“Toward a National Research Platform to Enable Data-Intensive Computing”

Virtual Data Science Seminar
Institute for Data Science
New Jersey Institute of Technology
October 27, 2021

Dr. Larry Smarr
Founding Director Emeritus, California Institute for Telecommunications and Information Technology;
Distinguished Professor Emeritus, Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD
http://lsmarr.calit2.net

PACIFIC RESEARCH PLATFORM

Abstract

Three current NSF grants [the Pacific Research Platform (PRP), the Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI), and Toward a National Research Platform (TNRP)] create a regional, national, to global-scale cyberinfrastructure, optimized for machine learning research and data analysis of large scientific datasets. This integrated system, which is federated with the Open Science Grid and multiple supercomputer centers, uses 10 to 100Gbps optical fiber networks to interconnect, across 30 campuses, nearly 200 Science DMZ Data Transfer Nodes (DTNs). The DTNs are rack-mounted PCs optimized for high-speed data transfers, containing multicore-CPUs, two to eight GPUs, and up to 256TB of disk each. Users' containerized software applications are orchestrated across the highly instrumented PRP by open-source Kubernetes, enabling easy access to commercial clouds as needed. I will describe several of the most active of PRP's 400 user namespaces, which support a wide range of data-intensive disciplines.

PRP
PACIFIC RESEARCH PLATFORM

36 Years Ago,
NSF Adopted a DOE High-Performance Computing Model







NCSA Was Modeled on LLNL

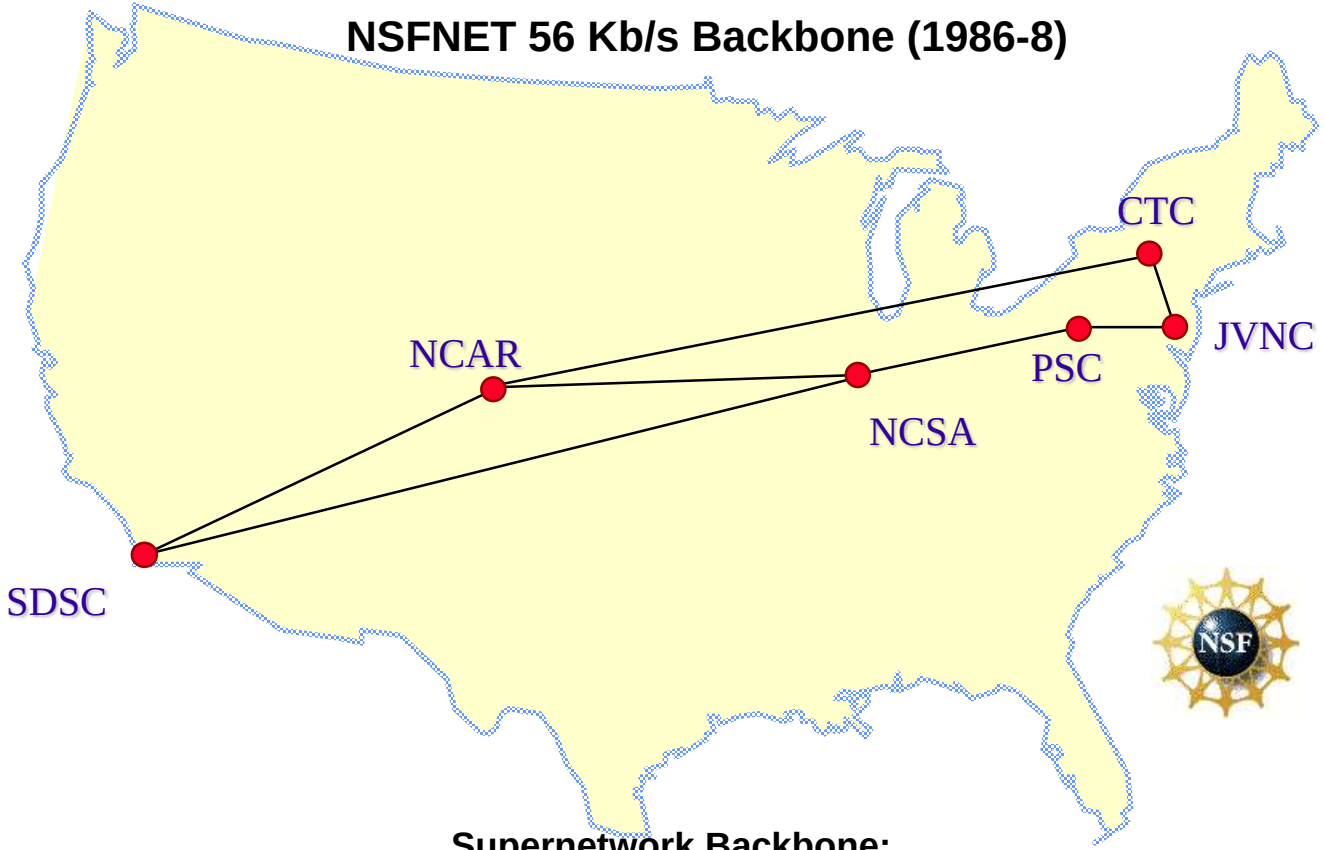
SDSC Was Modeled on MFEnet

PRP
PACIFIC RESEARCH PLATFORM

NCSA

SDSC

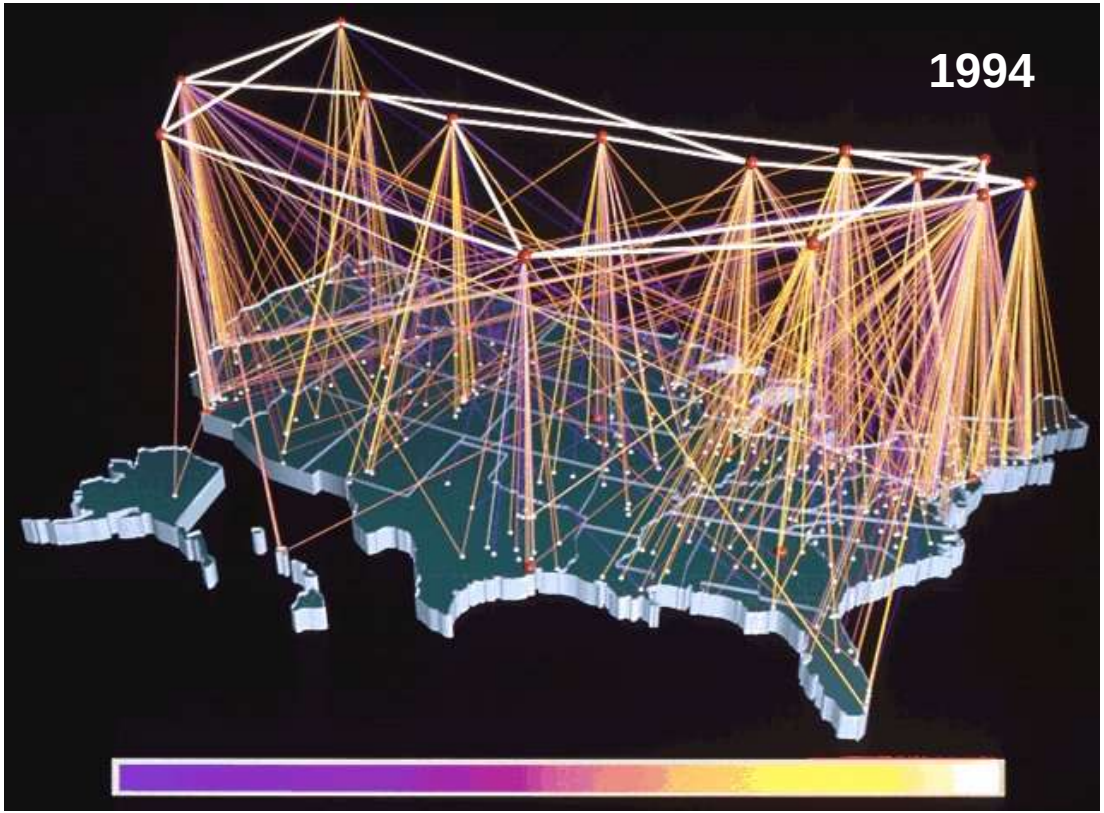
Launching the Nation's Information Infrastructure: NSFnet Supernetwork and the Six NSF Supercomputers



Supernetwork Backbone:
56kbps is 50 Times Faster than 1200 bps PC Modem!

PACIFIC RESEARCH PLATFORM

From Supercomputer Centers to the NSFnet to Today's Commercial Internet

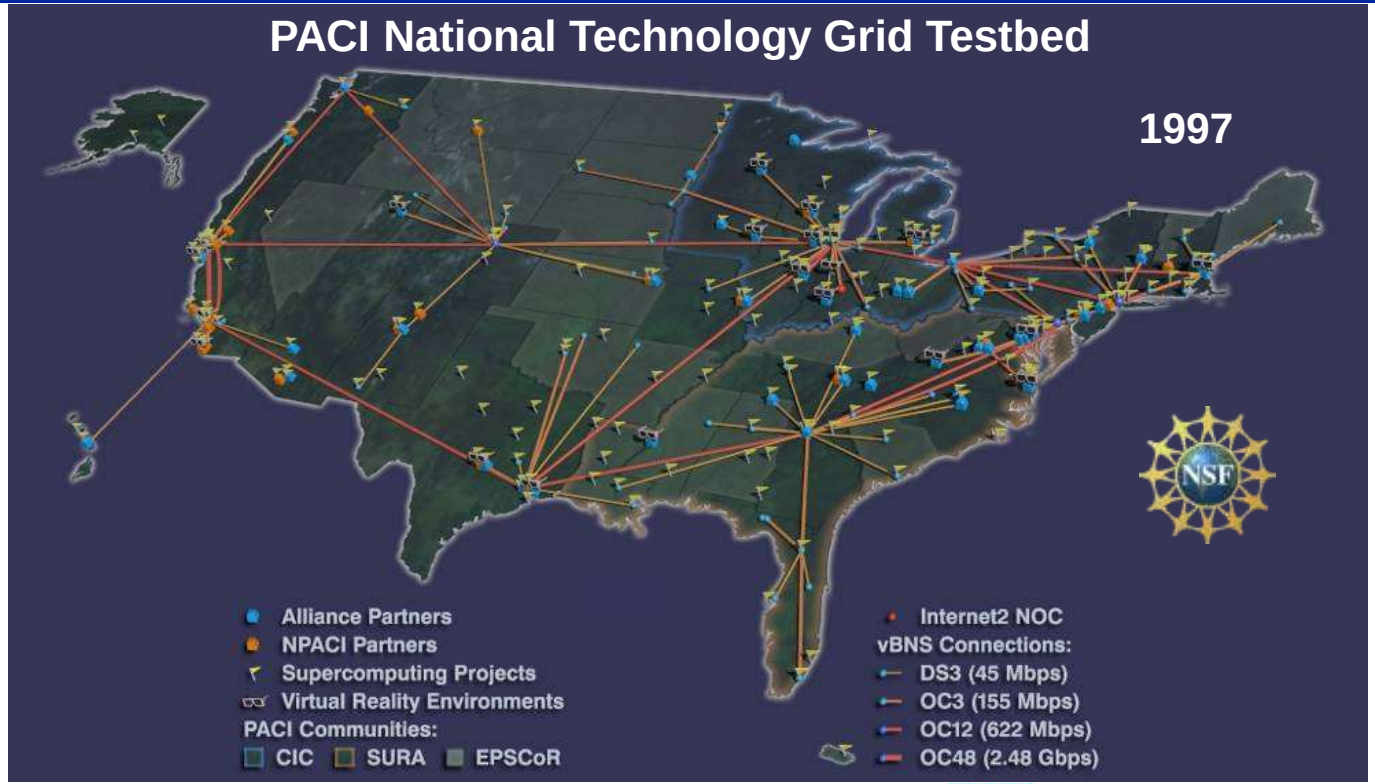


Visualization by NCSA's Donna Cox and Robert Patterson
Traffic on 45 Mbps Backbone December 1994

PACIFIC RESEARCH PLATFORM

NSF's PACI Program was Built on the vBNS to Prototype America's 21st Century Information Infrastructure

Key Role of Miron Livny & Condor



vBNS led to

INTERNET2

PRP PACIFIC RESEARCH PLATFORM

NPACI

National Computational Science ALLIANCE

Dave Bader Created the First Linux COTS Supercluster -*Roadrunner*- on the National Technology Grid, with the Support of NCSA and NSF

ALBUQUERQUE JOURNAL

YOUTH CRIME
Bill would fund state programs that crack down on juvenile offenders • D3

Metro & New Mexico

THURSDAY, APRIL 8, 1999

BUSINESS • 2 SECTION D
WEATHER • 8
DEATHS • 8

UNM To Crank Up $400,000 Supercomputer Today

Machine One of 100 Speediest in World

By John Fleck
Journal Staff Writer

A cluster of computers looking like a stack of oversized dorm room refrigerators is about to plug the University of New Mexico into an ambitious new supercomputer "mega-brain" spanning the nation.

Dignitaries from around the country will gather at UNM's Albuquerque High Performance Computing Center today to throw the switch on the "Roadrunner Supercluster."

Fresh out of the box, the Roadrunner is one of the 100 fastest computers in the world.

But its rocket-sled speed is just part of a more ambitious idea — an attempt to give scientists around the country access to a distributed supercomputing "brain" stretching from Boston to Maui, ready to solve their toughest problems.

The Roadrunner Supercluster will be part of what researchers are calling the "National Technology Grid," a collection of supercomputers across the country wired together to help handle scientists' growing demand for computer time.

Sitting at their own desks, the researchers could hook in to the grid and begin solving large scale problems without having to worry about where the computer doing the work is located, explained Larry Smarr, director of the National Computational Science Alliance, the federally funded group that paid for the new machine.

"The goal is to be completely seamless," the University of Illinois-based Smarr said in telephone interview this week.

UNM professor David Bader envisions a day when a researcher studying an ecosystem — Chesapeake Bay, say, or the Rio Grande Valley — can log in and rapidly spin through vast amounts of satellite data on climate and ground cover for their part of land.

"They wouldn't need to be located at a specialized supercomputer research center to do their work, explained Smarr — they could just plug into the grid and let the computers take care of the rest, much as you plug a toaster into the wall without having to worry about which power plant is generating the electricity.

"That's the vision," said Bader, who has spent the last week testing the brand new Roadrunner machine in preparation for today's unveiling.

Smarr, one of the leaders of the nation's supercomputer research community, is scheduled to join Sen. Pete Domenici, R-N.M., and others to formally throw the switch turning on the computer at a ceremony this afternoon.

The $400,000 computer, built by Alta Technology Corp., bears the mark of a new breed of mainstream priced machines that are making inroads in the high-performance research market.

Instead of using specialized high-performance computer chips made especially for supercomputing, it's built around 128 top-of-the-line Intel Pentiums, the same breed of computer chips used in desktop computers.

Programmers split up a large problem into smaller pieces, farming them out so that each of the 128 chips can work on a portion of the problem and then share results.

"You get to profit from the cultural of PC prices," explained Victor Eijkhout, a computer scientist at the New Mexico Institute of Mining and Technology who does research on a similar "cluster" machine at Tech.

Similar machines also have been built at Sandia and Los Alamos national labs for nuclear weapons and physics research, and the experience of the New Mexico research community is one reason UNM was chosen to host the new machine, Smarr said.

UNM is one of six "supernodes" on the National Technology Grid. In addition to the Roadrunner Cluster, the grid will be connected to UNM's Maui High Performance Computing Center, an Air Force-funded research complex in Hawaii run by the university for the military.



BIG BRAIN: University of New Mexico professor David Bader tests the Roadrunner Supercluster. It will be dedicated today.

NCSA Director Larry Smarr (left), UNM President William Gordon, and U.S. Sen. Pete Domenici turn on the Roadrunner supercomputer in April 1999



David Bader Selected to Receive the 2021 IEEE Computer Society Sidney Fernbach Award

PRP PACIFIC RESEARCH PLATFORM

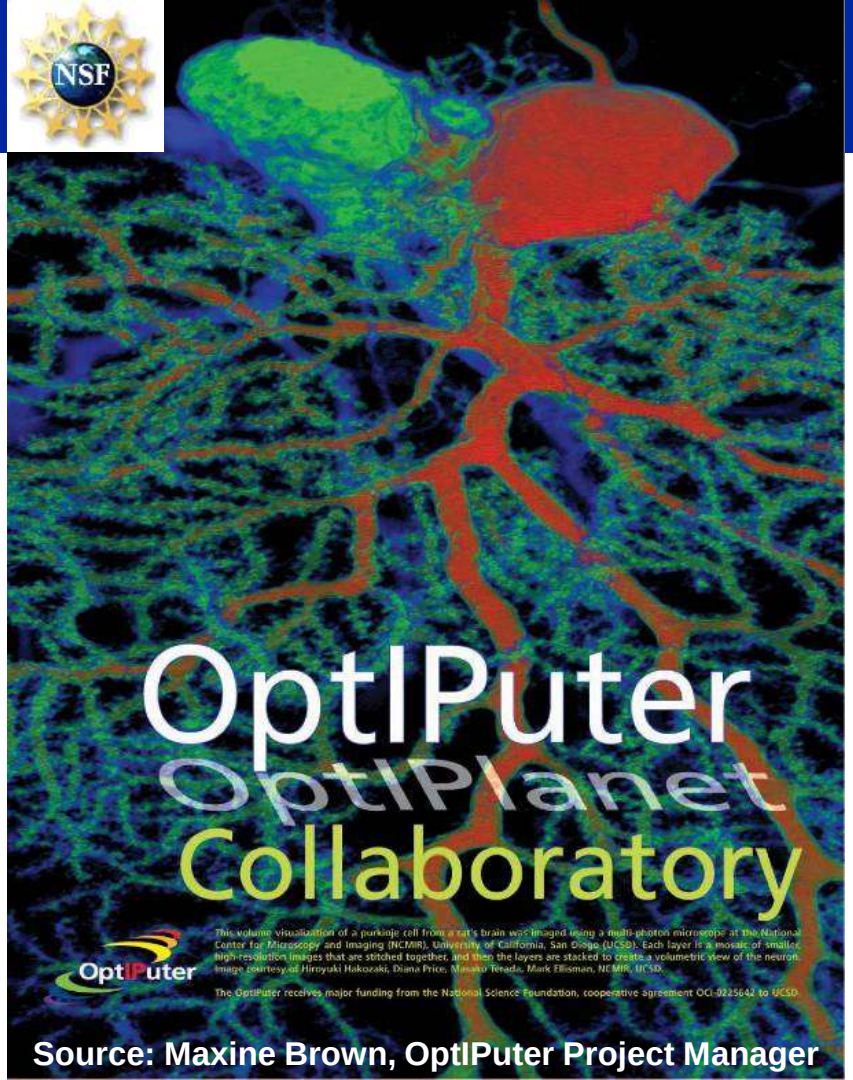
National Computational Science ALLIANCE

The 25 Years From the National Techology Grid To the National Research Platform

> The goal is a powerful, persistent, distributed U.S. national resource as usable in the future as the NSF Supercomputer Center standalone machines were in the 1980s and 1990s.

PACIFIC RESEARCH PLATFORM

From ***I-WAY to the National Technology Grid***, CACM, 40, 51 (1997)
Rick Stevens, Paul Woodward, Tom DeFanti, and Charlie Catlett



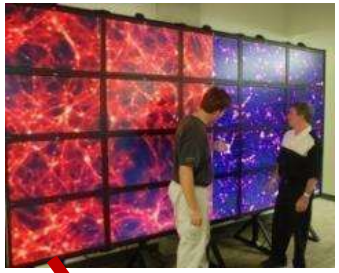
Source: Maxine Brown, OptIPuter Project Manager

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

Demonstrating That
Wide-Area Bandwidth
Can Equal
Local Cluster Backplane Speeds

PI Smarr,
2002-2009

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



End User OptIPortal



Instruments



HD/4k Telepresence



HD/4k Video Cams



Campus Optical Switch

10G Lightpath

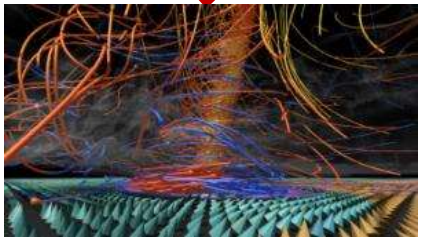


HPC

TeraGrid



Data Repositories & Clusters

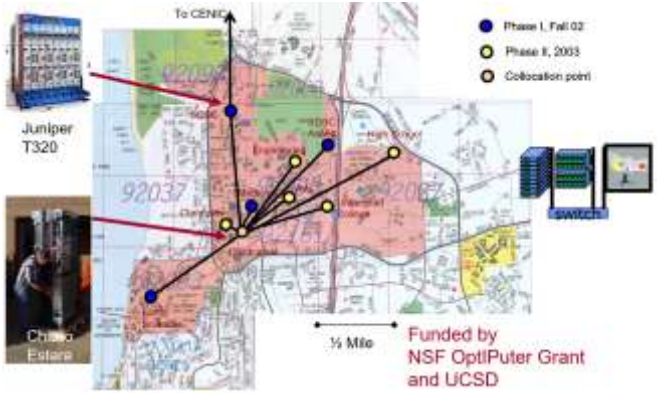


HD/4k Video Images

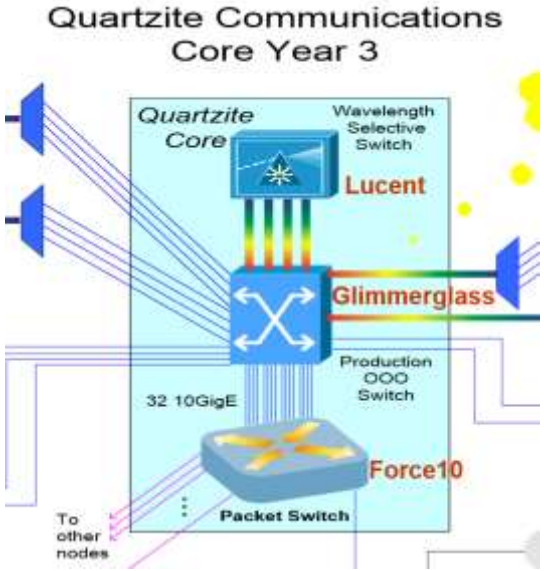
PRP
PACIFIC RESEARCH PLATFORM

PRP Was Built on 15 Years of NSF Awards: OptIPuter, Quartzite, & Prism

Precursors to DOE Defining DMZ in 2010



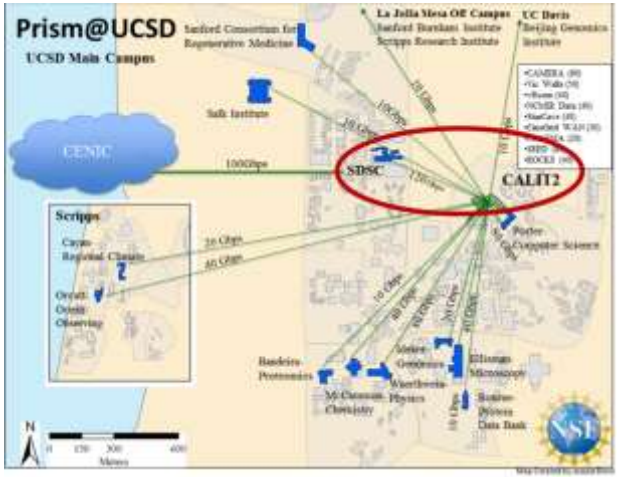
PI Smarr, 2002-2009



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

PI Papadopoulos, 2004-2007

Led to NSF CC* Award in 2013



PI Papadopoulos, 2013-2015

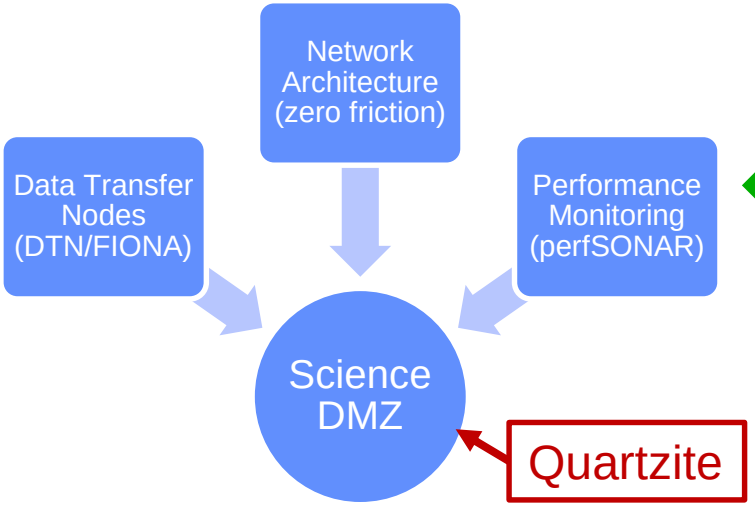
PRP PACIFIC RESEARCH PLATFORM

SDSC

20 YEARS QUALCOMM INSTITUTE

9 Years Ago,
NSF Adopted a DOE High-Performance Networking Model

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 340 Awards



Prism

2012-2020:
Across 50 States and Territories

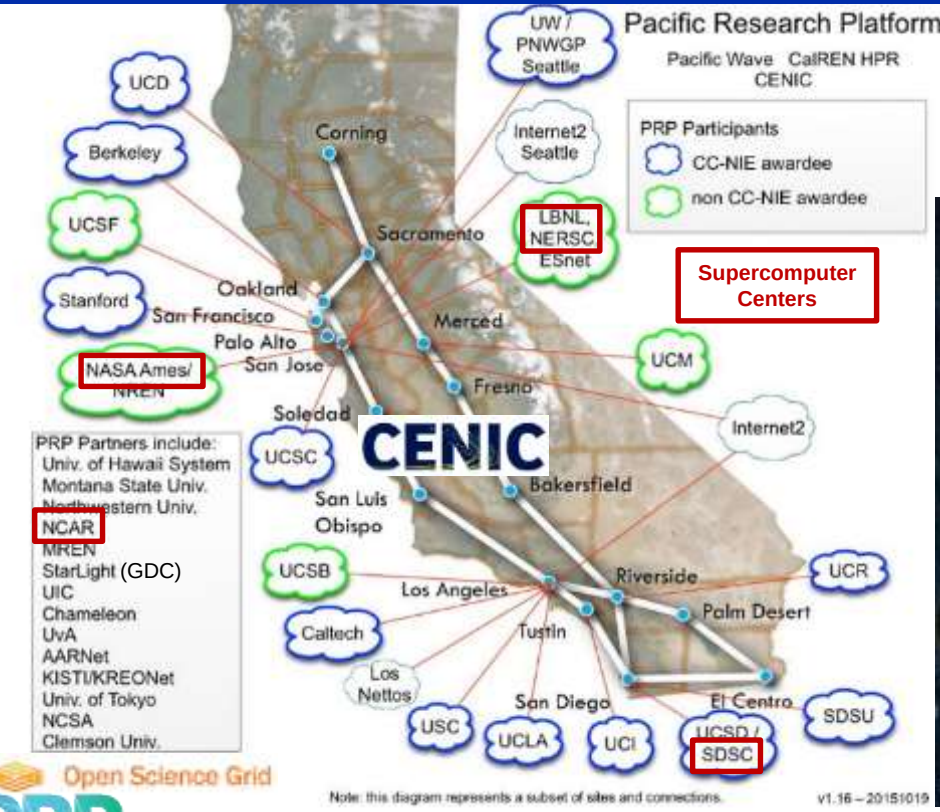
Slide Adapted From Inder Monga, ESnet

PRP PACIFIC RESEARCH PLATFORM

ESnet FASTERDATA KNOWLEDGEBASE

NSF

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



NSF CC*DNI Grant
$6.3M 10/2015-10/2020
In Year 6 Now, Year 7 is Funded



PRP PACIFIC RESEARCH PLATFORM

Source: John Hess, CENIC

CITRIS BANATAO INSTITUTE SDSC it²

2015-2021: UCSD Designs PRP Data Transfer Nodes (DTNs) -- Flash I/O Network Appliances (FIONAs)

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

Today's Roadrunner!

FIONA

(Flash Input/Output Network Appliance) FIONAs are rack-mounted PCs designed by our PRP team. They are Science DMZ Data Transfer Nodes (DTNs), optimized for 10-100Gbps data transfers and large data storage, and are also capable of holding up to eight GPUs or FPGA add-in boards or up to 256TB of disk each.

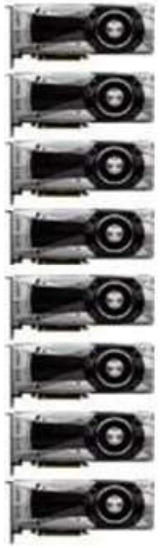
Software Specifications

Installation Instructions

Hardware Specifications

2021 – 2020 – 2019 – 2018 – 2017 – 2016 – 2015

www.pacificresearchplatform.org





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

Up to 192 TB Rotating Storage

PRP PACIFIC RESEARCH PLATFORM

NSF

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

SDSC

20 YEARS QUALCOMM INSTITUTE

PRP's Nautilus is a Multi-Institution Hypercluster Connected by Optical Networks

Nautilus

180 FIONAs on 25 Partner Campuses Networked Together at 10-100Gbps

Installed CPUs	Installed GPUs	Rotating Storage
7478	516	4000 TB

PRP PACIFIC RESEARCH PLATFORM

CENIC

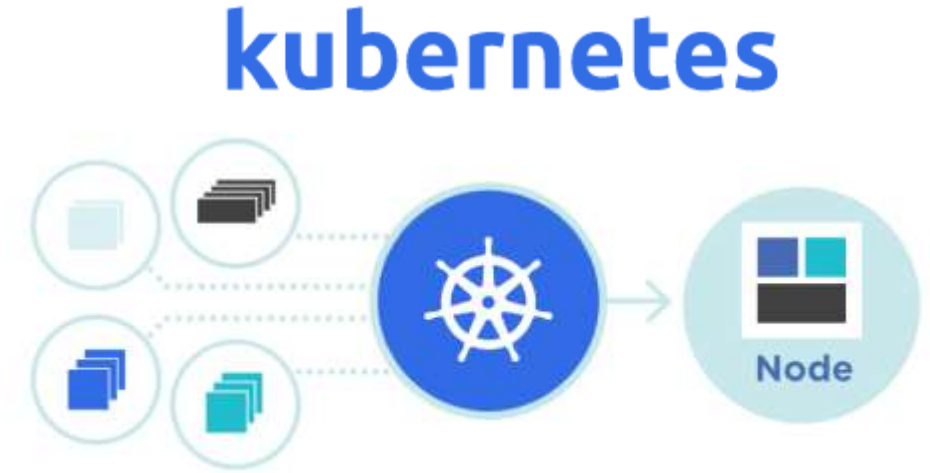
THE QUILT

INTERNET2

20 YEARS QUALCOMM INSTITUTE

2018/2019: PRP Game Changer!
Using Google's Kubernetes to Orchestrate Containers Across the PRP

User Applications



Containers

PRP PACIFIC RESEARCH PLATFORM

Clouds

PRP PACIFIC RESEARCH PLATFORM

CLOUD NATIVE COMPUTING FOUNDATION

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

Nautilus

20 YEARS QUALCOMM INSTITUTE

PRP Provides Widely-Used Kubernetes Services For Application Research, Development and Collaboration





PRP
PACIFIC RESEARCH PLATFORM

SDSC

Engaging More Scientists:
Newly Designed and Updated PRP Website

http://pacificresearchplatform.org



The PRP Web Site Has Detailed Information On How to Join PRP's Nautilus

START | TUTORIAL | RUNNING | STORAGE | DEVELOPMENT | BACK TO NAUTILUS | BACK TO PRP

Get access
Quick Start
Policies
Available Resources
FAQ
Contact

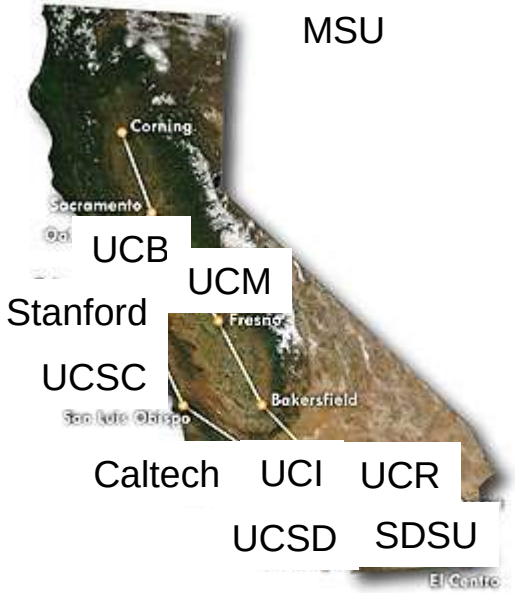
Where to start

www.pacificresearchplatform.org

- **Start** - how to get access to the cluster and intros
  - Get access: Getting authenticated in the cluster
  - Quick start: Basic info, links to external tutorials
  - Policies: **VERY IMPORTANT** section! Please read to avoid issues running in the cluster
  - Available resources: Additional resources you can get in the cluster
- **Tutorials** - step-by-step tutorials for working with the cluster
  - Docker: Using docker locally, finding and building containers
  - Basic kubernetes: Exrloding the system, running simple pods
  - Scaling and exposing: Scaling, load balancing, exposing services to outside
  - Scheduling: Scheduling and requesting resources
  - Batch jobs: Running batch jobs
  - Images: How to choose the right container image or build your own
  - Storage: Using persistent storage in the cluster
- **Running**: running various kinds of jobs in the cluster efficiently
  - Jupyter&Tensorflow2 example: A comprehensive example on how to run the image we provide, containing most popular python libraries for ML
  - Running batch jobs
  - Dealing with high I/O

PRP
PACIFIC RESEARCH PLATFORM

2017-2020: CHASE-CI Grant Adds a Machine Learning Layer Built on Top of the Pacific Research Platform for CISE Researchers



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

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

NSF Just Funded Two Extensions:
CHASE-CI ABR and ENS

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

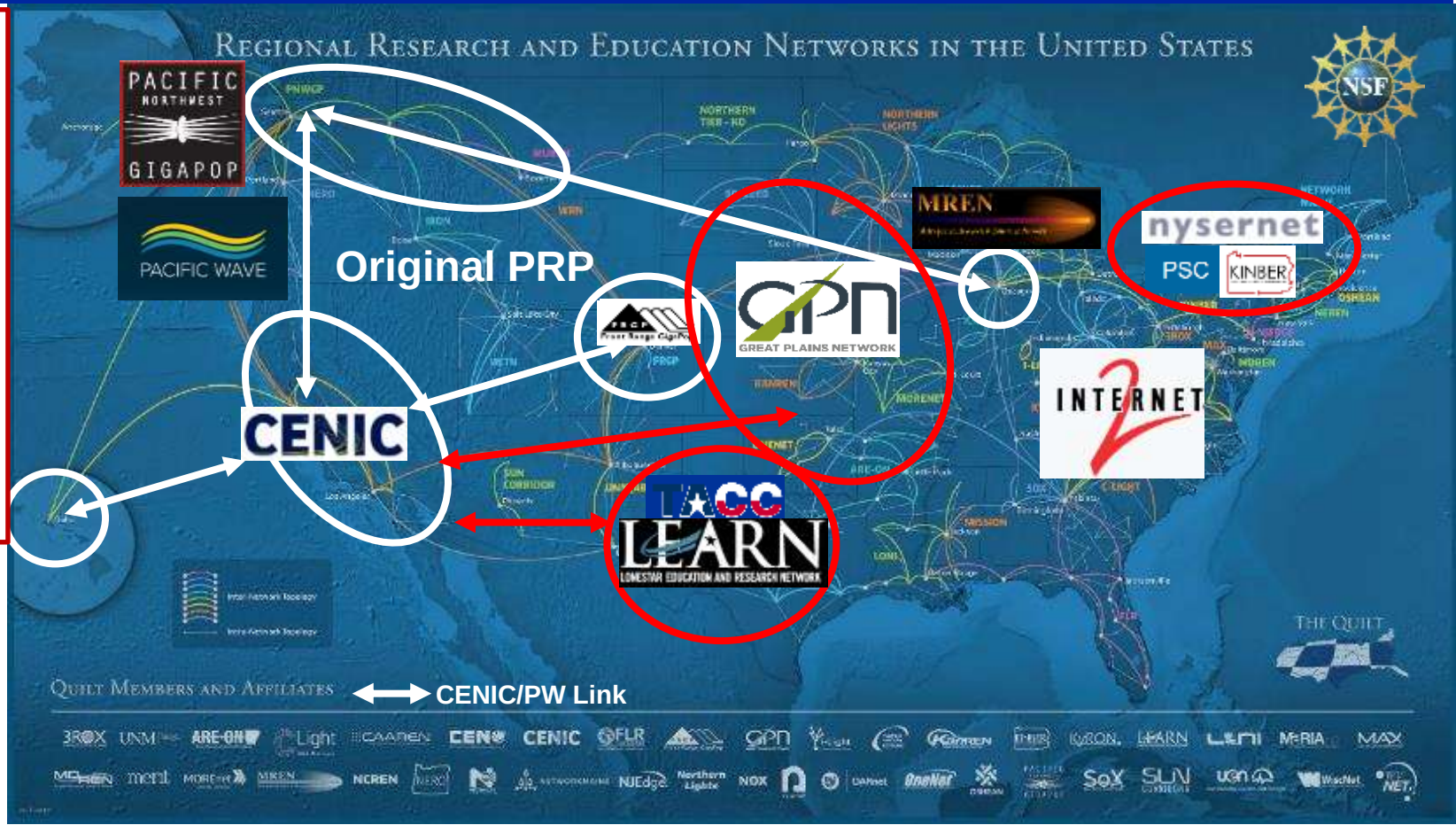
PRP PACIFIC RESEARCH PLATFORM

SDSC UC San Diego it²

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

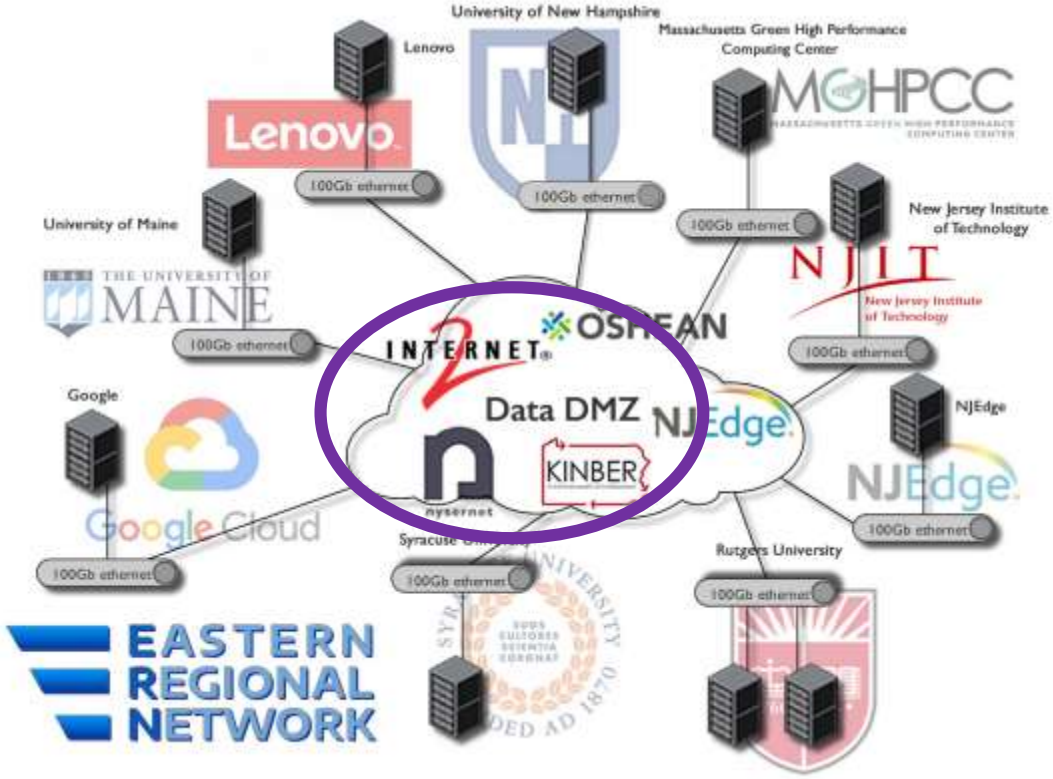
"Towards The NRP"
3-Year Grant Funded by NSF
$2.5M
October 2018
Award #1826967

PI Smarr
Co-PIs Altintas
Papadopoulos
Wuerthwein
Rosing
DeFanti



PACIFIC RESEARCH PLATFORM

Next Step?
Federate ERN with TNRP/PRP/CHASE-CI



PRP
PACIFIC RESEARCH PLATFORM

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



CMS

Particle Physics

Biomedical ‘omics

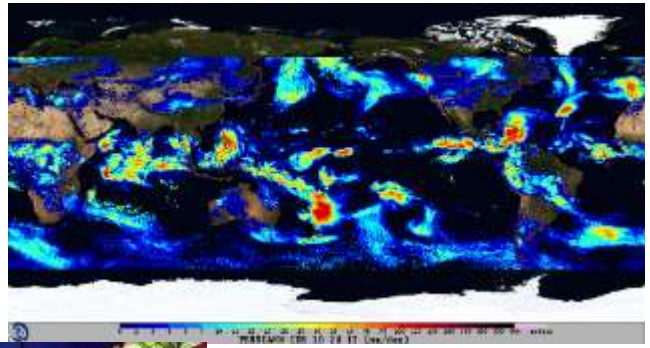


NSF

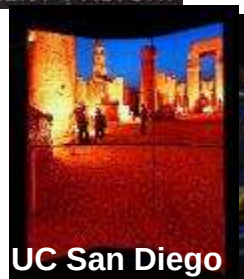


Telescope Surveys

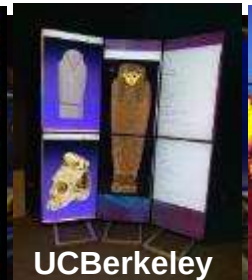
Earth Sciences



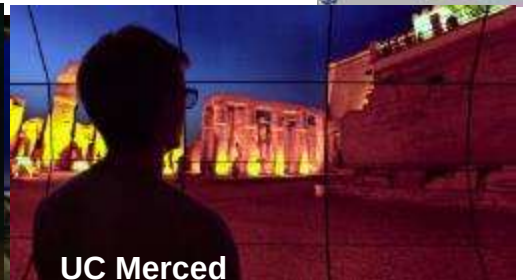
Visualization, Virtual Reality, Collaboration



UC San Diego



UCBerkeley



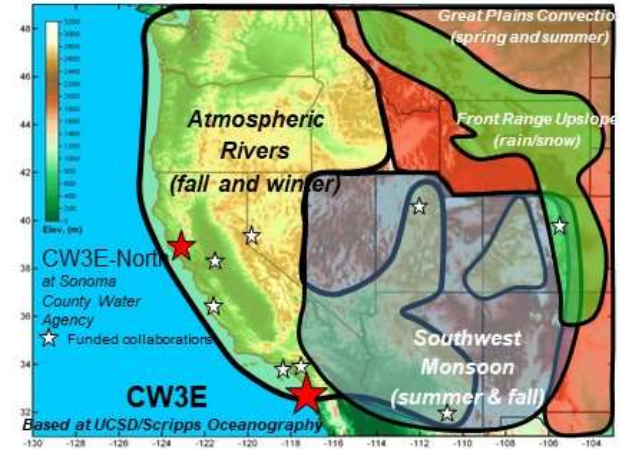
UC Merced

CITRIS THE BAN INST

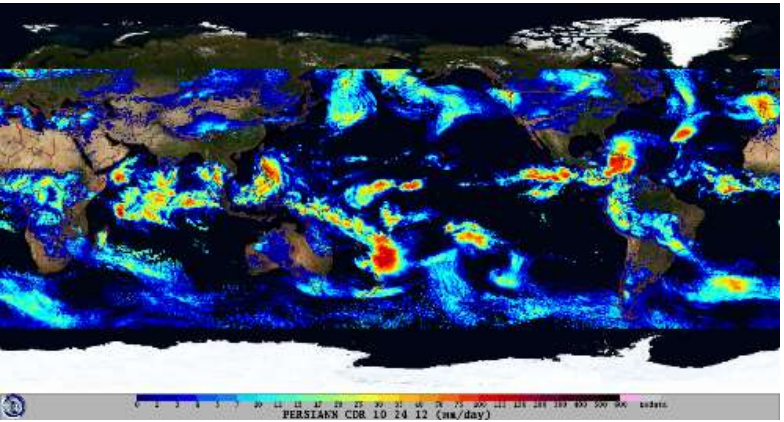
SDSC

it²

PRP Accelerates 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:

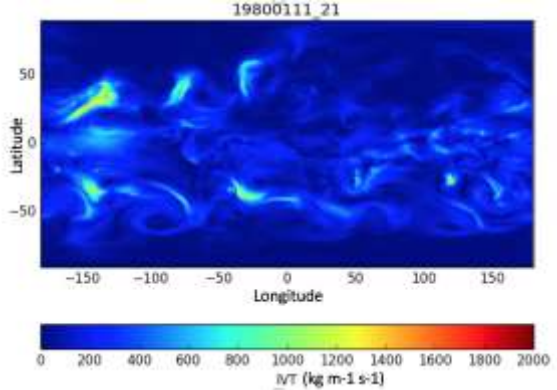
Center for Hydrometeorology & Remote Sensing
University of California, Irvine

Director, Soroosh Sorooshian, UCSD

Source: Scott Sellers, PhD CHRS; Postdoc CW3E

Scott Sellars Rapid 4D Object Segmentation of NASA Water Vapor Data - Machine Learning in Time and Space

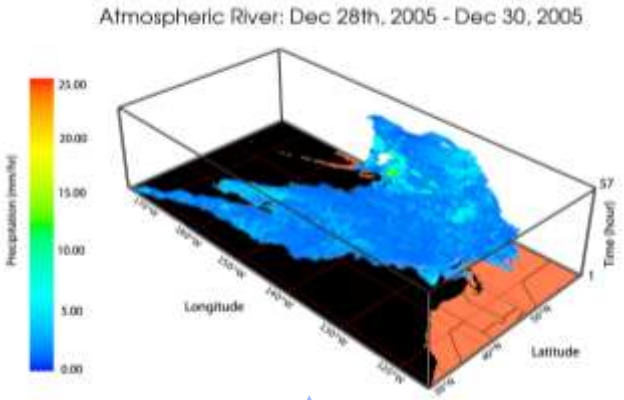
NASA *MERRA v2 – Water Vapor Data Across the Globe



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 Enabled Scott's Workflow to Run 532 Times Faster!

See Sellars, eScience 2019
https://ieeexplore.ieee.org/document/9041726

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

The New Pacific Research Platform Video Highlights 3 Different Applications



Pacific Research Platform Video:
www.thequilt.net/campus-cyberinfrastructure-program-resource/
www.pacificresearchplatform.org

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

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

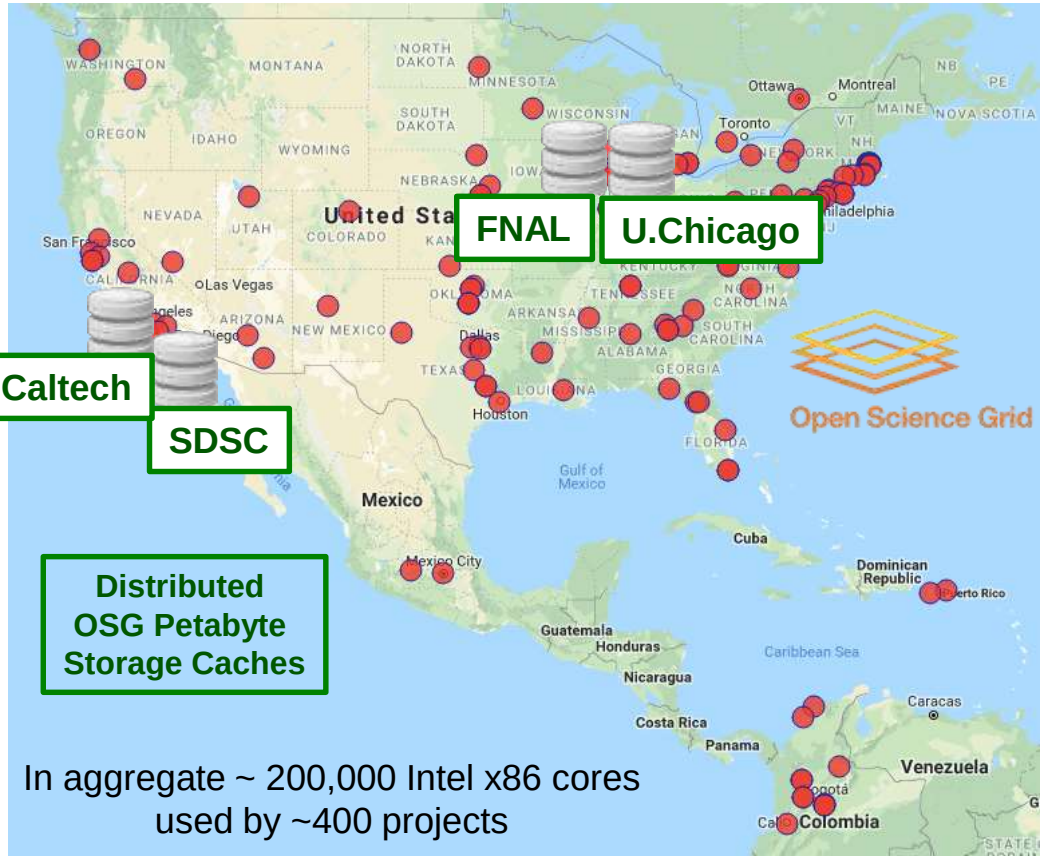
All OSG User Communities
Use HTCondor for Resource Orchestration

HTCondor
High Throughput Computing

NSF

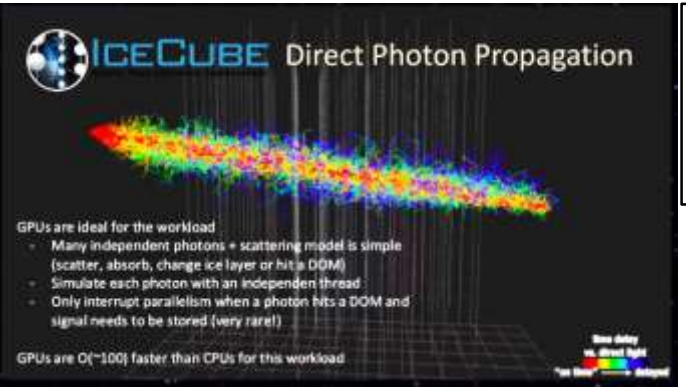
PRP
PACIFIC RESEARCH PLATFORM

OSG Federates ~100 Clusters Worldwide



NSF Large-Scale Observatories Are Using PRP and OSG as a Cohesive, Federated, National-Scale Research Data Infrastructure





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

WISCONSIN
WIPAC
Wisconsin IceCube Particle Astrophysics Center
NSF

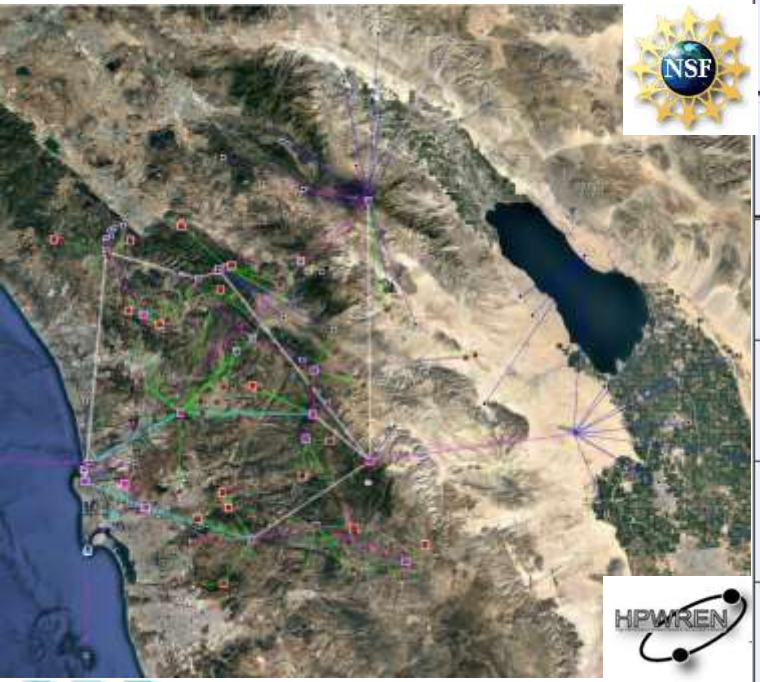
PRP
PACIFIC RESEARCH PLATFORM

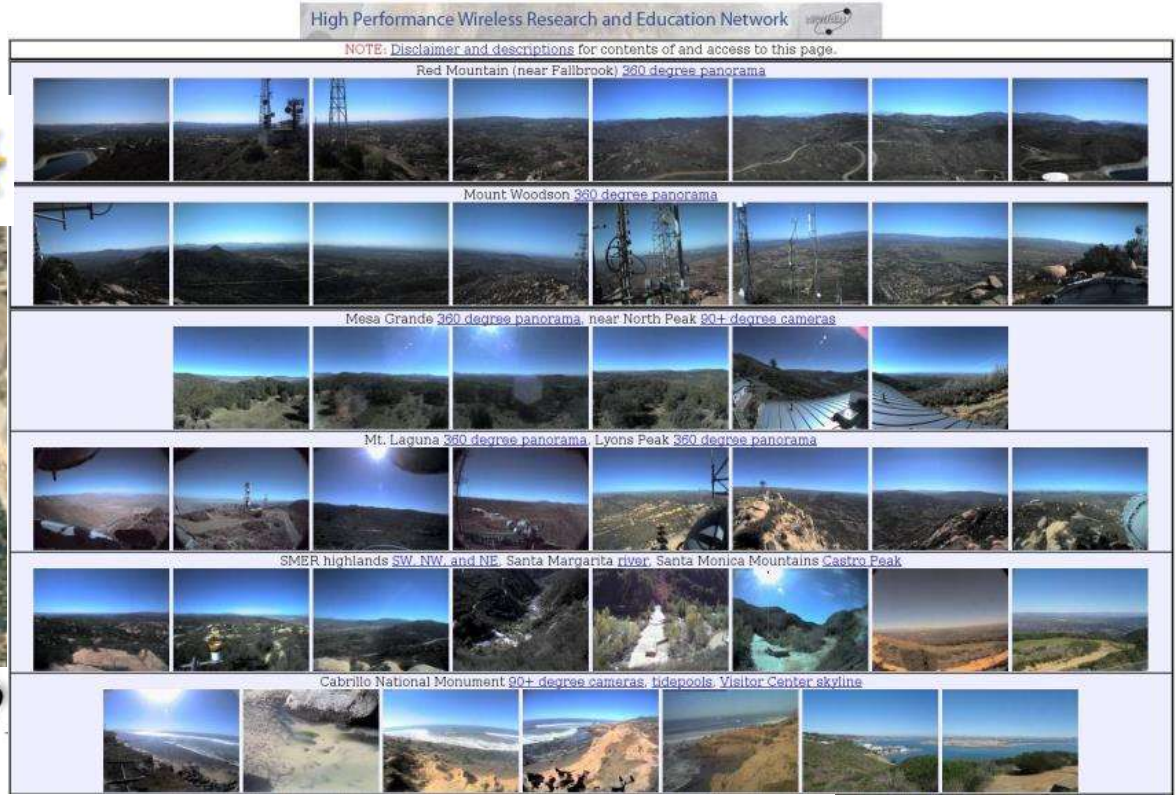
NSF's IceCube & LIGO Both See Nautilus as Just Another OSG Resource

Open Science Grid

2017-2019: HPWREN: 15 Years of NSF-Funded Real-Time Network Cameras and Meteorological Sensors on Top of San Diego Mountains for Environmental Observations

Source: Hans Werner Braun, HPWREN PI





NSF

HPWREN

PRP PACIFIC RESEARCH PLATFORM

UC San Diego

SDSC SAN DIEGO SUPERCOMPUTER CENTER

SCRIPPS INSTITUTION OF OCEANOGRAPHY

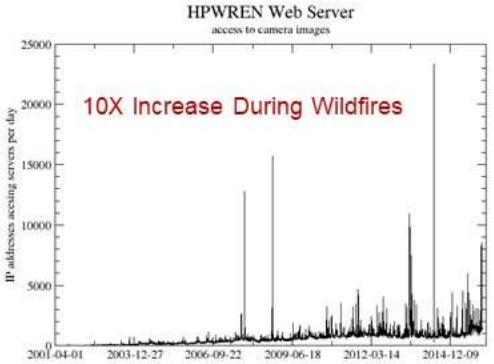
SAN DIEGO STATE UNIVERSITY

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



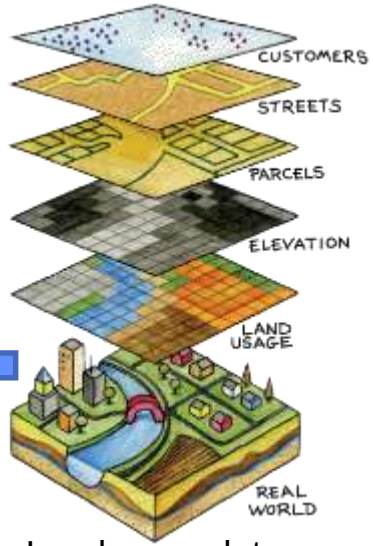
Real-Time Meteorological Sensors



Source: Ilkay Altintas, SDSC

WIFIRE

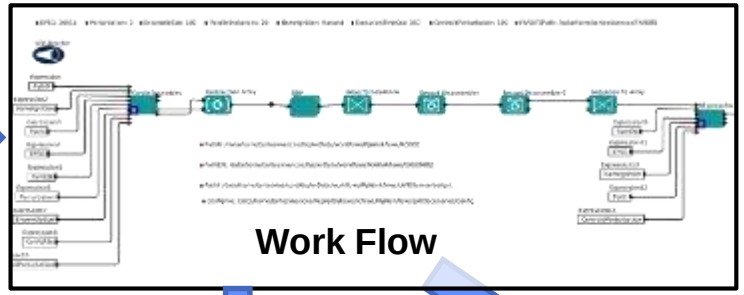




Landscape data

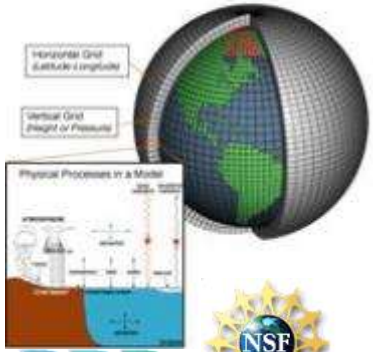
PRP





Work Flow

Weather Forecast







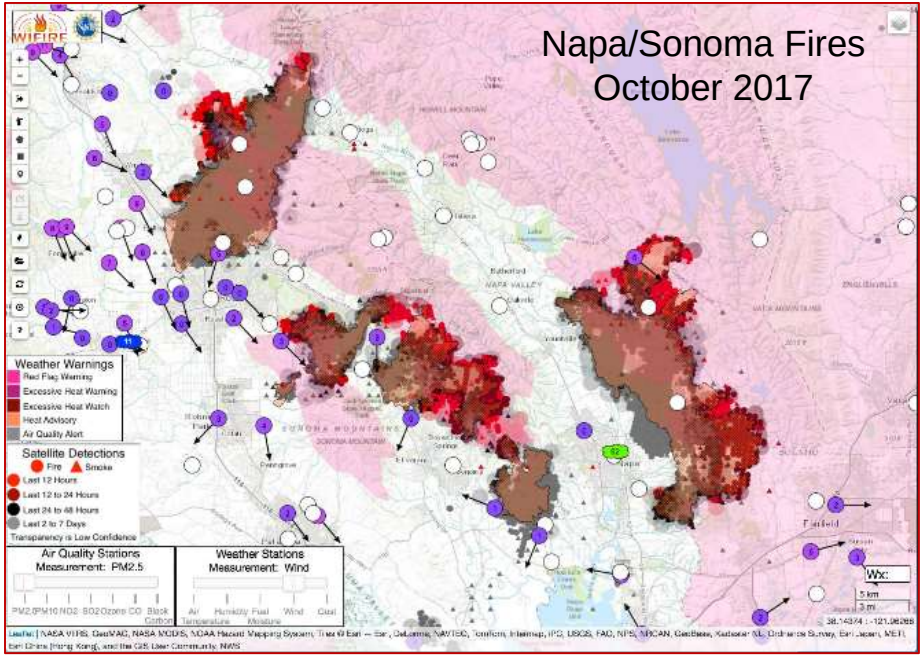


NSF

PRP
PACIFIC RESEARCH PLATFORM

WIFIRE's Firemap Was Heavily Used by Public For California Wildfires October 2017 through December 2017

800K+ unique visitors and 8M+ hits





PACIFIC RESEARCH PLATFORM

http://firemap.sdsc.edu

WIFIRE

NSF

NeuroKube: An Automated Neuroscience Reconstruction Framework Uses Nautilus for Large-Scale Processing & Labeling of Neuroimage Volumes

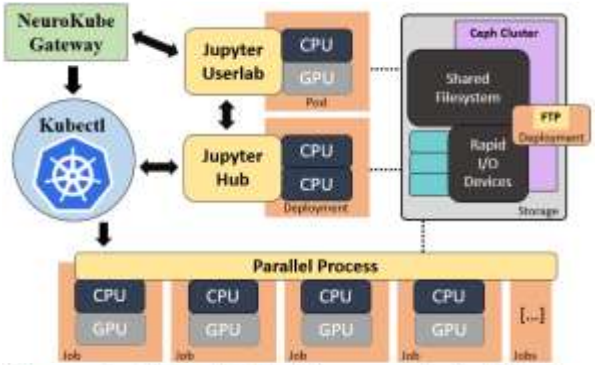


Figure 2. NeuroKube Framework Architecture

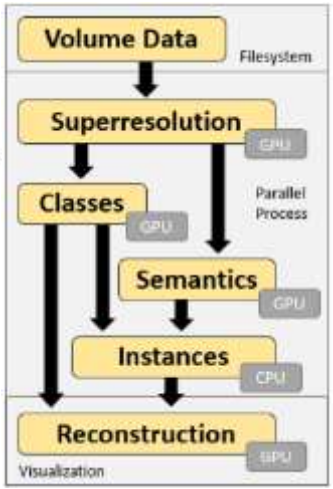


Figure 4. Workflow Steps

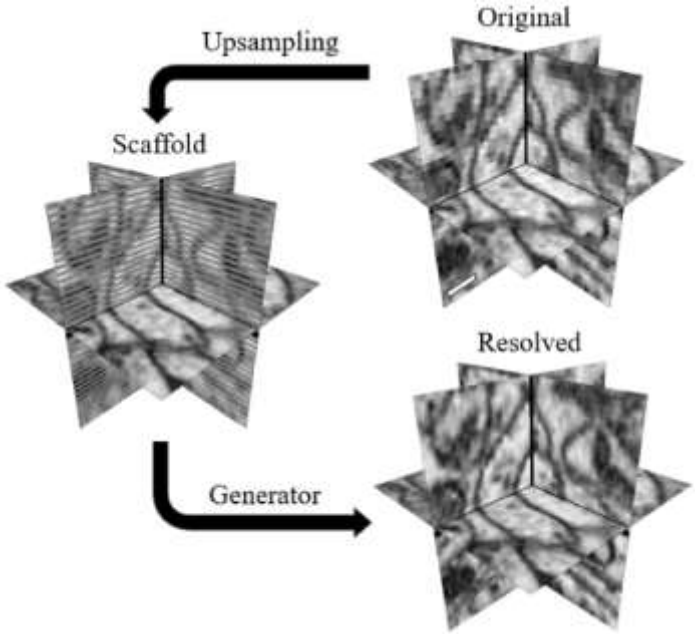
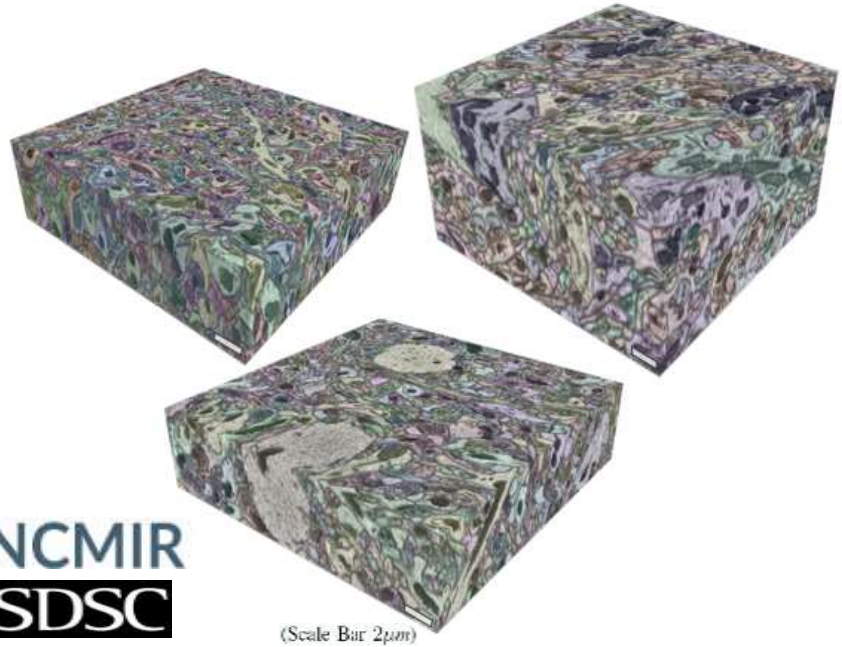


Figure 5. VEM Superresolution (Scale Bar 250nm)

Figures 2, 4, & 5 in “NeuroKube:
An Automated and Autoscaling Neuroimaging Reconstruction Framework
Using Cloud Native Computing and A.I.,”
Matthew Madany, et al. (accepted to IEEE Big Data ’20)

Computer Vision-Based Approach Provides the Potential to Automatically Generate Labels Using ML

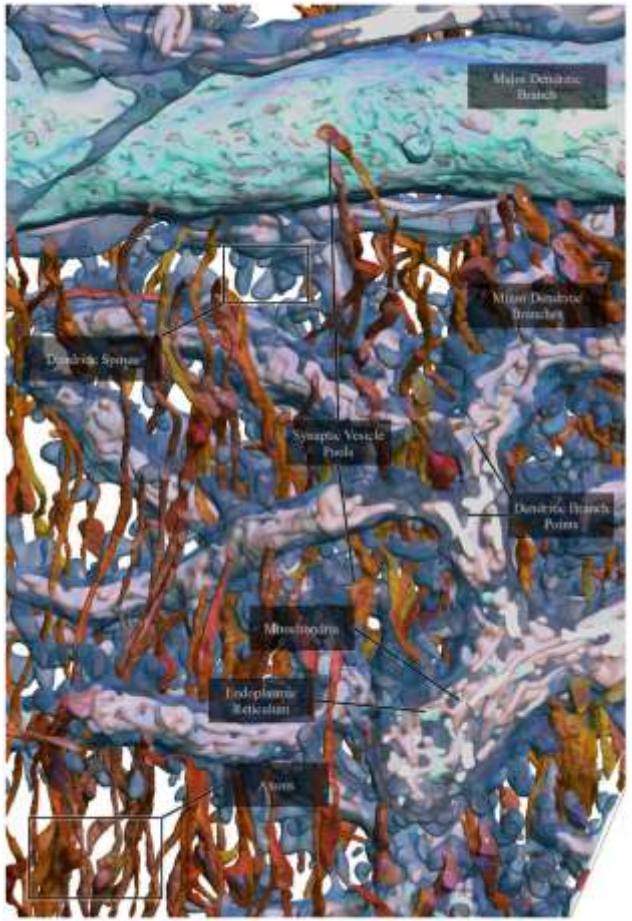


NCMIR

SDSC

PRP
PACIFIC RESEARCH PLATFORM

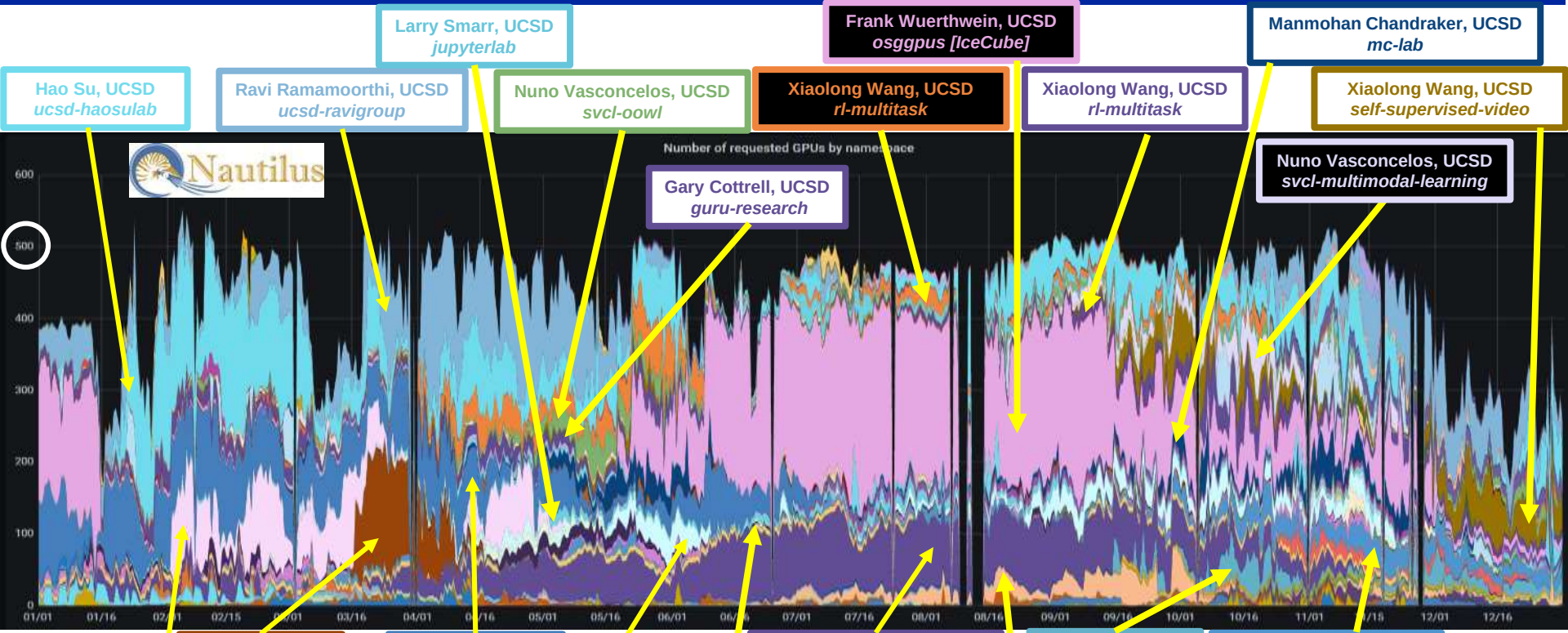
Volumetric Electron Microscopy (VEM) Data with Colorized Labels



Subset of Neurites from Cerebellum Neuropil Extracted & Rendered in 3D with Structures of Interest Labeled

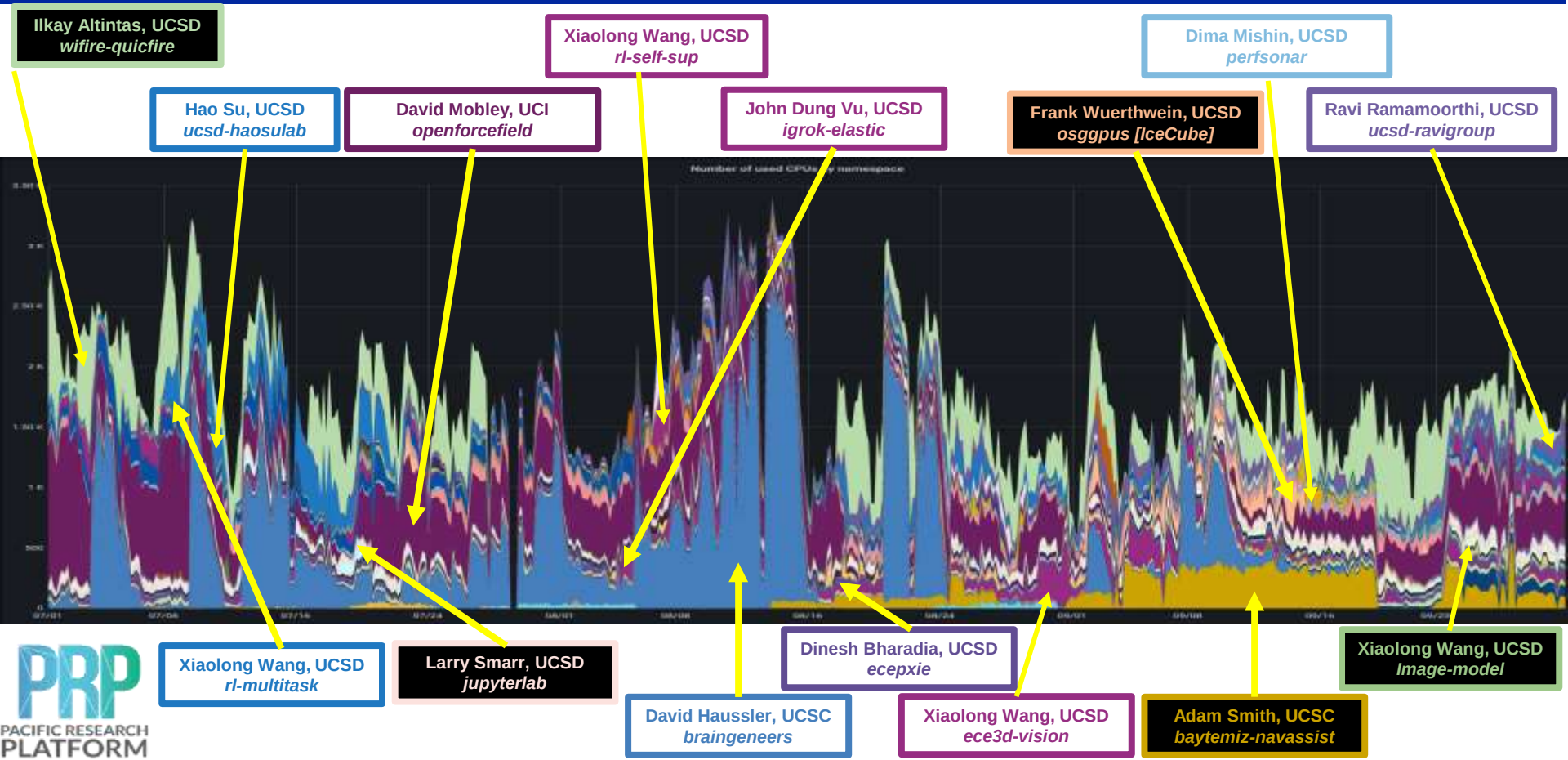
Figures 1 & 14 in “NeuroKube: An Automated and Autoscaling Neuroimaging Reconstruction Framework using Cloud Native Computing and A.I.,” Matthew Madany, et al. (accepted to IEEE Big Data ’20)

Top 20 GPU Users Out of 400 Nautilus Namespace Applications: Together They Consumed Nearly 500 GPUs in 2020



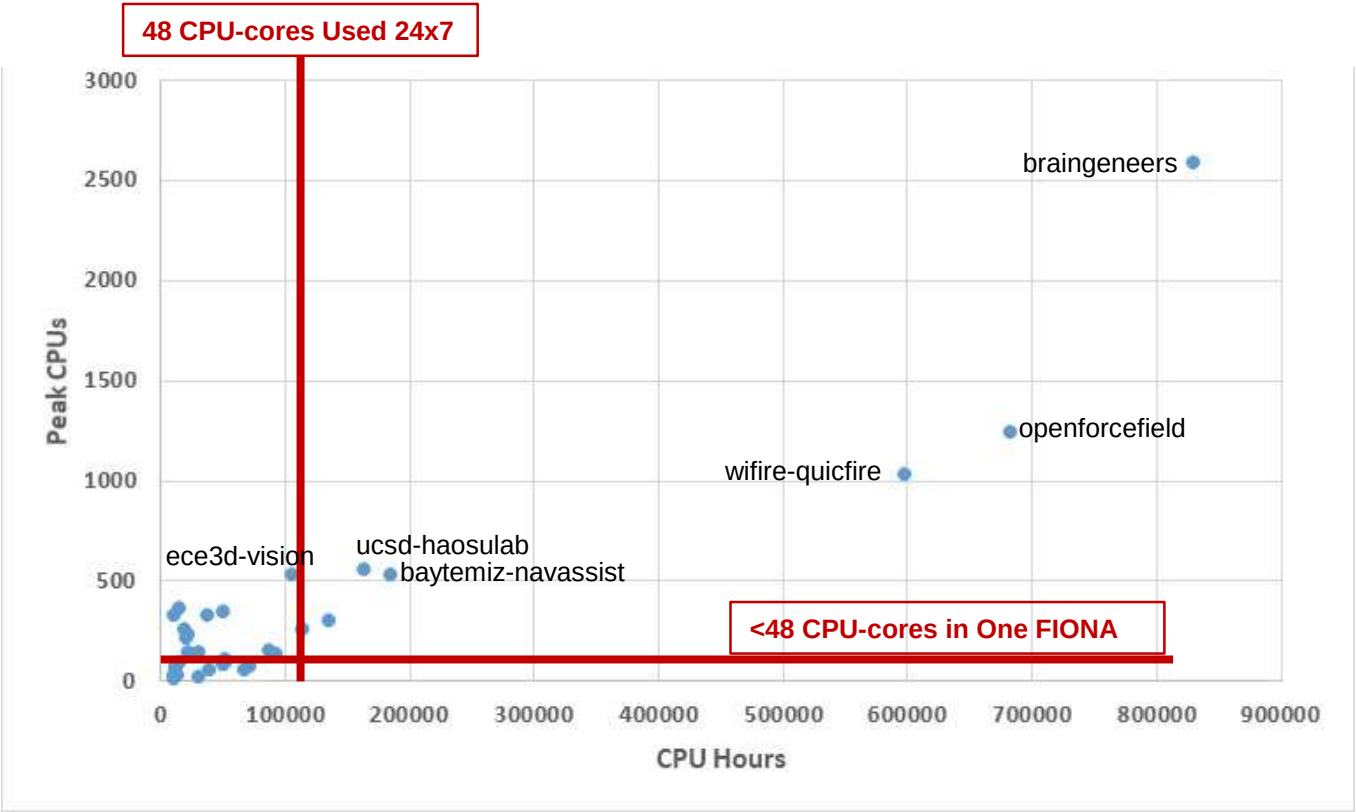
PRP
PACIFIC RESEARCH PLATFORM

PRP Y6Q4
Top 15 CPU Nautilus Namespace Users (>50,000 CPU Core Hours)



PACIFIC RESEARCH PLATFORM

Peak vs. Total CPU Nautilus Namespace Usage
Y6Q4



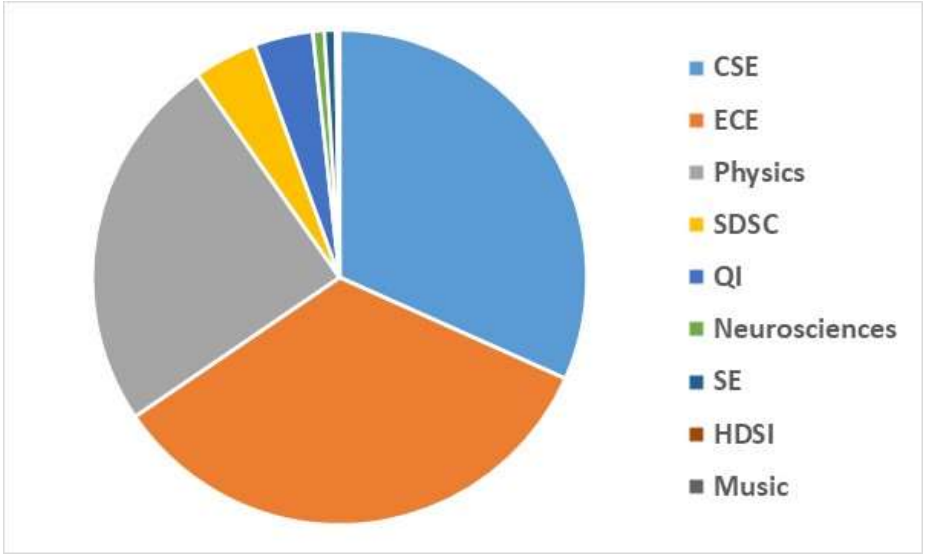
PACIFIC RESEARCH PLATFORM

ML/AI Namespace examples

PACIFIC RESEARCH PLATFORM

PRP's Nautilus GPUs
Supports a Broad Set of Science and Machine Learning Applications

3,110,765 GPU-Hours



Total Usage is Equivalent to Running
355 GPUs 24/7 for 12 Months

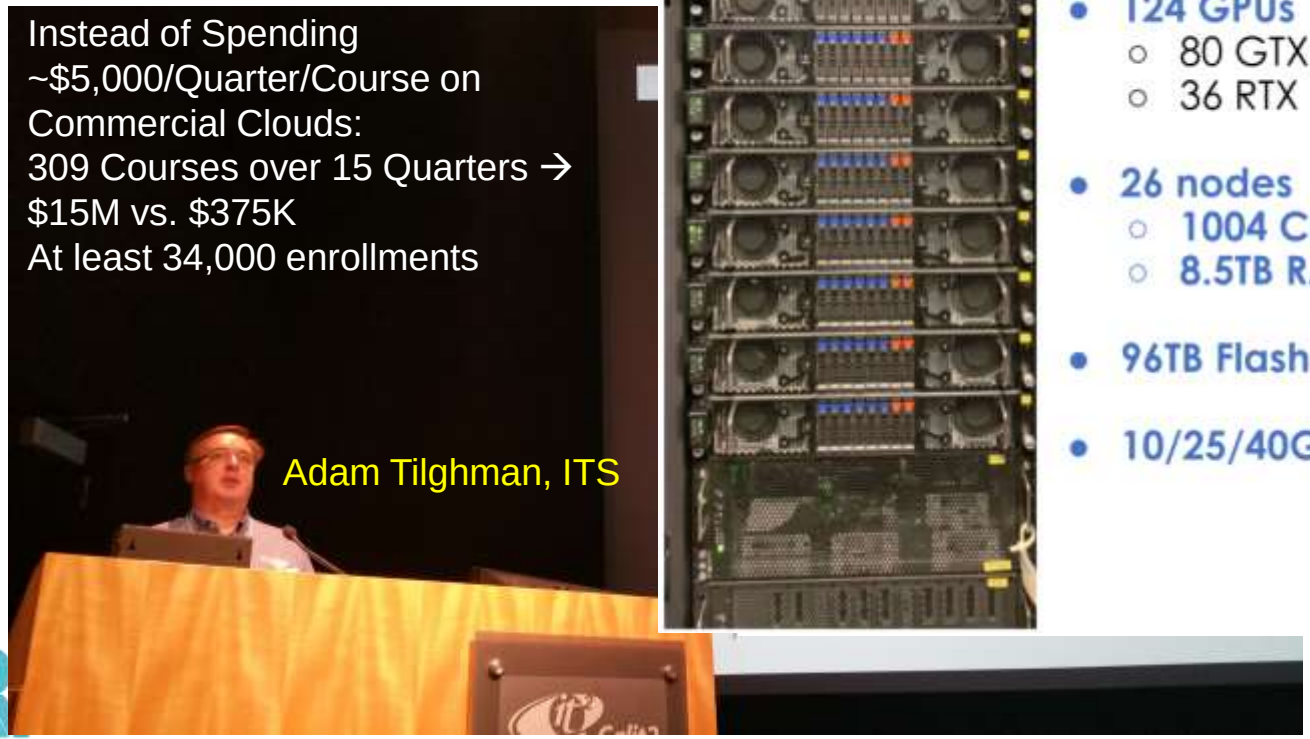
UC San Diego by Department in 2020

- Physics Usage is *Community* Data Analysis of NSF Major Facilities:
  - Large Hadron Collider
  - IceCube South Pole Neutrino Detector
  - LIGO Gravitational Wave Observatory
- SDSC and Qualcomm Institute Usage is *Community* Software Support
- CSE, ECE, SE, Neurosciences, & Music Department Usage - Individual Machine Learning Faculty Research Projects

PRP
PACIFIC RESEARCH PLATFORM

UCSD's Information Technology Services Adapted PRP FIONA8s To Support Data Science Courses

Instead of Spending ~$5,000/Quarter/Course on Commercial Clouds:
309 Courses over 15 Quarters → $15M vs. $375K
At least 34,000 enrollments



- 124 GPUs
  - 80 GTX 1080Ti
  - 36 RTX 2080Ti & RTX Titan
- 26 nodes
  - 1004 CPU cores
  - 8.5TB RAM
- 96TB Flash Storage
- 10/25/40G Networking

PACIFIC RESEARCH PLATFORM

UC San Diego DSMLP
Data Science / Machine Learning Platform

- **Student-focused GPU/CPU cluster for:**
  - **Undergraduate & Graduate Coursework**
  - **For-Credit Independent Study**
  - **Thesis/Dissertation Research**
  - **Capstones & Projects**
- **Research-Driven Architecture**
- **Managed by Central IT Services**



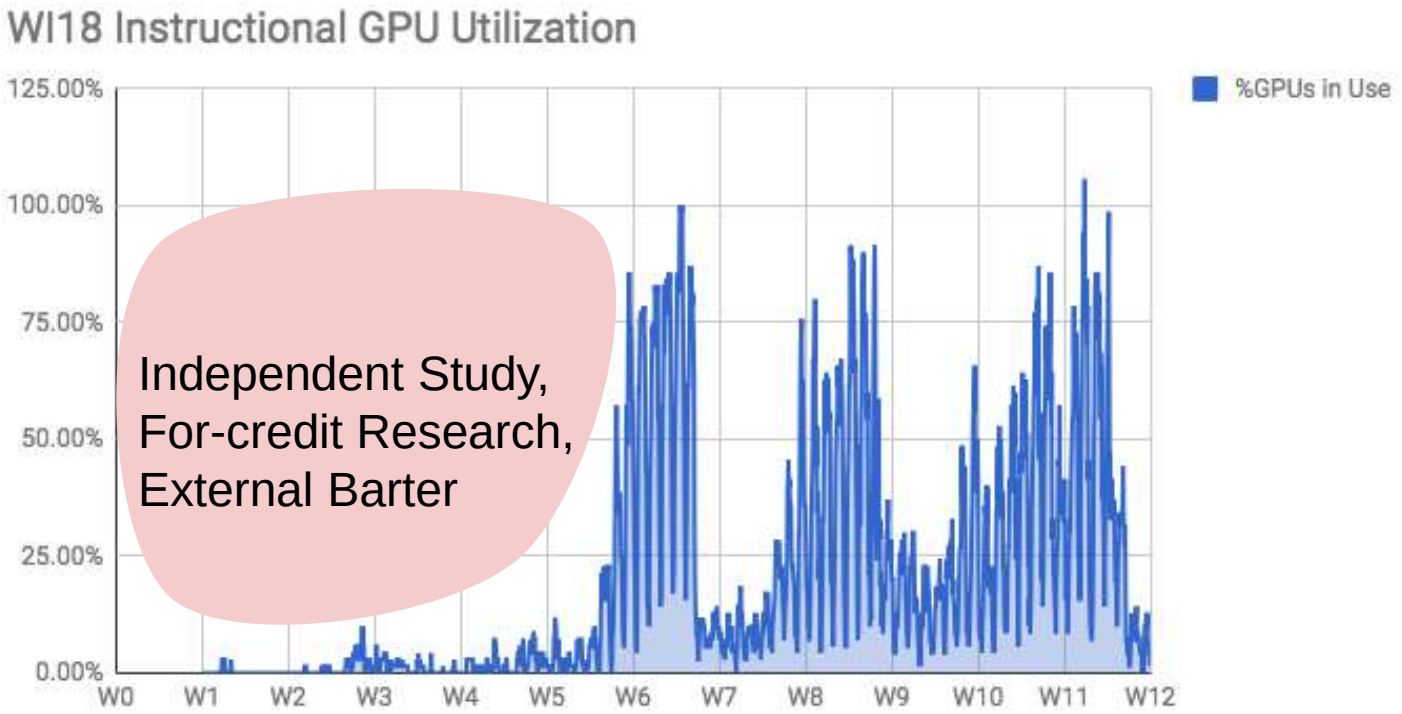
PACIFIC RESEARCH PLATFORM

IT SERVICES

UC San Diego
Jacobs School of Engineering

UCSD Cognitive Science

Coursework Activity Patterns



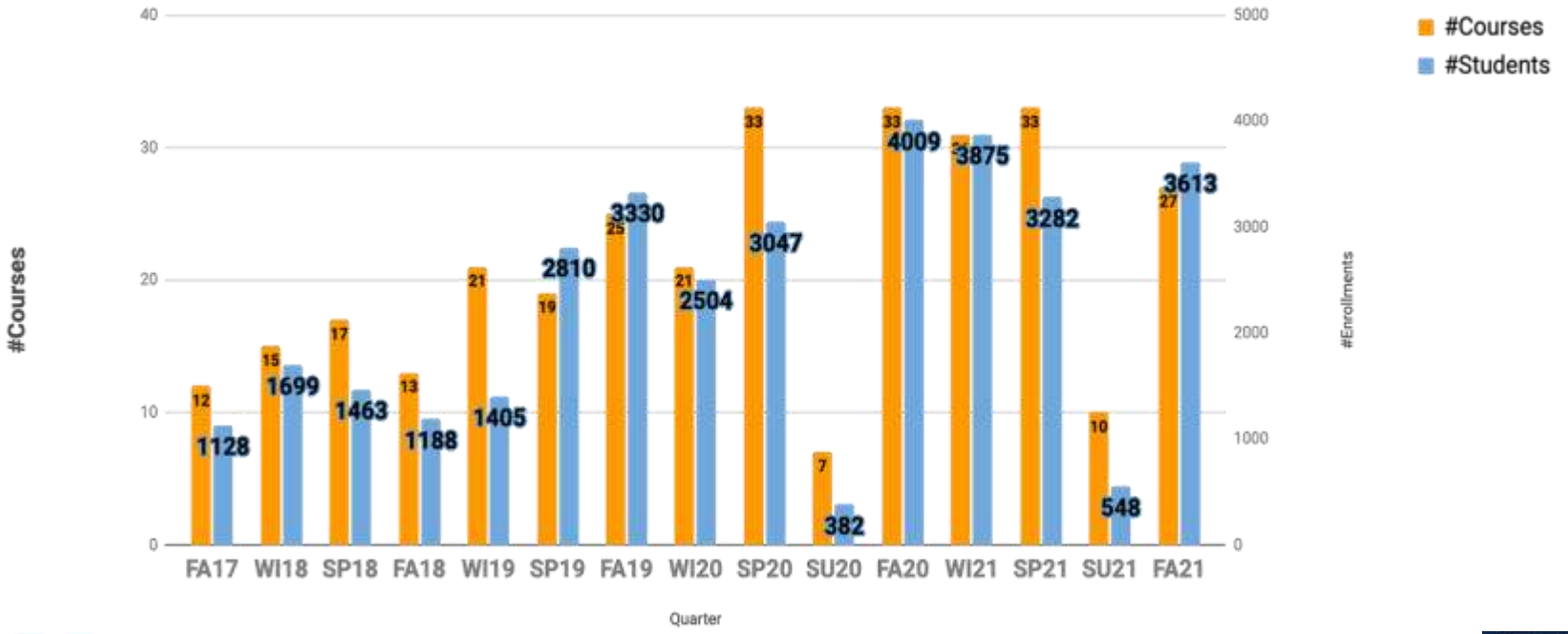
PACIFIC RESEARCH PLATFORM

DSMLP Courses by Division, Term



PACIFIC RESEARCH PLATFORM

DSMLP Courses, Enrollments by Term



PRP PACIFIC RESEARCH PLATFORM

IT SERVICES UC San Diego

Community Building
Through Large-Scale Workshops









PRP
PACIFIC RESEARCH PLATFORM

2nd Global Research Platform (2GRP) Workshop
September 20-24, 2021

Community Building Though Inclusion and Diversity: Workshops With Minority Serving Universities

2NRP Montana State University







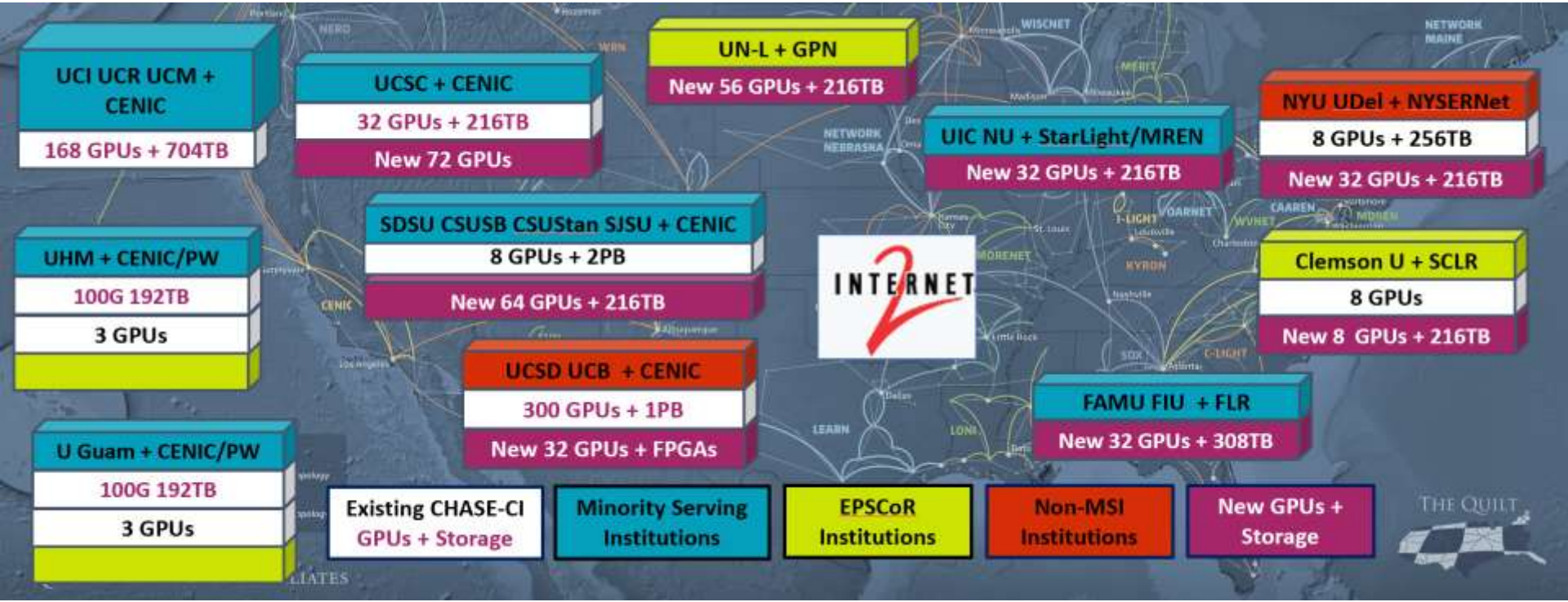
PRP PACIFIC RESEARCH PLATFORM

CITRIS BANATAO INSTITUTE

The Next Three Phases
As We Approach a National Research Platform

PACIFIC RESEARCH PLATFORM

2021-2024 NRP Future I: *Proposed* Extension of Nautilus



PACIFIC RESEARCH PLATFORM

CHASE-CI ENS, Tom DeFanti PI (NSF Award # 2120019)
CHASE-CI ABR, Larry Smarr PI (NSF Award # 2100237)
$2.8M

20 YEARS QUALCOMM INSTITUTE

2021-2026 NRP Future II: PRP Federates with SDSC's EXPANSE Using CHASE-CI Developed Composable Systems

EXPANSE

COMPUTING WITHOUT BOUNDARIES
5 PETAFLOP/S HPC and DATA RESOURCE

SDSC

XSEDE

HPC RESOURCE
13 Scalable Compute Units
728 Standard Compute Nodes
52 GPU Nodes: 208 GPUs
4 Large Memory Nodes

DATA CENTRIC ARCHITECTURE
12PB Perf. Storage: 140GB/s, 200k IOPS
Fast I/O Node-Local NVMe Storage
7PB Ceph Object Storage
High-Performance R&E Networking



~$20M over 5 Years
PI Mike Norman, SDSC

REMOTE CI INTEGRATION

PRP PACIFIC RESEARCH PLATFORM

CLOUD

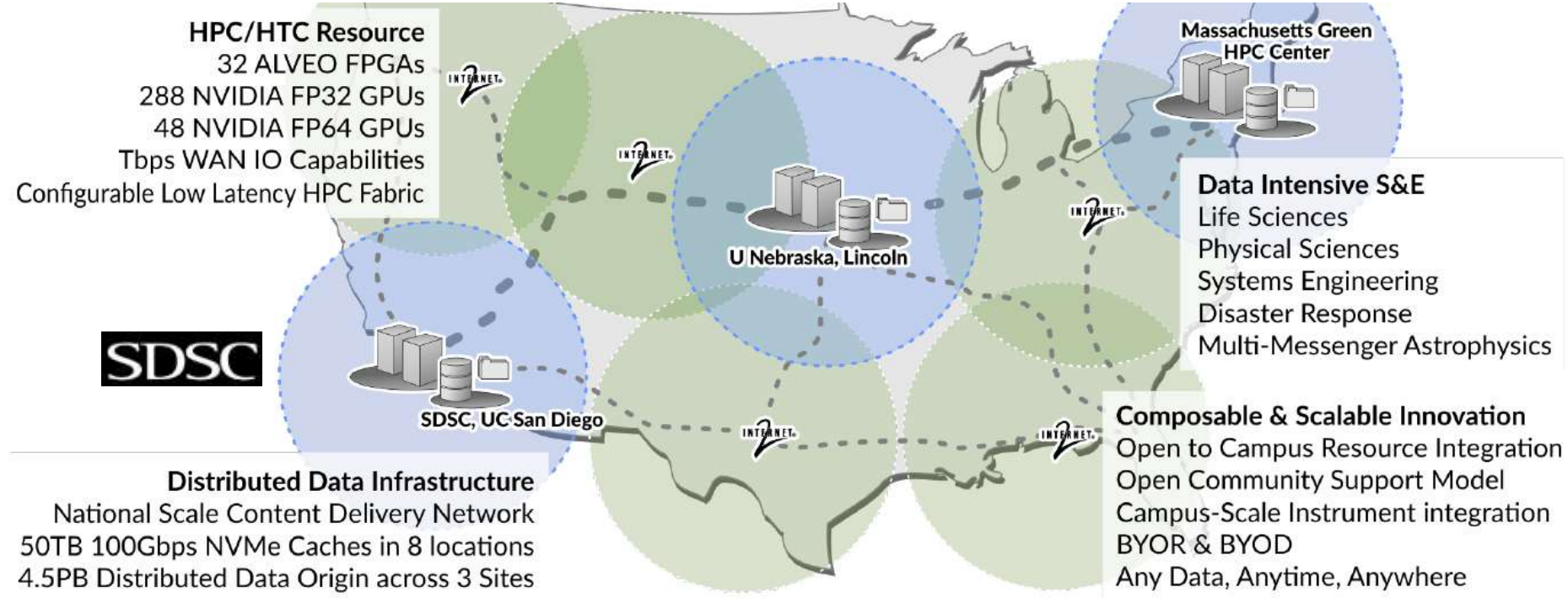
Open Science Grid

Heterogeneous Resources

LONG-TAIL SCIENCE
Multi-Messenger Astronomy
Genomics
Earth Science
Social Science

INNOVATIVE OPERATIONS
Composable Systems
High-Throughput Computing
Science Gateways
Interactive Computing
Containerized Computing
Cloud Bursting

2021-2026 NRP Future III: PRP Federates with NSF-Funded *Prototype National Research Platform*



PRP PACIFIC RESEARCH PLATFORM

NSF Award OAC #2112167 (June 2021) [$5M Over 5 Years]
PI Frank Wuerthwein (UCSD, SDSC)
Co-PIs Tajana Rosing (UCSD), Thomas DeFanti (UCSD), Mahidhar Tatineni (SDSC), Derek Weitzel (UNL)

PRP's Support and Community:



- National Science Foundation (NSF) awards to UCSD:
  - CNS (1456638, 1730158, 2120019, 2100237) OAC (1540112, 1541349, 1826967, 2112167)
- UCSD; Calit2/Qualcomm Institute; and UCSD's Research IT and Instructional IT
- UCB CITRIS and the Banatao Institute
- UC Office of the President
- Partner Campuses: UCSC, UCI, UCR, UCLA, UCD, UCM, UCSB, USC, Caltech, Stanford, NU, UW, UChicago, UIC, UIUC, UHM, UWM, IU, NPS, CSUSB, CSUS, SDSU, SJSU, UMC, UM, MSU, NYU, UNL, UNM, UNC, UTA, WSU, FAMU, FIU, Clemson, UD, UG, UU, JCU, KISTI, UVA, AIST, NTU, UQ, UTokyo
- Computing Partners: San Diego Supercomputer Center, LBNL/NERSC, NCAR/UCAR & Wyoming Supercomputing Center, NASA NAS/USRA, Texas Advanced Computing Center, NSCC, Open Science Grid, Chameleon Cloud, SLATE, AWS, Google Cloud, Microsoft
- Network Partners: CENIC, Pacific Wave/PNWGP, FRGP, StarLight/MREN, HPWREN, The Quilt, Great Plains Network, KINBER, LEARN, NYSERNet, OARnet, FLR, Internet2, DOE Esnet, AMPATH, AARNet, CESnet, KREOnet, PIREN, SURFnet, SCLR, SingAREN

PRP
PACIFIC RESEARCH PLATFORM