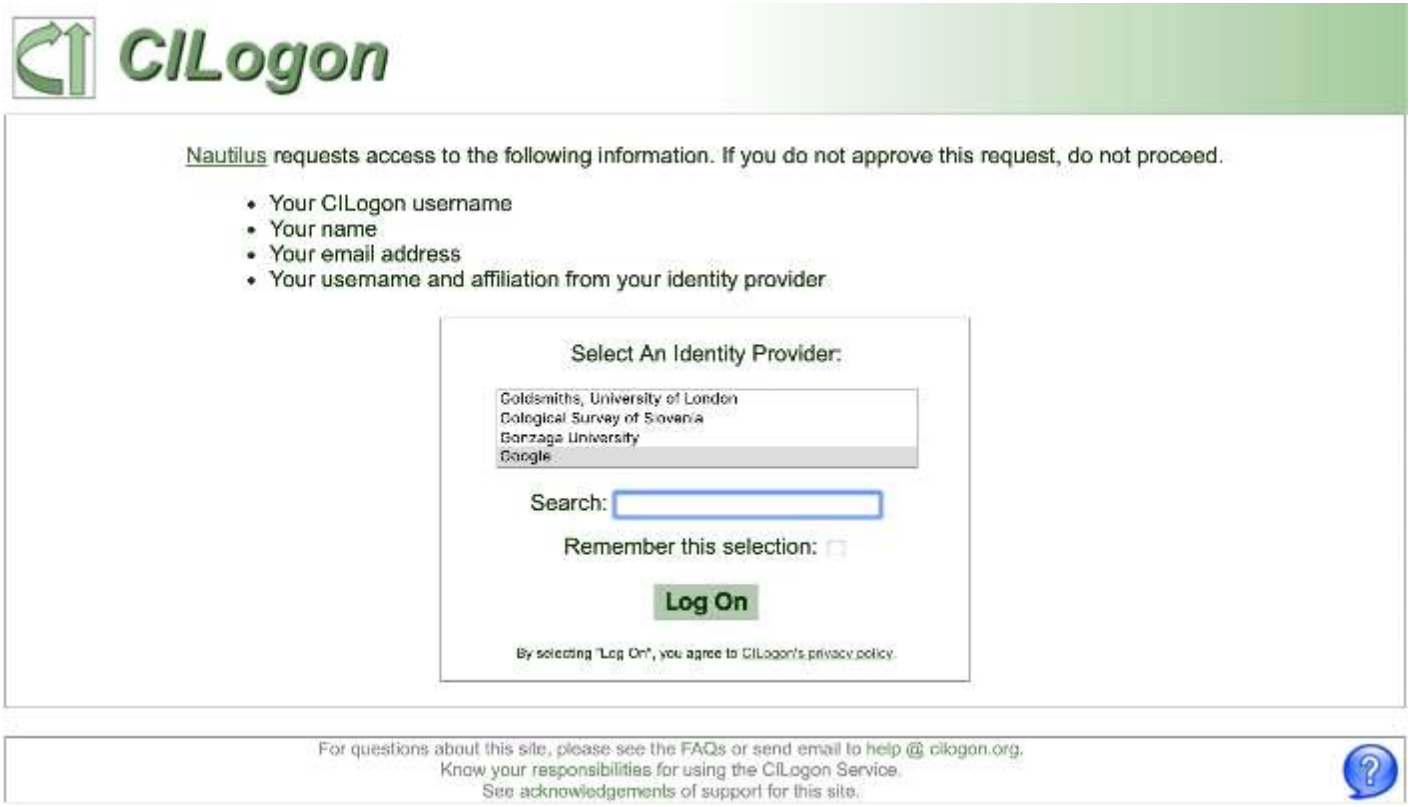
Security Technologies Utilized in the PRP / CHASE-CI Nautilus Hypercluster

- CILogon Federated Authentication
- Secure Namespaces for Multi-Institution Collaborations
- Calico Policy-Driven Network Security

PRP PACIFIC RESEARCH PLATFORM

it²

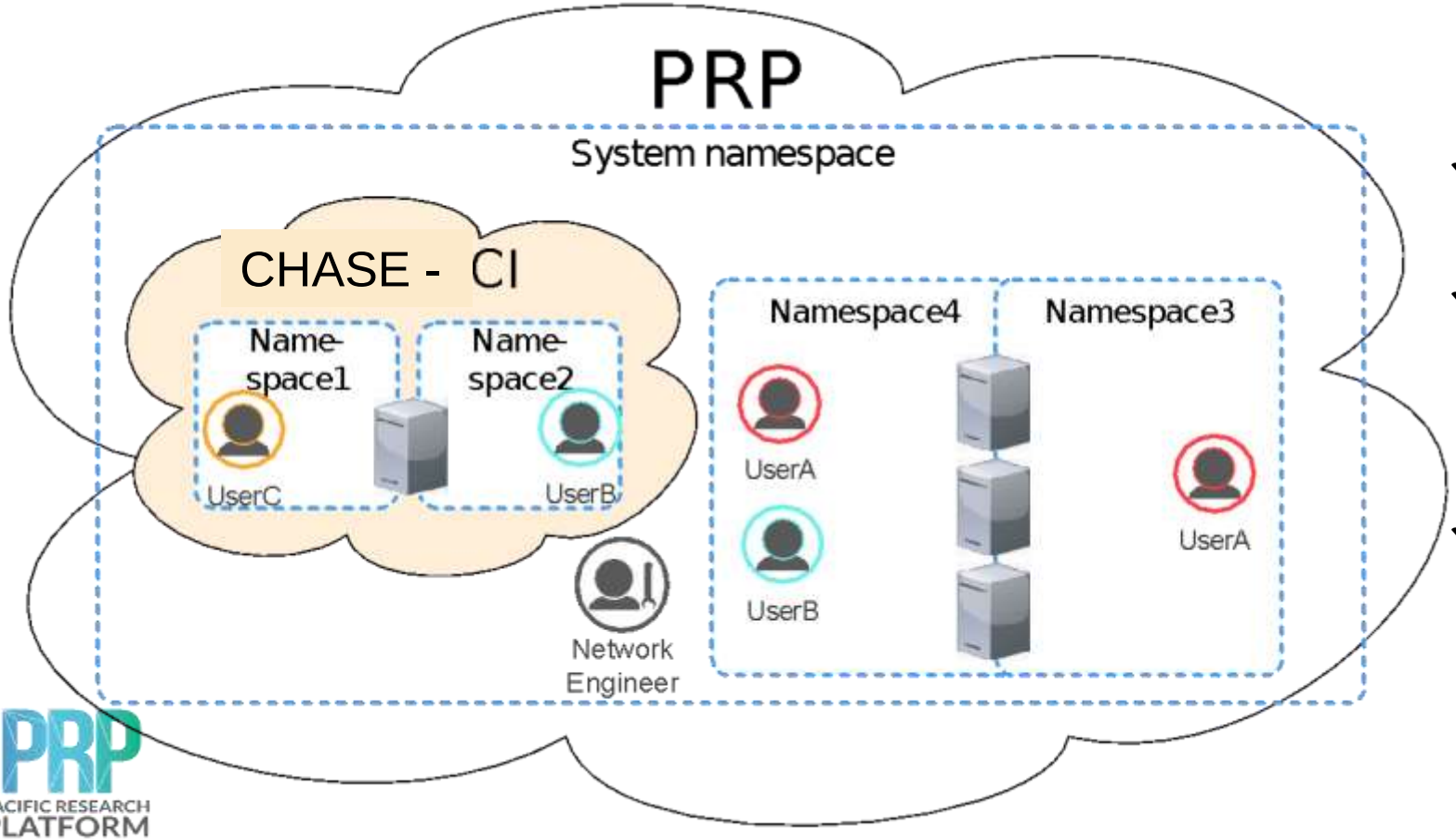
Nautilus Uses CILogon for Federated Identity Already Adopted by Campuses



PRP PACIFIC RESEARCH PLATFORM

it²

Nautilus Namespaces
Enable Secure for Multi-Institution Collaborations



- ✓ Isolate user space
- ✓ Create environment to collaborate
- ✓ Define policies

PRP PACIFIC RESEARCH PLATFORM

Nautilus is Using Calico
For Policy-Driven Network Security

Policy-driven network security

"A micro-firewall for every workload" minimizes attack surface



Perimeter security (edge firewalls) has been demonstrated time and again to be insufficient. That's why we built a security layer into Calico that enables developers and operations staff to easily define with fine granularity which connections are allowed, and which are not. These rules implement and extend the Kubernetes Network Policy API – but also work on all other platforms supported by Calico. They might separate development from production workloads, or limit access to a specific restricted service to ensure regulatory compliance. A distributed algorithm calculates which rules are required on each node in the cluster and updates them dynamically as workloads are created and terminated. As a result, malicious actors – or just errant applications – are detected and stopped before they can cause damage.

PRP
PACIFIC RESEARCH PLATFORM

it²

Installing FIONAs Across California in Late 2018 and Early 2019
To Enhance User's CPU and GPU Computing, Data Posting, and Data Transfers



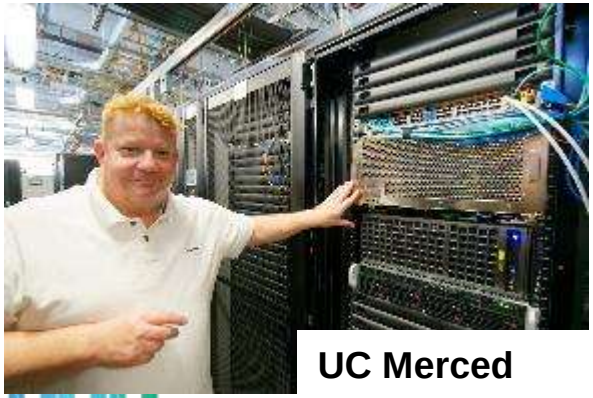
UC Irvine



Stanford



UC Santa Barbara



UC Merced



UC Santa Cruz



UC Riverside

PACIFIC RESEARCH PLATFORM

PRP's California Nautilus Hypercluster Connected by Use of CENIC 100G Network

Nautilus

Minority Serving Institution

PRP Disks

CHASE-CI

*= July RT

Calit2/UCI
4 FIONA8s*
40G 160TB
40G 160TB HPWREN

UCR
40G 160TB
1 FIONA8

USC
40G 160TB

UCLA
2x40G 160TB
100G NVMe 6.4TB

Caltech
100G NVMe 6.4TB

USD
40G 192TB

CSUSB
10G 3TB

UCSB
40G 192TB
2 FIONA8s*

CENIC

15-Campus Nautilus Cluster:
4360 CPU Cores 134 Hosts
~1.7 PB Storage
407 GPUs, ~4000 cores each

UCSC
40G 160TB
100G NVMe 6.4TB
4.5 FIONA8s

NPS
100G 48TB

SDSC @ UCSD
8 FIONA8s + 5 FIONA8s
100G Gold NVMe
100G Epyc NVMe

Stanford U
40G 160TB
1 FIONA8*

UCSD
2x40G 160TB HPWREN
17 FIONA8s
35 FIONA2s

SDSU
FPGAs + 2PB BeeGFS
1 FIONA8* 2 FIONA4s
100G NVMe 6.4TB
40G 160TB HPWREN

UCM
40G 160TB
2 FIONA8
10 FIONA2s

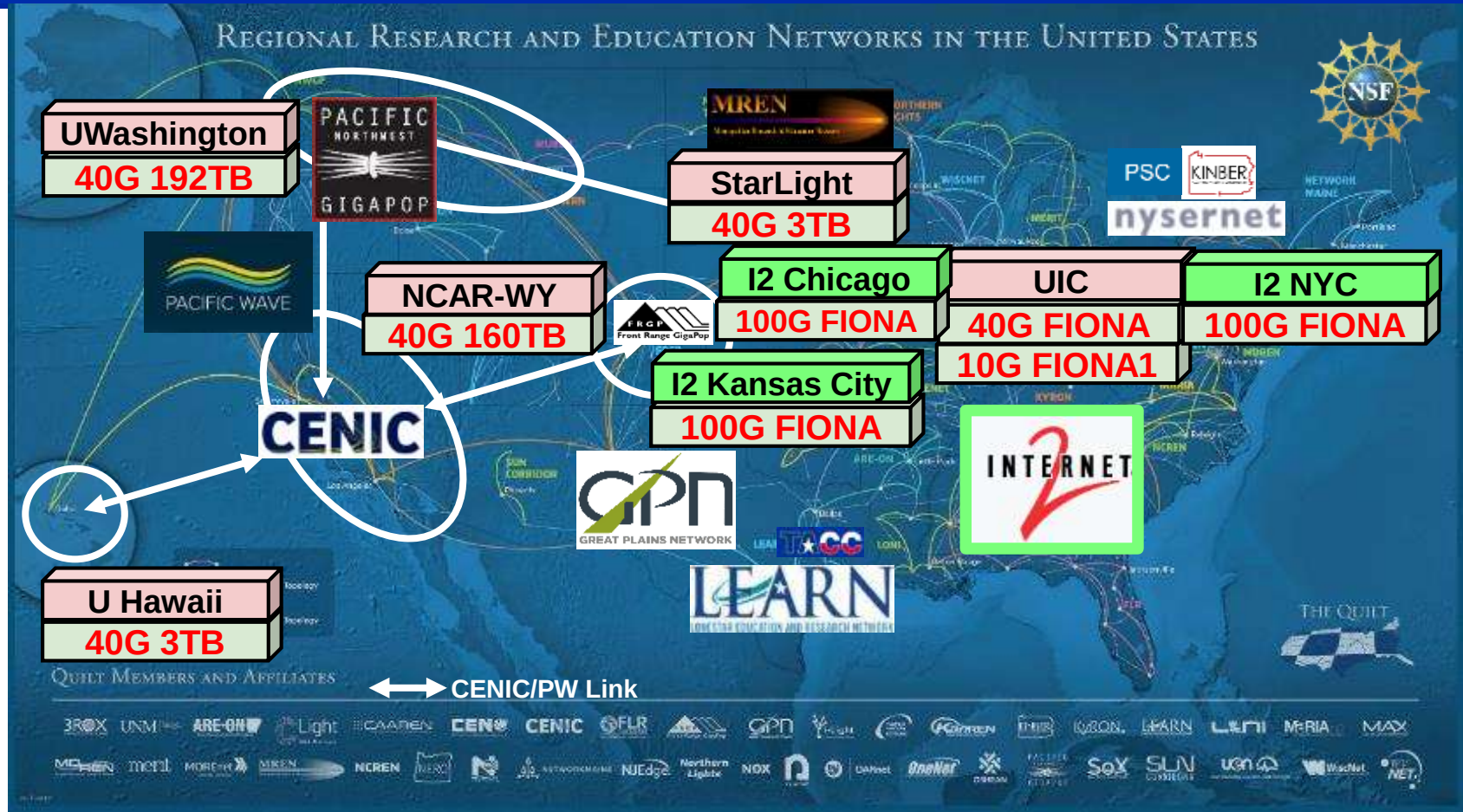
UCSF
40G 192TB

CITRIS BANATAO INSTITUTE

SDSC

it²

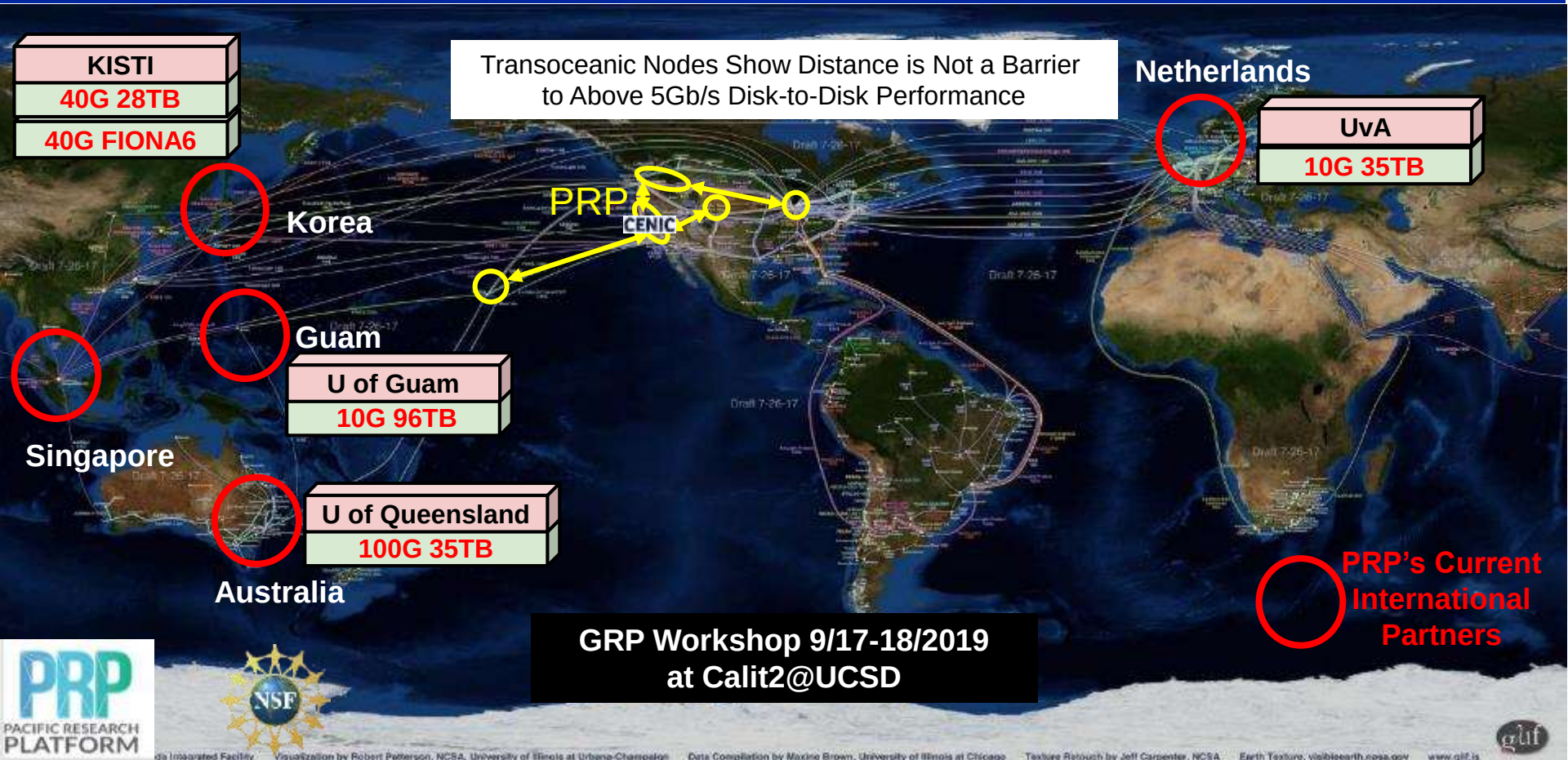
PRP/TNRP's United States Nautilus Hypercluster FIONAs Now Connects 4 More Regionals and 3 Internet2 Storage Sites



ESnet
CENIC
SDSC
CITRIS BANATAO INSTITUTE

PRP
PACIFIC RESEARCH PLATFORM

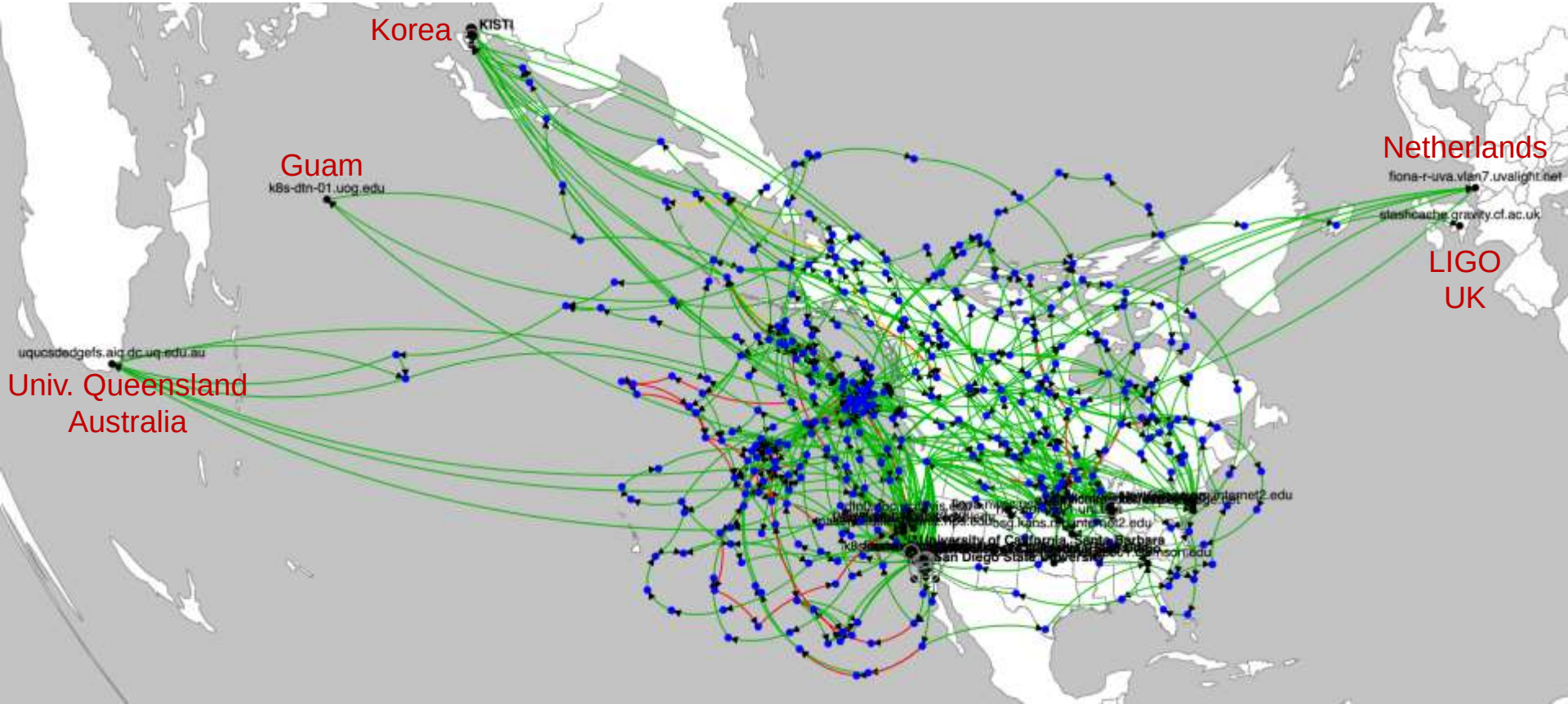
PRP Global Nautilus Hypercluster Is Rapidly Adding International Partners Beyond Our Original Partner in Amsterdam



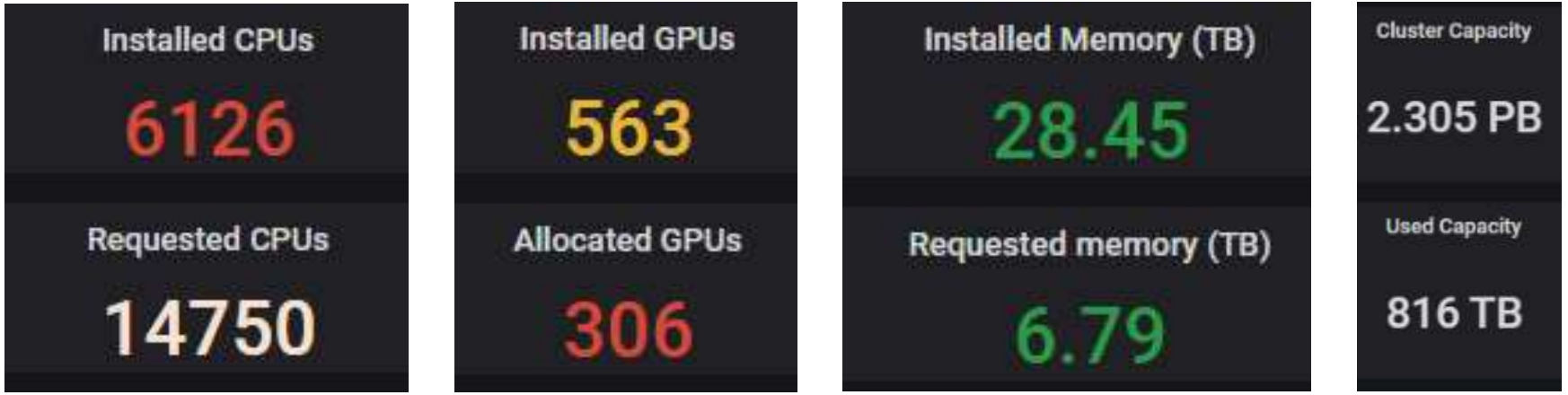
Operational Metrics: Containerized Trace Route Tool Allows Realtime Visualization of Status of PRP Network Links on a National and Global Scale

9/16/2019

Source: Dima Mishin, SDSC



PRP's Nautilus Forms a Multi-Application Powerful Distributed "Big Data" Storage and Machine-Learning Computer



Source: grafana.nautilus.optiputer.net on 1/27/2020

PRP is Science-Driven:
Connecting Multi-Campus Application Teams and Devices



CMS

Particle Physics

Biomedical ‘omics

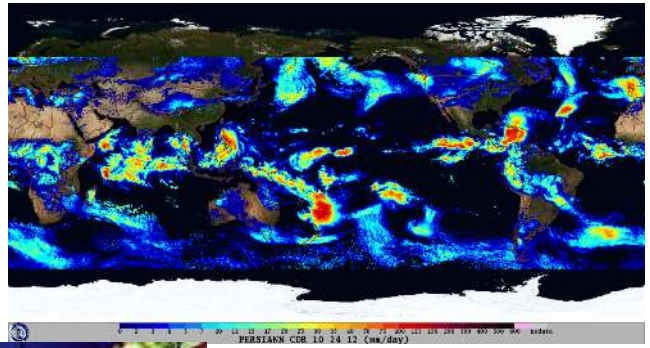


NSF



Telescope Surveys

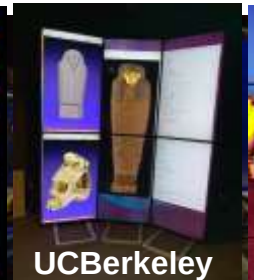
Earth Sciences



Visualization, Virtual Reality, Collaboration



UC San Diego



UCBerkeley



UC Merced

CITRIS
SDSC

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

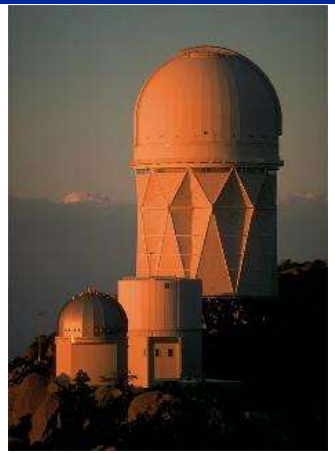
120GB per night

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

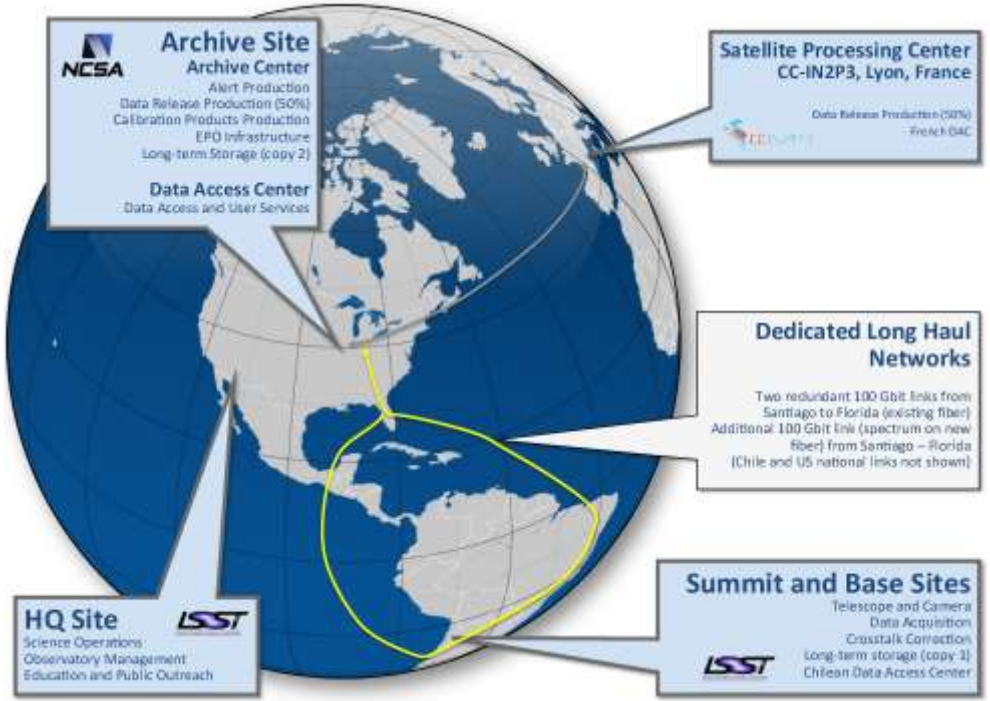
it²

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

Large Synoptic Survey Telescope (LSST)

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

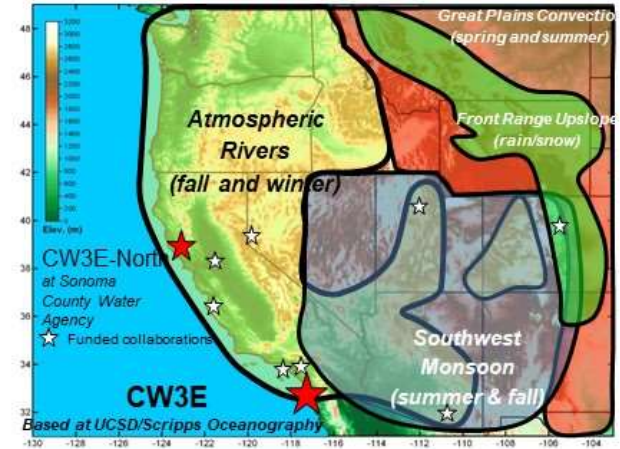




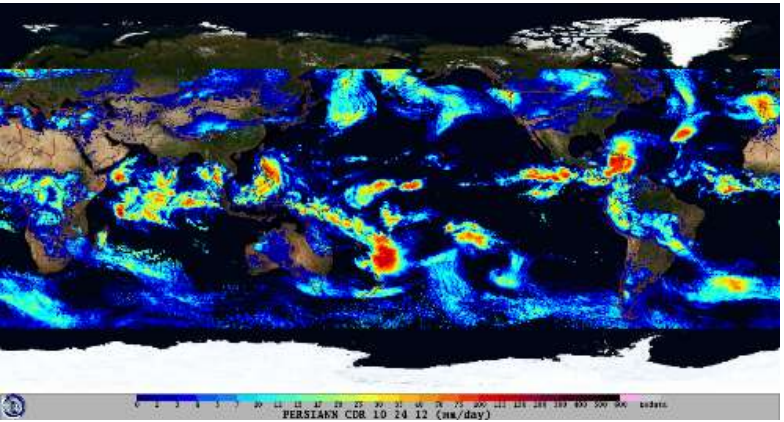
PRP
PACIFIC RESEARCH PLATFORM

Use PRP-Like CI to Connect NCSA Repository To Remote Astronomy Big Data Analysis Users?

Collaboration on Atmospheric Water in the West Between UC San Diego and UC Irvine



Key Phenomena Causing Extreme Precipitation in the Western U.S. (Ralph et al. 2014)



Center for Western Weather and Water Extremes
SCRIPPS INSTITUTION OF OCEANOGRAPHY AT UC SAN DIEGO

Director: F. Martin Ralph

Big Data Collaboration with:

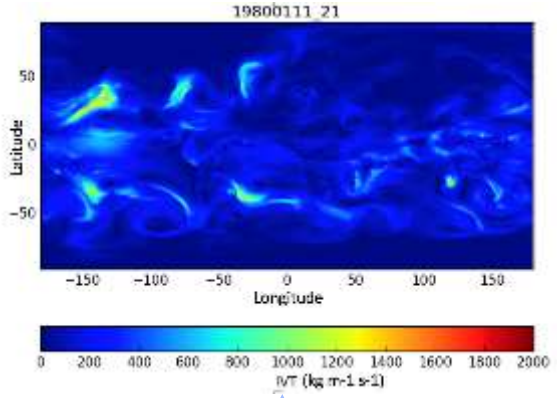
Center for Hydrometeorology & Remote Sensing
University of California, Irvine

Director, Soroosh Sorooshian, UCSD

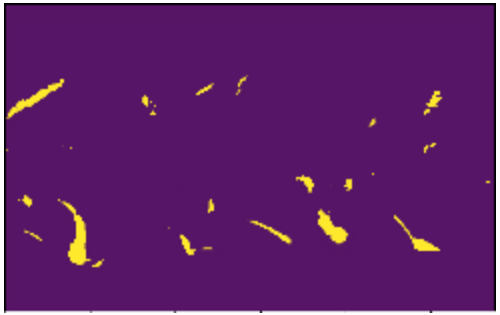
Source: Scott Sellers, PhD CHRS; Postdoc CW3E

Rapid 4D Object Segmentation of NASA Water Vapor Data - "Stitching" in Time and Space

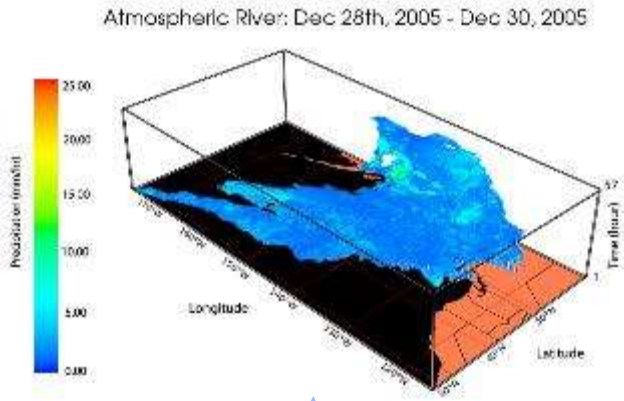
NASA *MERRA v2 – Water Vapor Data Across the Globe



Object Detection, Segmentation and Tracking



4D Object Constructed (Lat, Lon, Value, Time)



Scott L. Sellars[1], John Graham[1], Dima Mishin[1], Kyle Marcus[2] , Ilkay Altintas[2], Tom DeFanti[1], Larry Smarr[1], Joulien Tatar[3], Phu Nguyen[4], Eric Shearer[4], and Soroosh Sorooshian[4]

[1]Calit2@UCSD; [2]SDSC; *[3]Office of Information Technology, UCI; [4]Center for Hydrometeorology and Remote Sensing, UCI*

PRP Shortened Workflow From 19.2 Days to 52 Minutes - 532 Times Faster!

Pacific Research Platform (10-100 Gb/s)

Complete workflow time: 19.2 days→52 Minutes!

UC, Irvine

GPUs



Calit2's FIONA

SDSC's COMET



Source: Scott Sellers, CW3E

UC, San Diego

GPUs



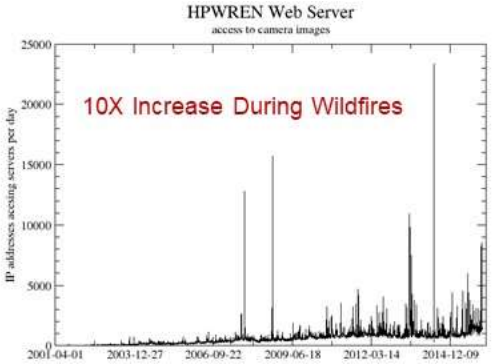
Calit2's FIONA

PRP Optical Fiber Connects Data Servers for High Performance Wireless Research and Education Network (HPWREN)

- **PRP Uses CENIC 100G Optical Fiber to Link UCSD, SDSU & UCI HPWREN Servers**
  - **Data Redundancy**
  - **Disaster Recovery**
  - **High Availability**
  - **Kubernetes Handles Software Containers and Data**



Source: Frank Vernon, Hans Werner Braun HPWREN



Data From Hans-Werner Braun



UCI Antenna Dedicated June 27, 2017

PRP PACIFIC RESEARCH PLATFORM
HPWREN
UC San Diego
SDSC SAN DIEGO SUPERCOMPUTER CENTER
SCRIPPS INSTITUTION OF OCEANOGRAPHY
Calit2
SAN DIEGO STATE UNIVERSITY

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



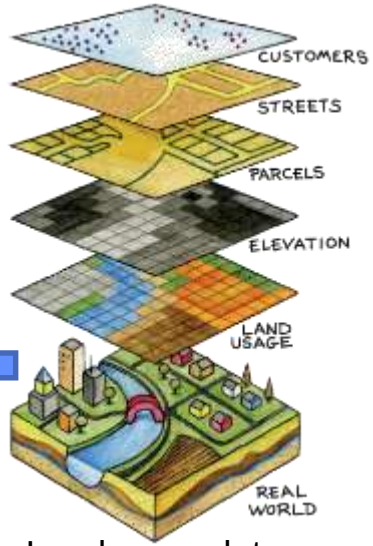
Real-Time Meteorological Sensors



Source: Ilkay Altintas, SDSC

WIFIRE

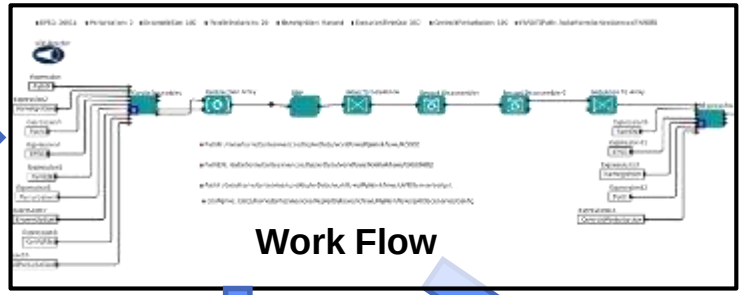




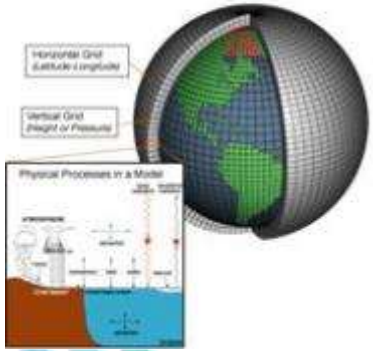
Landscape data

PRP





Weather Forecast



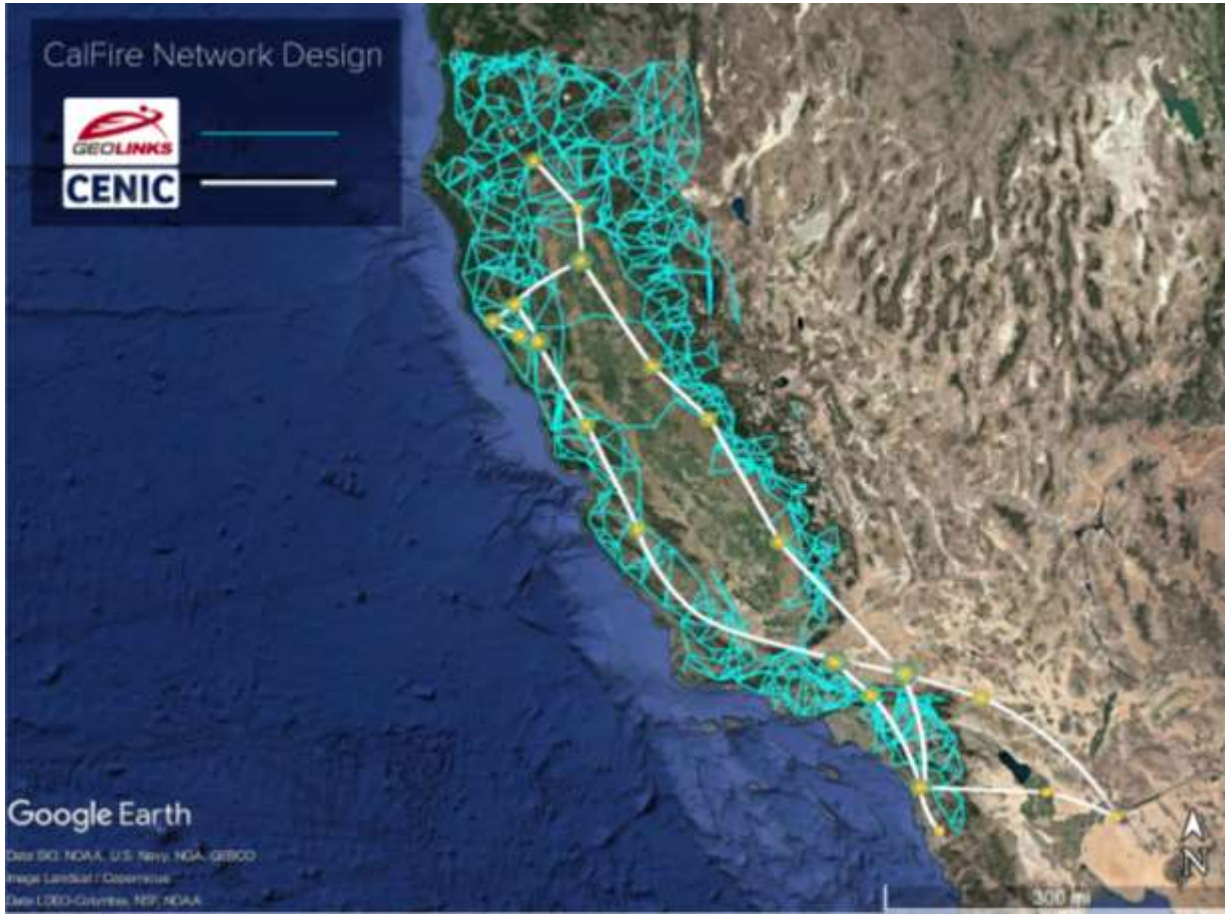






PRP
PACIFIC RESEARCH PLATFORM

California Public-Private Partnership Plan
To Scale Fixed Wireless With CENIC Optical Fiber Backhaul

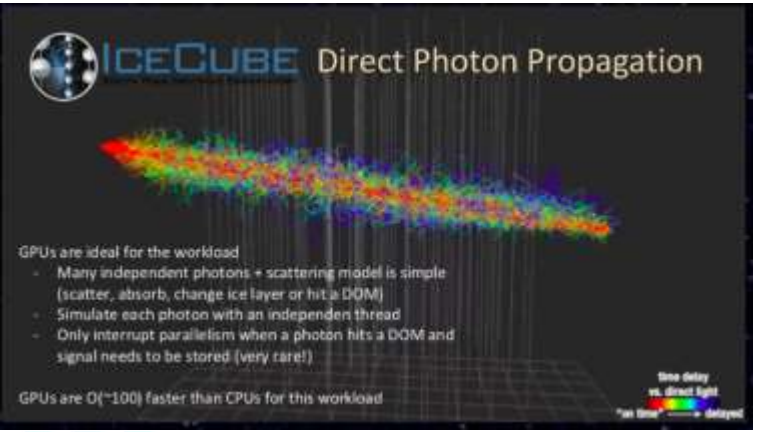


PRP
PACIFIC RESEARCH PLATFORM

it²

NSF Large-Scale Observatories
Are Beginning to Utilize PRP Compute Resources



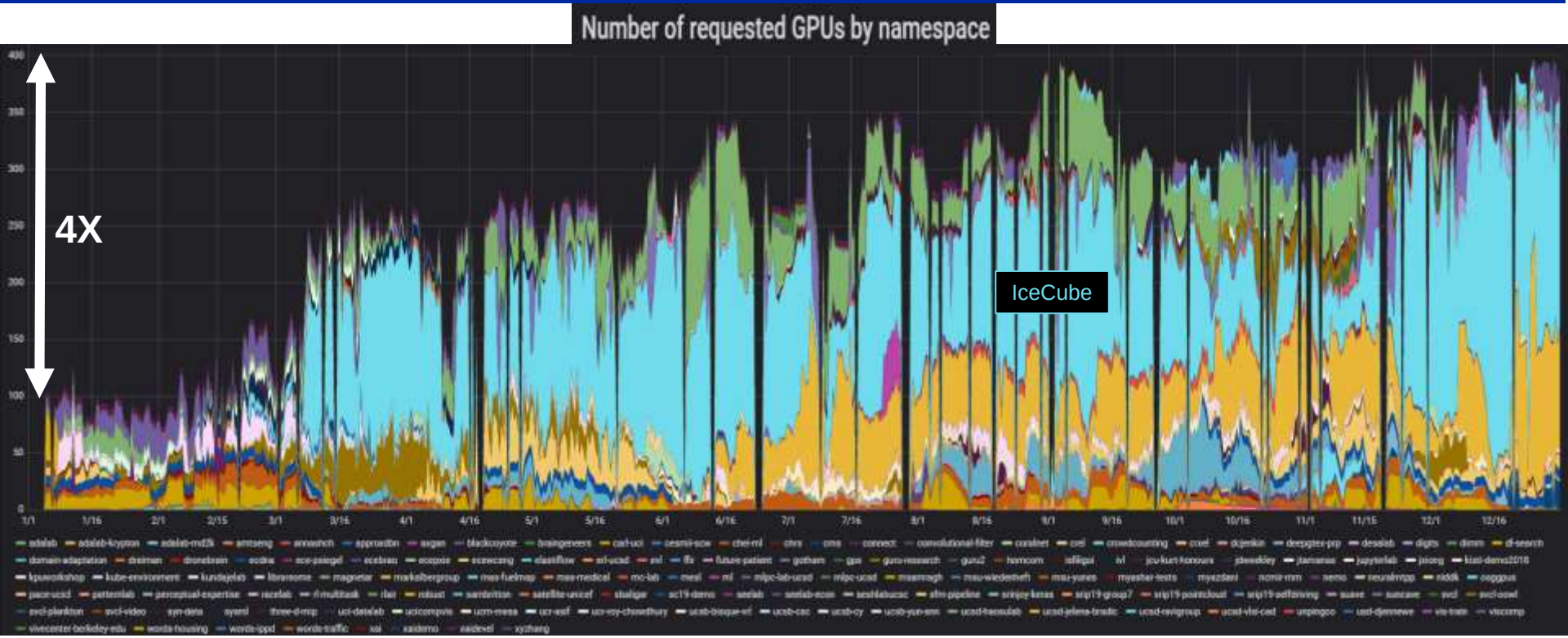


GPU Simulations Needed *to Improve Ice Model*.
=> Results in *Significant Improvement*
in Pointing Resolution for Multi-Messenger Astrophysics

PRP
PACIFIC RESEARCH PLATFORM

it²

Number of Requested PRP Nautilus GPUs Has Gone Up 4X in 2019



https://grafana.nautilus.optiputer.net/d/fHSeM5Lmk/k8s-compute-resources-cluster-gpus?orgId=1&fullscreen&panelId=2&from=1546329600000&to=1577865599000

Using Grafana to Track PRP Requested GPUs 10/1-12/31/19

Ravi Ramamoorthi, UCSD
Viscomp ML

Hao Su, UCSD
AI in Robotics

Nuno Vasconcelos, UCSD
OOWL Synthetic Data ML

Volkan Vural, UCSD
Workflow Analysis

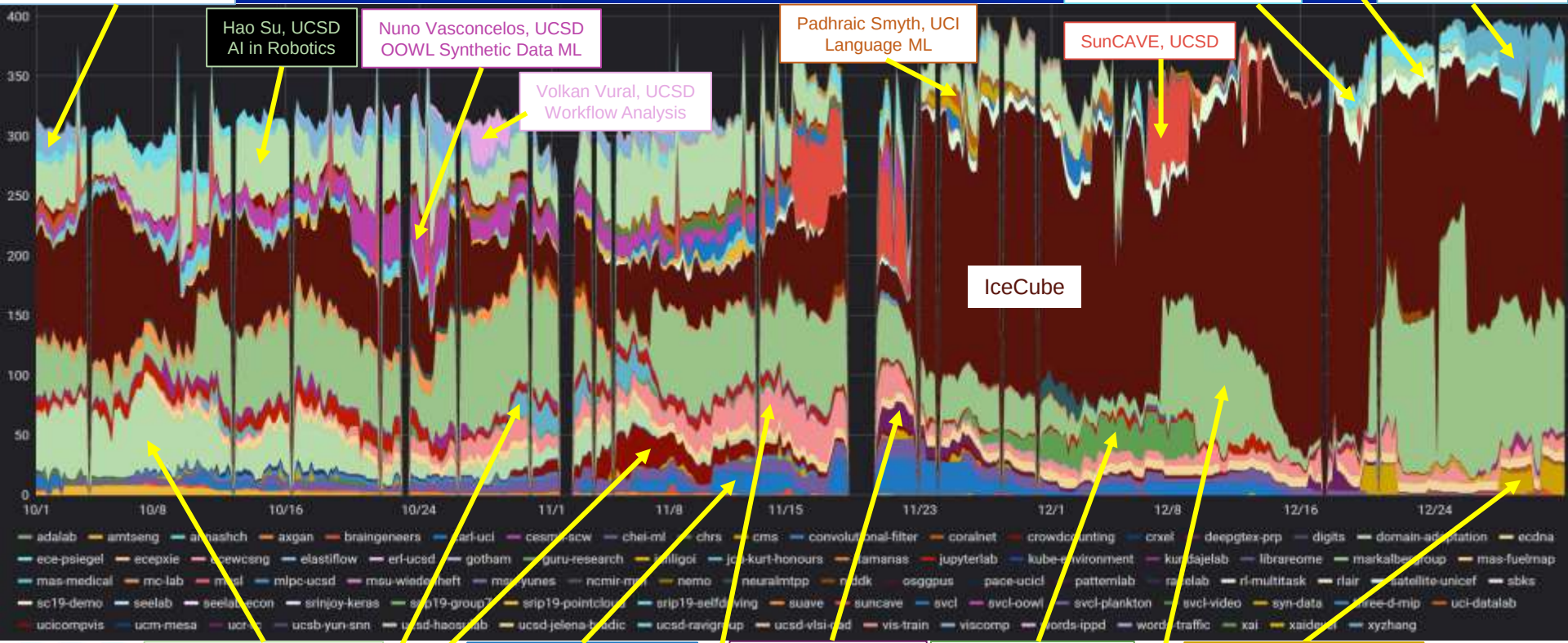
Padhraic Smyth, UCI
Language ML

SunCAVE, UCSD

Ravi Ramamoorthi, UCSD
CG ML

Xiaolong Wang, UCSD
Robot ML

Xinyu Zhang, UCSD
Wireless ML



Nuno Vasconcelos, UCSD
Domain-adaptation ML

Jeff Krichmar, UCI
Reinforcement Learning

Alex Feltus, Clemson
Oncogenomics

Gary Cottrell, UCSD
Deep Learning

Frank Wuerthwein, UCSD
LHC CMS

Kurt Schoenhoff, JCU
Semantic ML

Nuno Vasconcelos, UCSD
Crowdcounting Image ML

Dinesh Bharadia, UCSD
Autonomous Driving

Mark Alber Group, UCR
Quantitative Modeling in Biology

PRP
PACIFIC RESEARCH PLATFORM

it²

The Open Science Grid (OSG) Has Been Integrated With the PRP

Source: Frank Würthwein,
OSG Exec Director; PRP co-PI; UCSD/SDSC

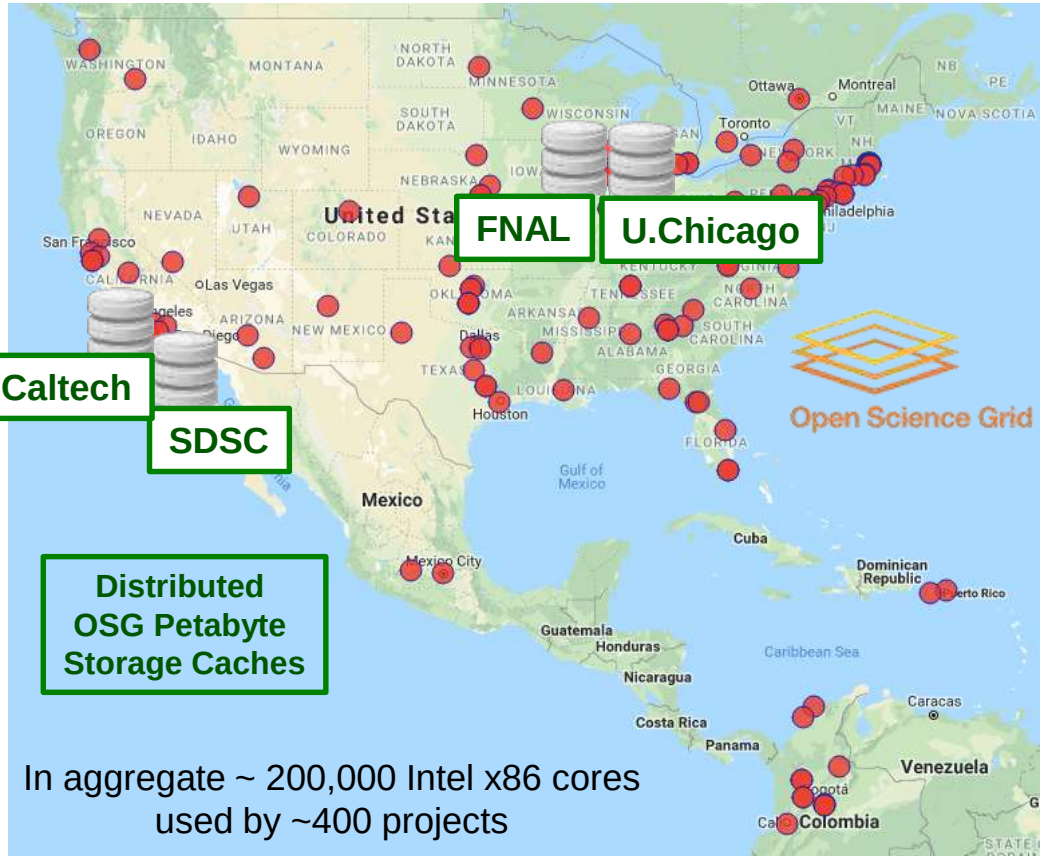
All OSG User Communities
Use HTCondor for Resource Orchestration

HTCondor
High Throughput Computing

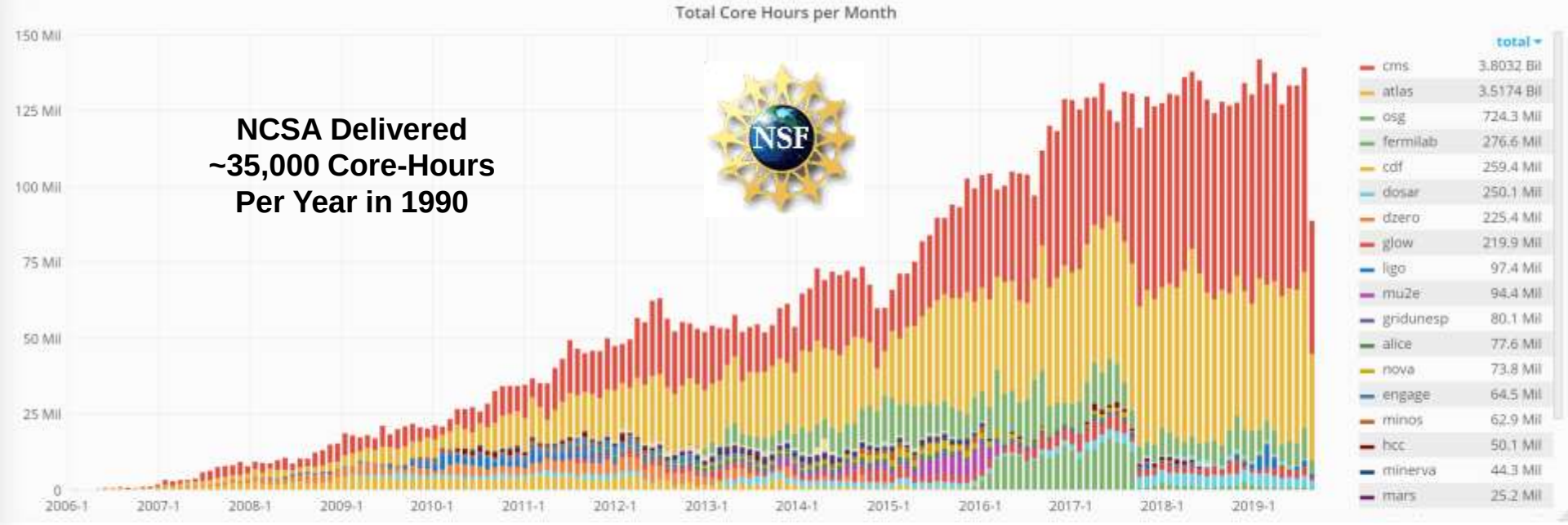
NSF

PRP
PACIFIC RESEARCH PLATFORM

OSG Federates ~100 Clusters Worldwide



The Open Science Grid Delivers
1.8 Billion Core-Hours Per Year to Over 50 Fields of Science



NCSA Delivered ~35,000 Core-Hours Per Year in 1990

Active Virtual Organizations	Active Sites	Represented Fields of Science
20	149	50

PRP PACIFIC RESEARCH PLATFORM

GRACC Grid Accounting

https://gracc.opensciencegrid.org/dashboard/db/gracc-home

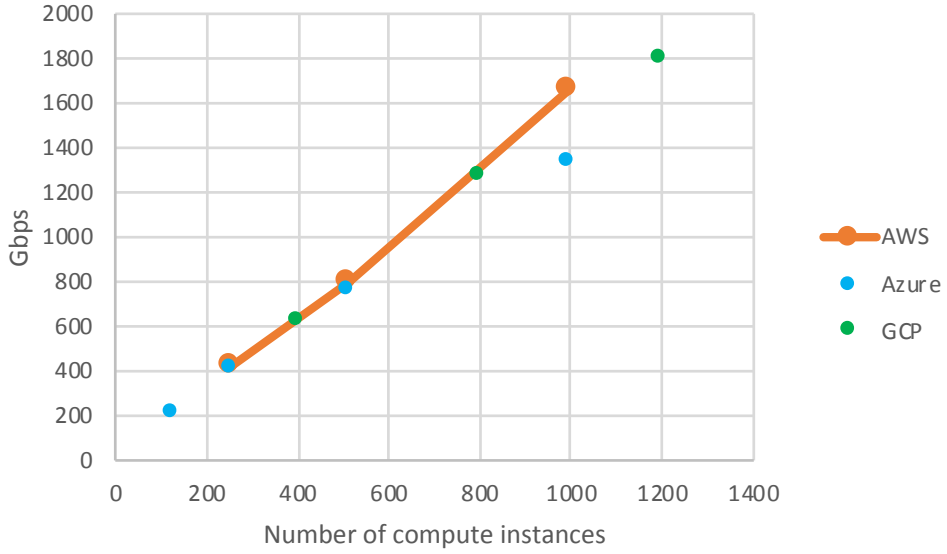
Running a 51k GPU Burst for Multi-Messenger Astrophysics with IceCube Across All Available GPUs in the Cloud

Frank Würthwein - OSG Executive Director, PRP co-PI
Igor Sfiligoi - Lead Scientific Researcher
UCSD/SDSC

The In-Cloud Network Testing Ahead of Time

In-Cloud Storage Showed Great Scalability

Exceeded 1 Tbps

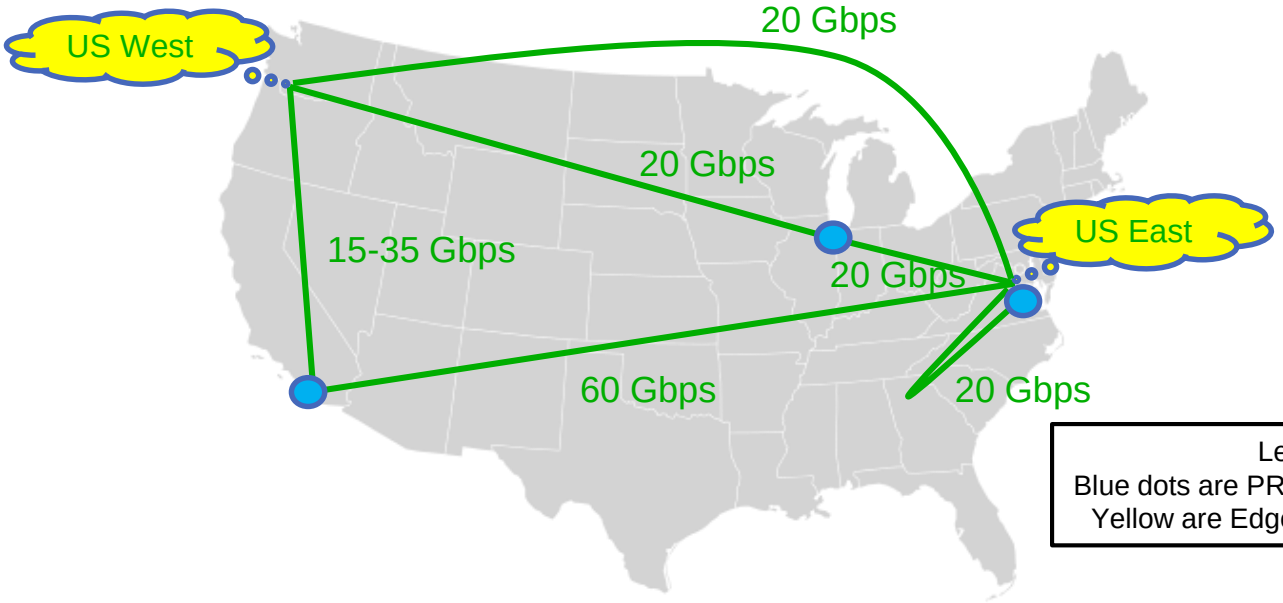


Probably Can Scale Much Higher

Networking Inside a Cloud Region Is Not a Concern!

The U.S. Network Testing Ahead of Time

Cloud to PRP/TNRP/OSG/I2 Networking in US is Good But Not Great.
At Least as Compared to the In-Cloud Networking

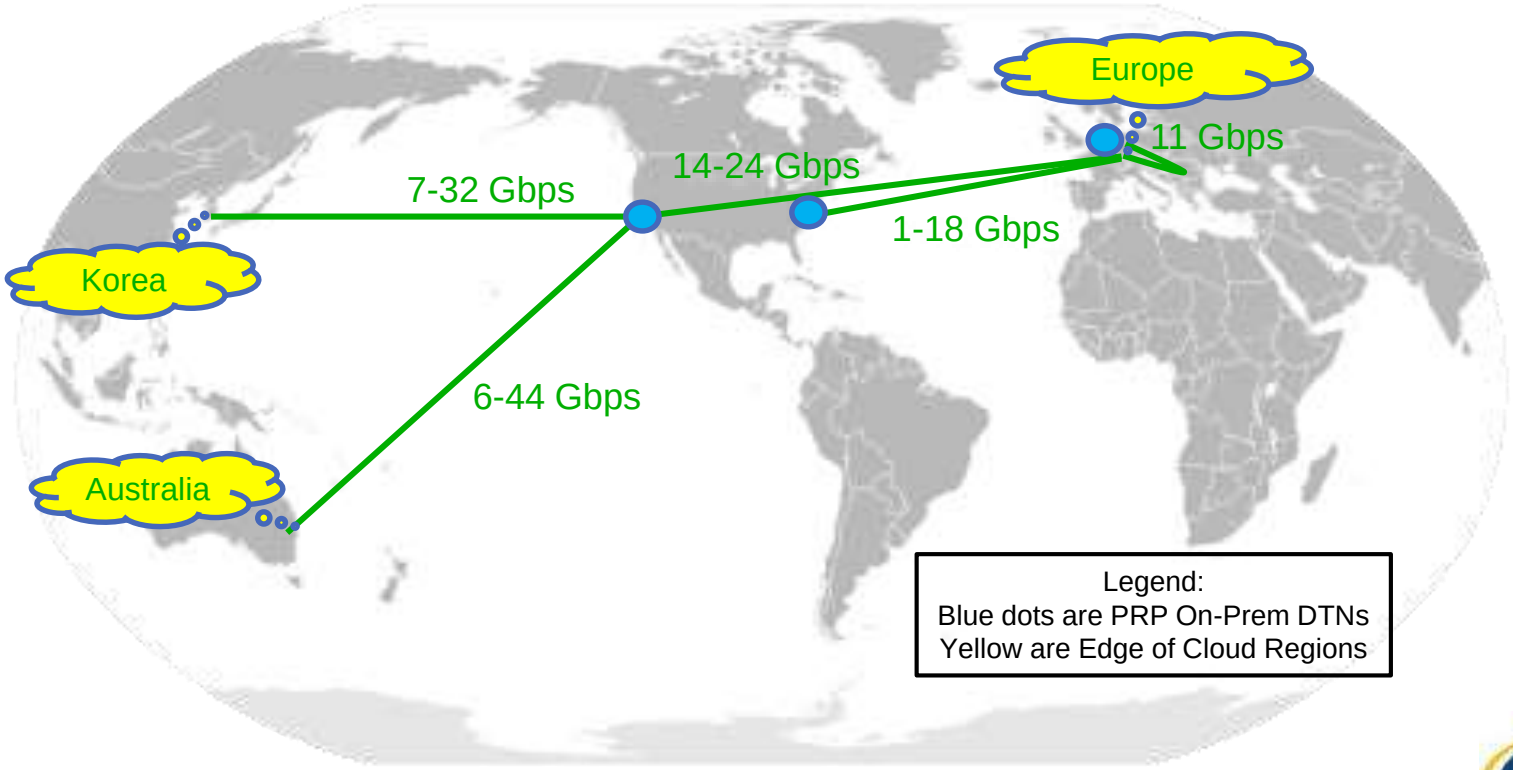


Legend:
Blue dots are PRP/I2 On-Prem DTNs,
Yellow are Edge of Cloud Regions

One of the Reasons for Pre-Staging the Inputs Inside of the Clouds

The Global Network Testing Ahead of Time

Cloud to On-Prem International Networking



Slide Source: Frank Würthwein, Igor Sfiligoi

The Idea

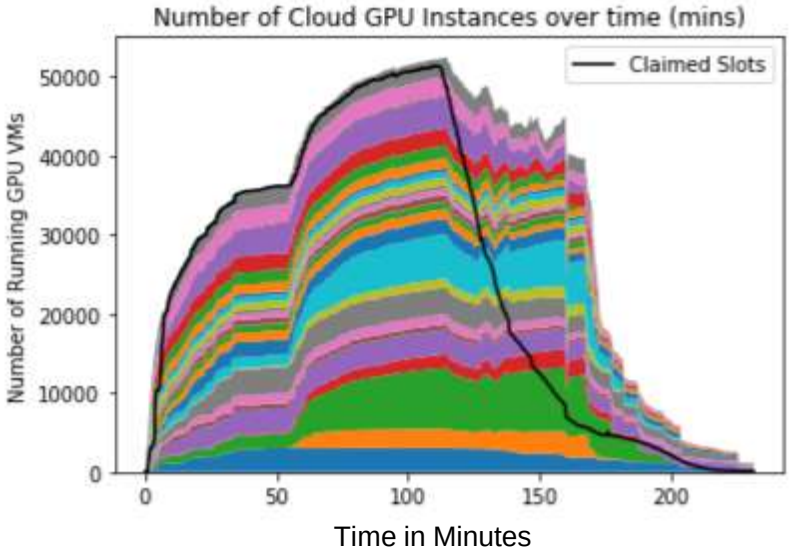
- **Integrate All GPUs Available for Sale Worldwide into a Single HTCondor Pool**
  - **Use 28 Regions Across AWS, Azure, and Google Cloud for a Burst of a Couple Hours, or so**
- **IceCube Submits Their Photon Propagation Workflow to this HTCondor Pool.**
  - **The Input, Jobs on the GPUs, and Output are All Part of a Single Globally Distributed System**
  - **This Demo Used Just the Standard HTCondor Tools**

Run a GPU Burst Relevant in-Scale for Future Exascale HPC Systems

PRP PACIFIC RESEARCH PLATFORM

it²

Science with 51,000 GPUs Achieved as Peak Performance



Each Color is a Different Cloud Region in US, EU, or Asia.

Total of 28 Regions in Use

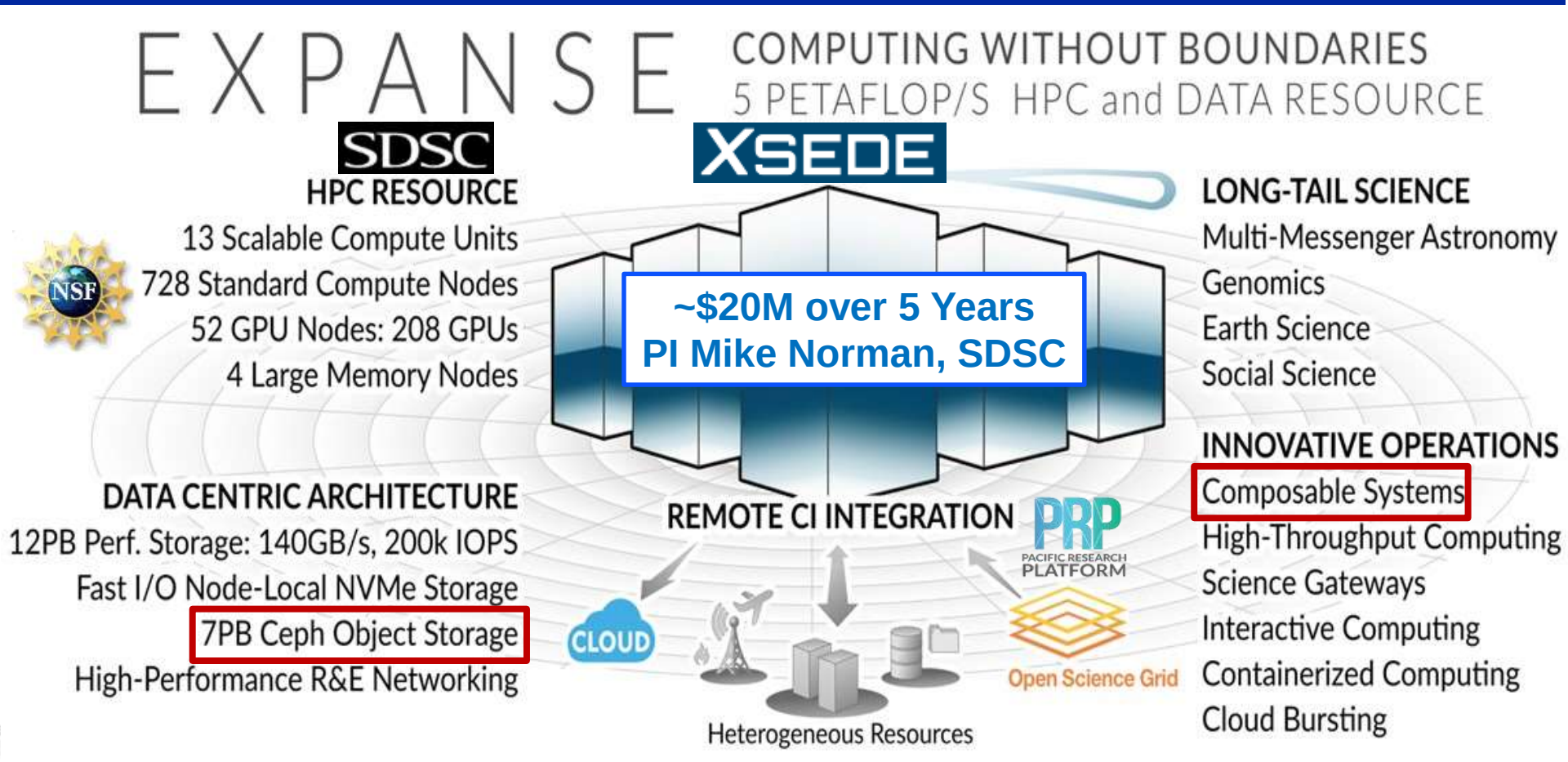
Peaked at 51,500 GPUs

~380 Petaflops of FP32

Summary of Stats at Peak - 8 Generations of NVIDIA GPUs Used

	V100	P100	P40	P4	T4	M60	K80	K520
Num GPUS	9.2k	7.2k	2.1k	0.5k	4.6k	10.1k	12.5k	5.4k
PFLOP32s	132.2	68.1	25.2	2.5	38.6	48.8	51.6	12.4

2020-2025 NRP Future:
SDSC's EXPANSE Will Use CHASE-CI Developed Composable Systems



PRP/TNRP/CHASE-CI Support and Community:

- US National Science Foundation (NSF) awards to UCSD, NU, and SDSC
  - CNS-1456638, CNS-1730158, ACI-1540112, ACI-1541349, & OAC-1826967
  - OAC 1450871 (NU) and OAC-1659169 (SDSU)
- UC Office of the President, Calit2 and Calit2's UCSD Qualcomm Institute
- San Diego Supercomputer Center and UCSD's Research IT and Instructional IT
- Partner Campuses: UCB, UCSC, UCI, UCR, UCLA, USC, UCD, UCSB, SDSU, Caltech, NU, UWash UChicago, UIC, UHM, CSUSB, HPWREN, UMo, MSU, NYU, UNeb, UNC,UIUC, UTA/Texas Advanced Computing Center, FIU, KISTI, UVA, AIST
- CENIC, Pacific Wave/PNWGP, StarLight/MREN, The Quilt, Kinber, Great Plains Network, NYSERNet, LEARN, Open Science Grid, Internet2, DOE ESnet, NCAR/UCAR & Wyoming Supercomputing Center, AWS, Google, Microsoft, Cisco

PRP PACIFIC RESEARCH PLATFORM

it²