“The Pacific Research Platform: Building a Distributed Big-Data Machine-Learning Cyberinfrastructure”

Briefing
Jacobs School of Engineering
University of California San Diego
July 18, 2019

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

UC San Diego

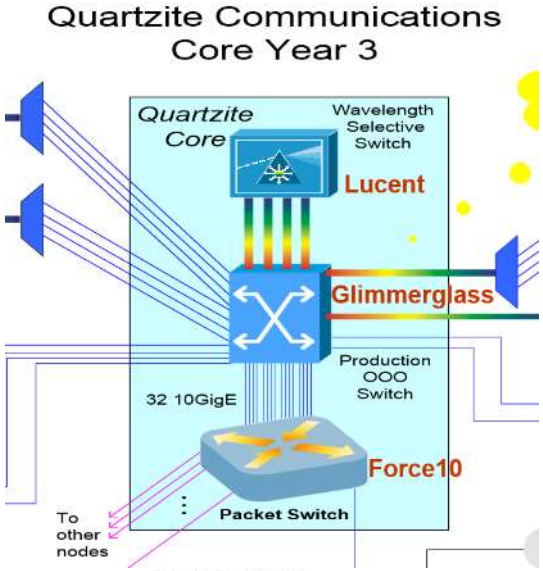
it²

UC San Diego's Calit2 & SDSC Have Pioneered Big-Data Cyberinfrastructure for 17 Years with NSF Grants: OptIPuter, Quartzite, Prism, CHERuB, PRP, CHASE-CI, TNRP



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

OptIPuter
PI Smarr,
Co-PI DeFanti
Co-PI Papadopoulos, Ellisman
2002-2009



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

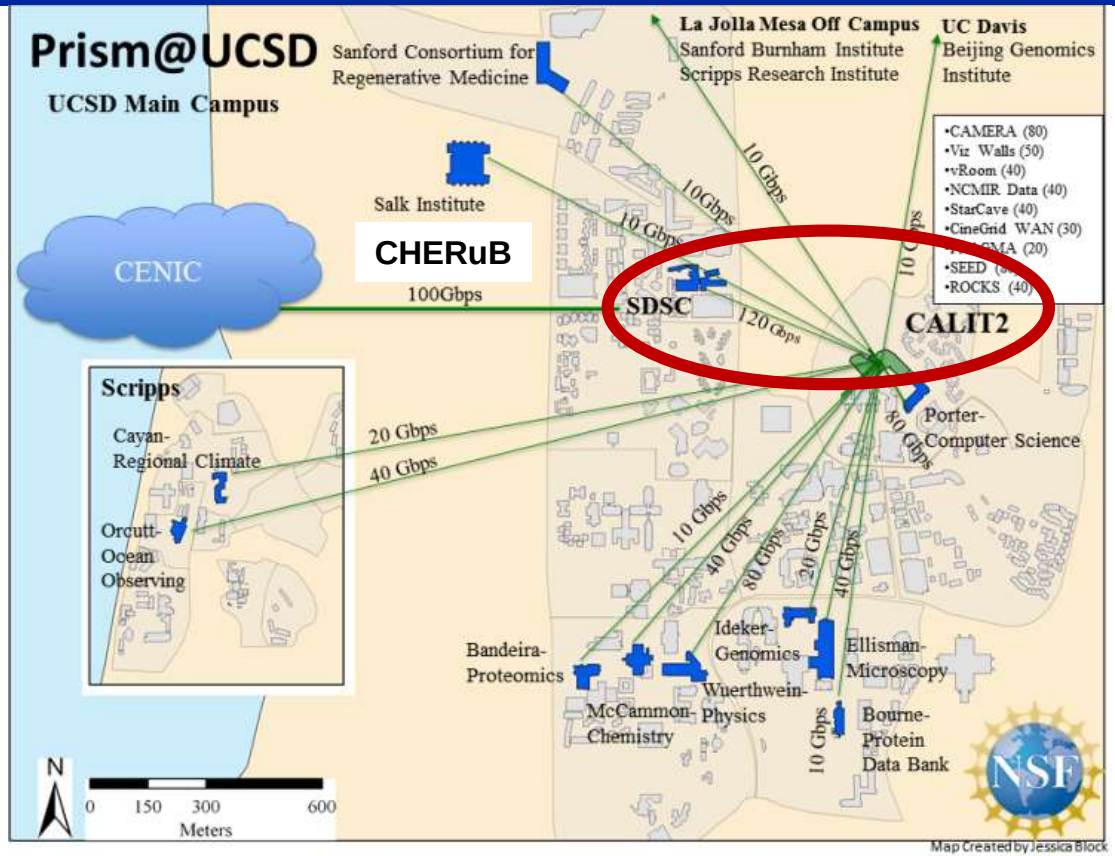
Quartzite
PI Papadopoulos,
Co-PI Smarr, Ford,
Fainman

UC San Diego

SDSC

it²

2013-2015: Creating a "Big Data" Backplane on Campus: NSF CC-NIE Funded Prism@UCSD and CHERuB



Prism@UCSD, Phil Papadopoulos, SDSC, Calit2, PI; Smarr co-PI
CHERuB, Mike Norman, SDSC PI

UC San Diego

SDSC

it²

2015-2020: The Pacific Research Platform Connects Campus "Big Data Freeways" to Create a Regional End-to-End Science-Driven "Big Data Superhighway" System



Source: John Hess, CENIC

NSF CC*DNI Grant
$6M 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**

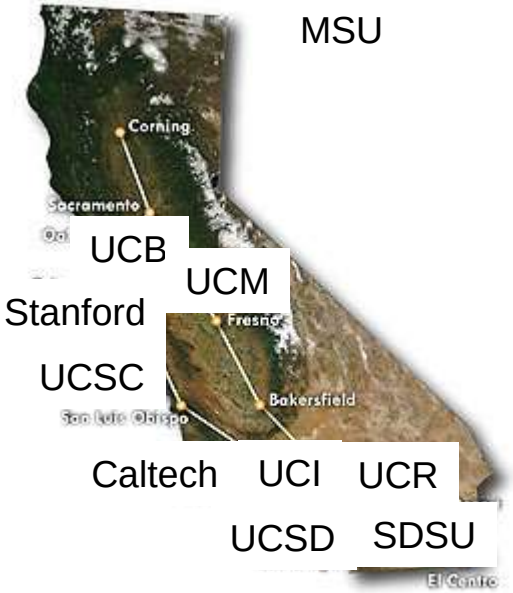
Letters of Commitment from:

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

UCOP CIO Tom Andriola
Provided Funds and ITLC Support
for Using Ten UC Campuses
For Advanced Technology Testing

UC San Diego
SDSC
CITRIS BANATAO INSTITUTE

2017-2020: CHASE-CI Adds Machine-Learning to the Data-Science Community Cyberinfrastructure



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

UC San Diego

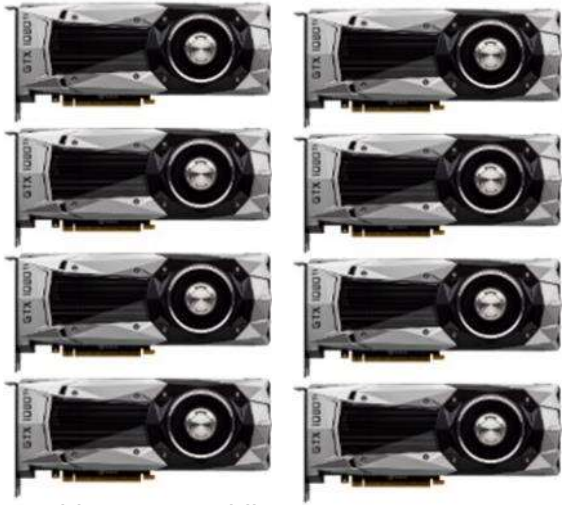
SDSC UC San Diego it²

PRP Engineers Designed and Built Several Generations of Optical-Fiber Big-Data 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 200 TeraByte Rotating Storage



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



FIONette—
1G, $250
Used for
Training 50
Engineers in
2018-2019

Over 100 Now Deployed on PRP

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

UC San Diego

NSF

SDSC

it²

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

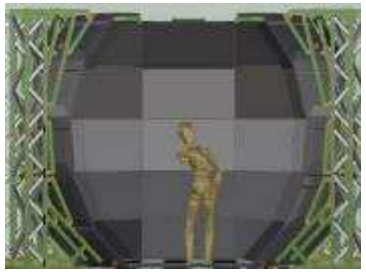
UC San Diego

IT SERVICES

104 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



FIONA with
8-Game GPUs



Plus 288 64-bit GPUs
On SDSC's Comet

Qi QUALCOMM INSTITUTE

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

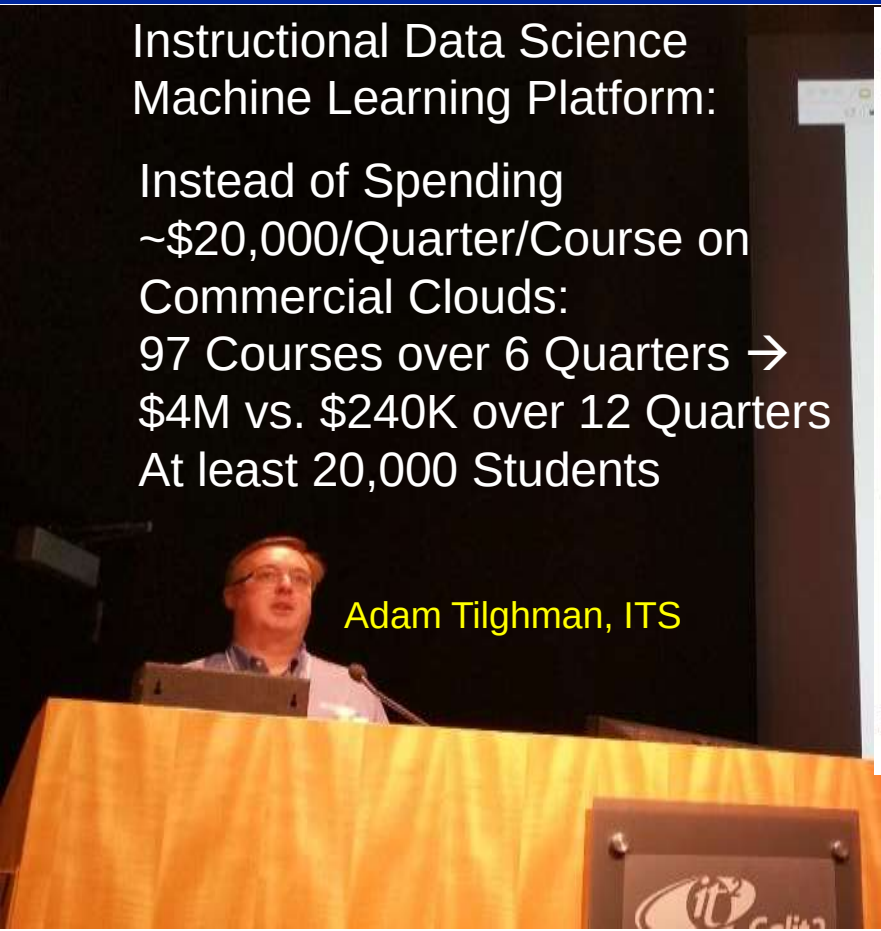
UC San Diego

it²

UCSD's ITS Adapted PRP FIONA8s To Support Data Science Courses

Instructional Data Science Machine Learning Platform:

Instead of Spending ~$20,000/Quarter/Course on Commercial Clouds:
97 Courses over 6 Quarters → $4M vs. $240K over 12 Quarters
At least 20,000 Students

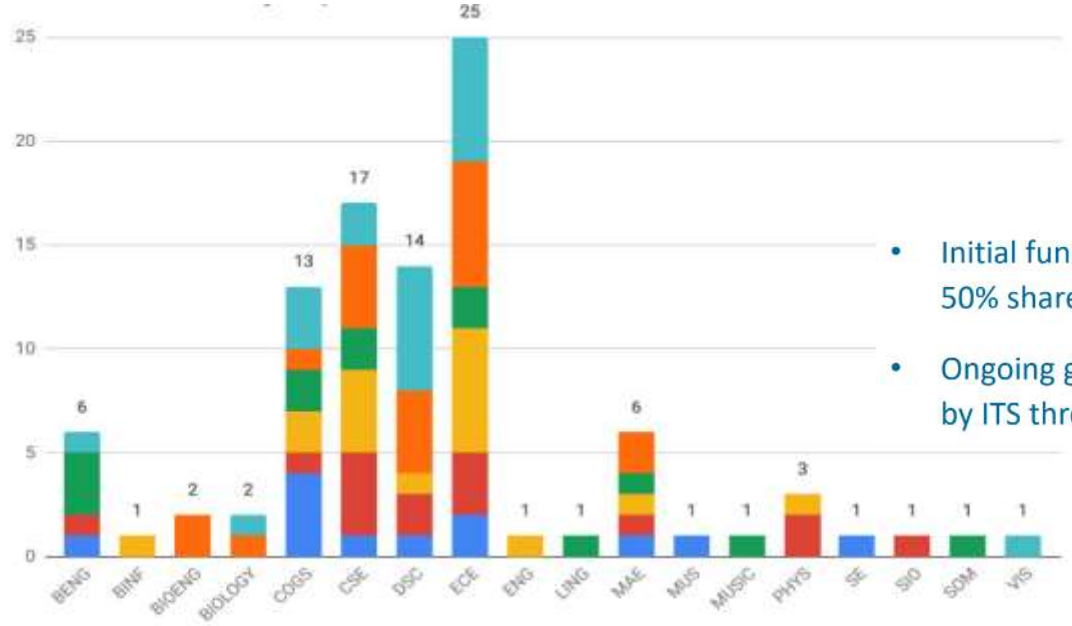




- 104 GPUs
  - 80 GTX 1080Ti
  - 16 RTX 2080Ti
  - 1.05 Petaflops
- 28 nodes
  - 544 CPU cores
  - 6.5TB RAM
- 40TB Flash Storage
- 10G networking

Source: UCSD ITS

The Student GPUs
Have Supported a Broad Set of Courses Across Campus



- Initial funding: ~$150K, 50% from ITS & 50% shared by JSOE and Cog Sci
- Ongoing growth: $50K/yr commitment by ITS through FY20/21

Selected Spring 2018 Courses

Advanced Statistical NLP
Vision & Learning
Comp Phys II:PDE & Matrix Modeling
Deep Learning and Applications
Robot Reinforcement Learning
Intelligent Vehicles and Assistance Systems
Intro to Autonomous Vehicles
Optimization and Acceleration of DL

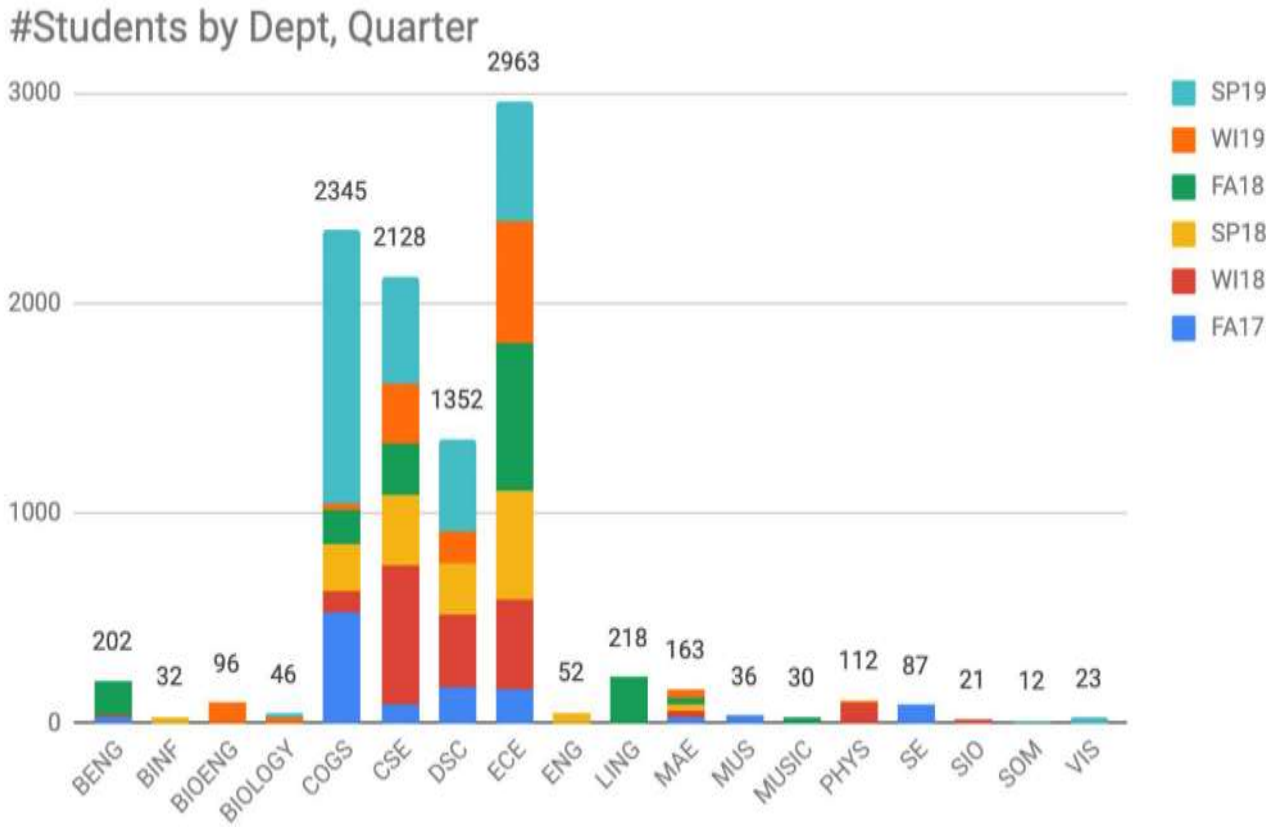
11

UC San Diego

Source: UCSD ITS

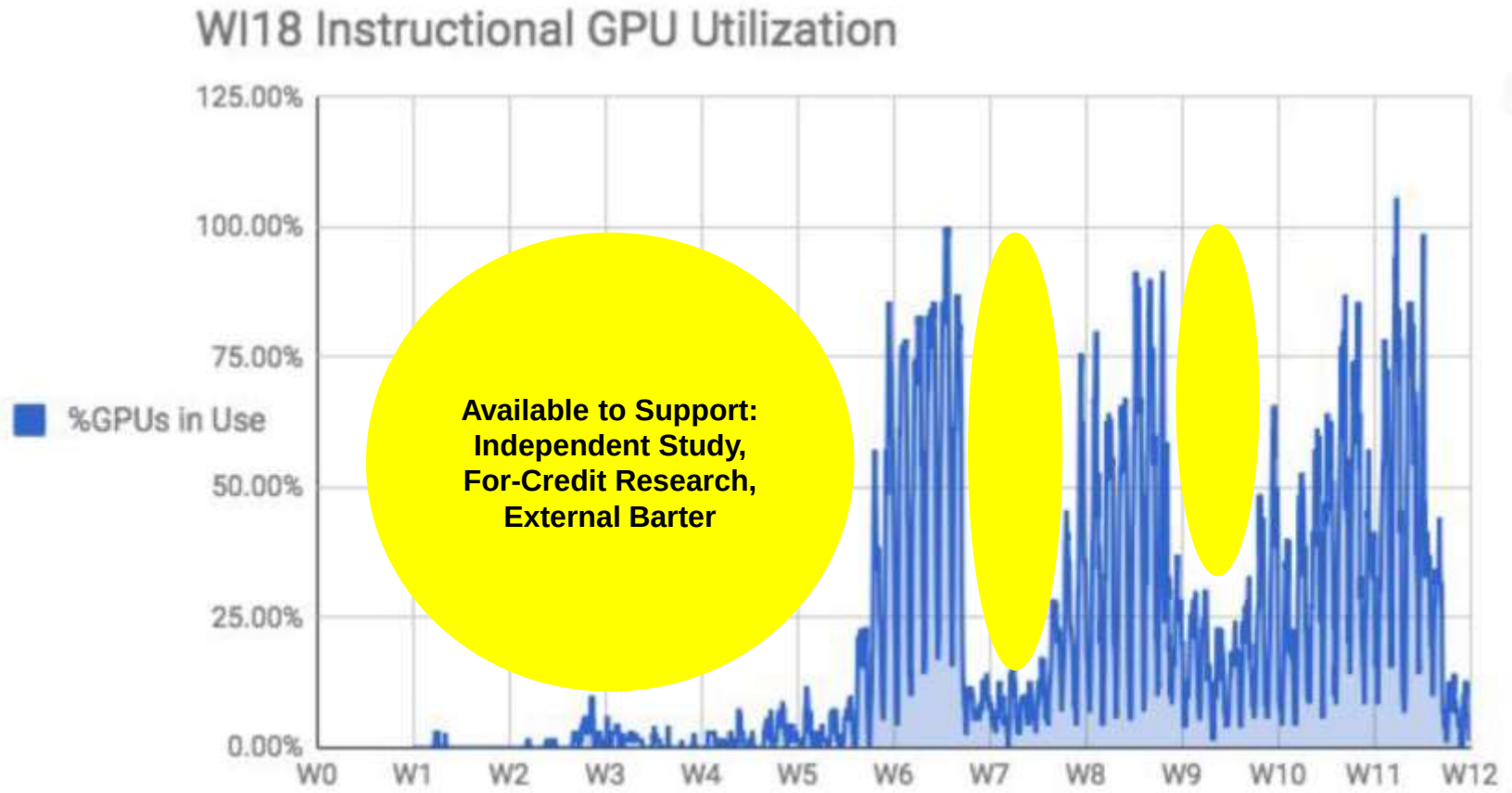
The ITS GPUs
Have Supported Thousands of Students

Source: UCSD ITS



UC San Diego

Student GPU Demand Is Variable Allowing for Other Student Uses



Source: UCSD ITS

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

How Kubernetes Conquered 2017 (and is Positioned for 2018)





Dan Meyer
December 25, 2017

sdx central

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,

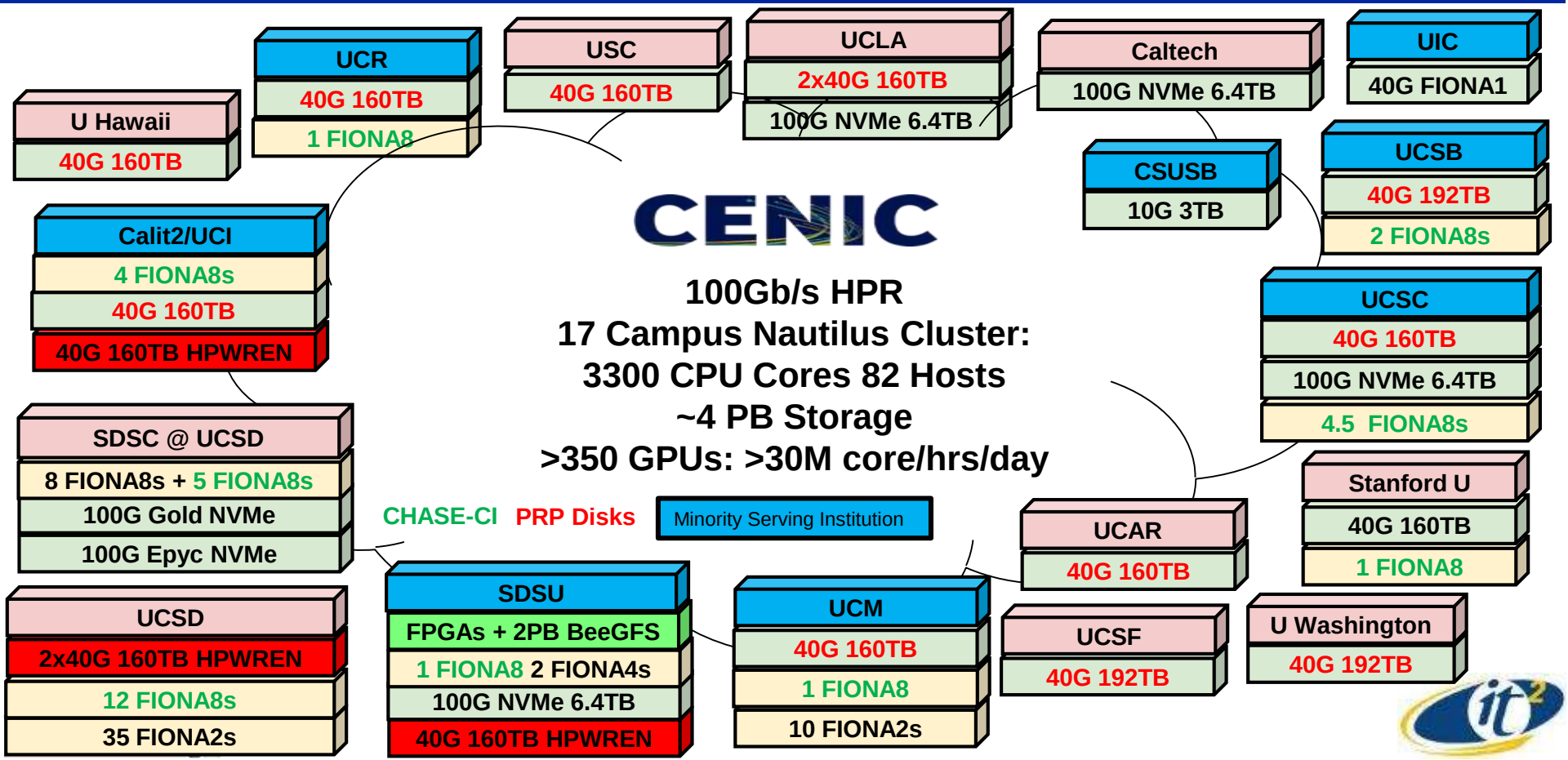
UC San Diego

CLOUD NATIVE COMPUTING FOUNDATION

SDSC

it²

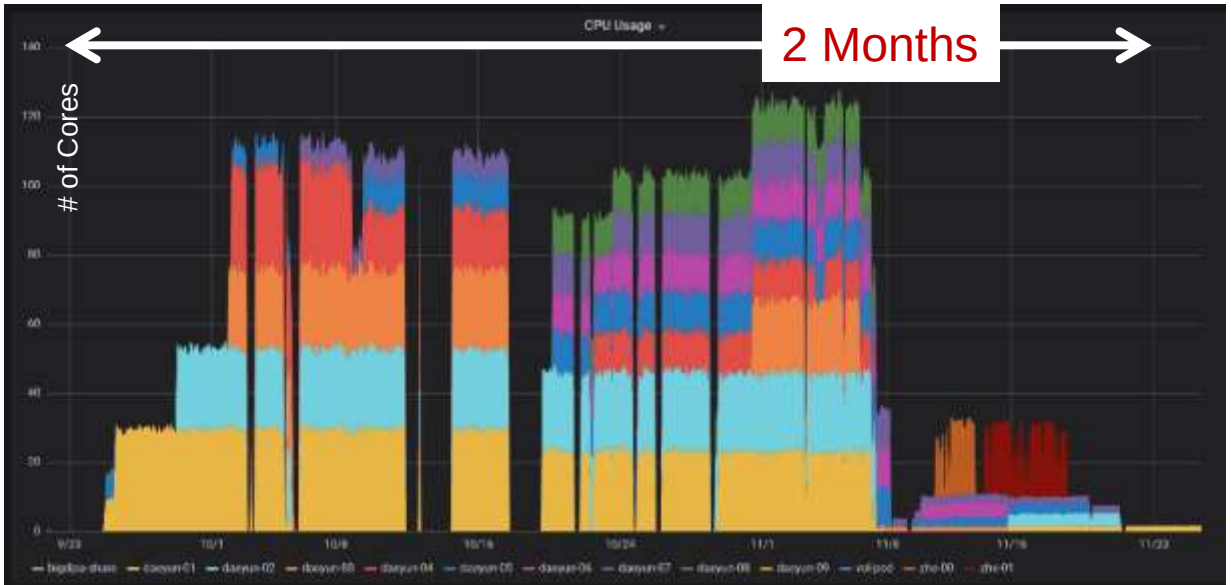
Nautilus Kubernetes Cluster
Connected by CENIC in California



Major CHASE-CI Usage by UCI Over PRP to UCSD CPUs/GPUs

Cognitive Anteater Robotics Laboratory (CARL) supervised by Prof. Jeff Krichmar





UCICompVis Group supervised by Prof. Charless Fowlkes

Demo Last Night From Data Think Tank Lab

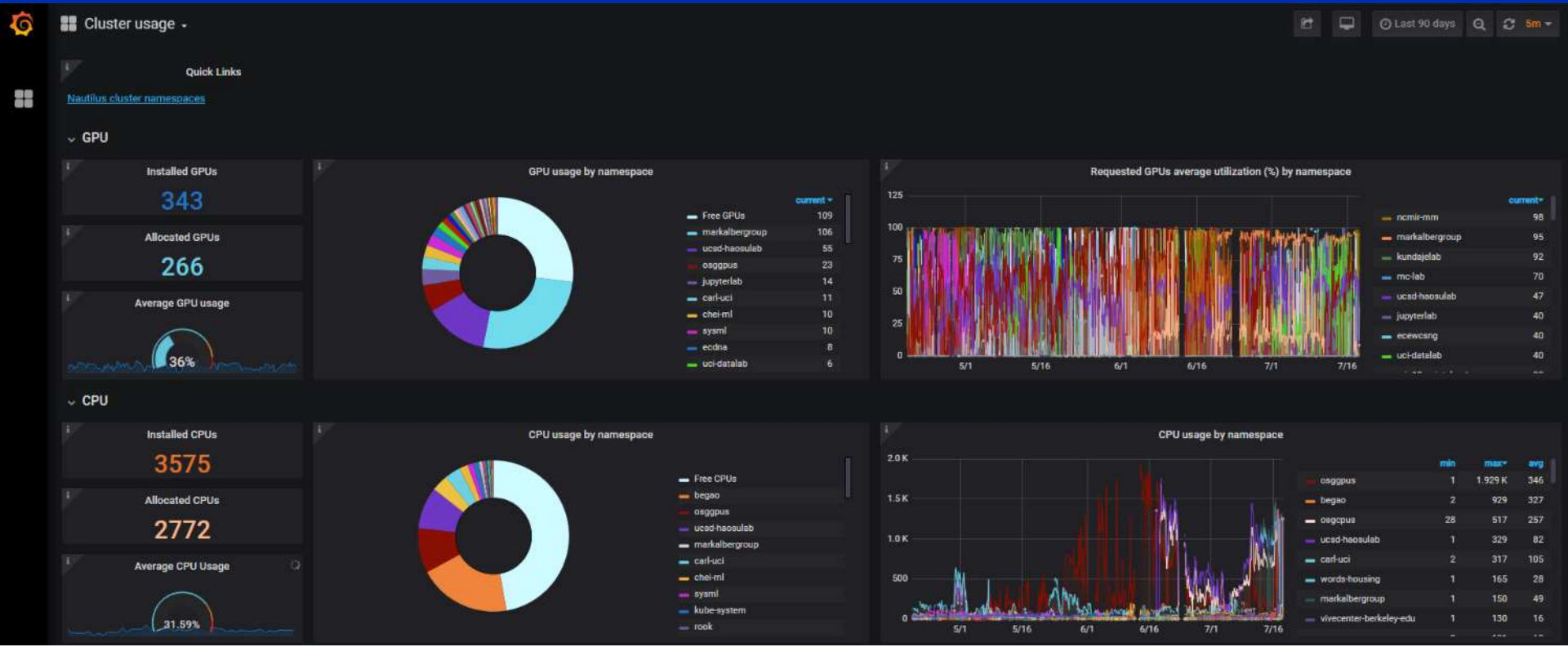
UC San Diego

Very Cost-Effective for Academic Machine Learning and Data Sharing

- **Data science researchers need DTNs with lots of storage, encryption and lots of GPUS**
- **One UC spends $40,000 in cloud GPU per published grad student paper**
- **Another spends $20,000 for undergrad ML AWS access in just one course**
- **Instead, add to our Nautilus hypercluster (or clone it & federate):**
  - **UCSD ECE Department bought 4 FIONA8s, buying 4 more**
  - **UCSD Physics Department. bought 3 FIONA8s, buying 3 more**
  - **UCSD CSE researchers bought/are buying FIONA8s to add to Nautilus**
  - **UCSD Instructional IT has 13 FIONA8s for Machine Learning/AI class labs**
- **Working Storage on Nautilus FIONAs is**
  - **very inexpensive (12TB drives are ~$430 each—16 per FIONA. FISA encrypted drives @ same cost)**
  - **and very high speed (most FIONAs are 40/100G and are located in ScienceDMZs)**

Clemson's Alex Feltus: "I cannot wait to add a node to the Nautilus compute fabric!" 5/22/2019

Nautilus Usage
April 17, 2019 to July 17, 2019



UC San Diego

it²

Biggest Nautilus GPU Users
December – April, 2019

Nautilus Namespace Dec 18-Apr 19	GPU Hours	GPU Core Hrs	Institution	PI	Admins
osggpus	4622.79	15,953,248	University of California, San Diego	Frank Wuerthwein	"Igor Sfiligoi" <isfiligoi@ucsd.edu>
guru-research	1442.00	4,976,342	University of California, San Diego	Gary Cottrell	"Sumit Binnani" <sbinnani@eng.ucsd.edu>
viscomp	1370.86	4,730,838	University of California, San Diego	Ravi Ramamoorthi	"Alexandr Kuznetsov" <a1kuznetsov@ucsd.edu>
ucsd-haosulab	1062.50	3,666,688	University of California, San Diego	Hao Su	"Jiayuan Gu" <jigu@eng.ucsd.edu>
cari-uci	962.32	3,320,966	University of California, Irvine	Jeff Krichmar	"Hirak Jyoti Kashyap" <kashyaph@uci.edu>
ecewcsng	682.71	2,356,032	University of California, San Diego	Dinesh Bharadia	"Dinesh Bharadia" <dineshb@eng.ucsd.edu>
kundajelab	639.76	2,207,812	Stanford University	Anshul Kundaje	"Anna Shcherbina" <annashch@stanford.edu>
chei-ml	563.08	1,943,189	University of California, San Diego and SIO	Falko Kuester	"Vid Petrovic" <vipetrovic@ucsd.edu>
ecdna	533.32	1,840,487	University of California, San Diego	Vineet Bafna	"Utkrisht Rajkumar" <urajkumar@ucsd.edu>
sysml	457.84	1,580,006	University of California, Santa Barbara	Paul Weakliem	"Zhang, chris" <zhizhouzhang@ucsb.edu>
ucicompvis	403.02	1,390,822	University of California, Irvine	Charless Fowlkes	"Charless C Fowlkes" <fowlkes@uci.edu>
domain-adaptation	241.21	832,416	University of California, San Diego	Nuno Vasconcelos	"Yunsheng Li" <yul554@eng.ucsd.edu>
suncave	204.33	705,143	University of California, San Diego	Tom DeFanti	"John Graham" <jjgraham@ucsd.edu>
xai	201.71	696,101	University of California, San Diego	Jurgen Schulze	"Jurgen Schulze" <jschulze@ucsd.edu>
xaidevel	190.09	656,001	University of California, San Diego	Jurgen Schulze	"Kamran Alipour" <kalipour@eng.ucsd.edu>
kube-environment	175.33	605,064	University of California, San Diego	Bill Lin	"Weijia Wang" <wweijia@eng.ucsd.edu>
mesl	162.39	560,408	University of California, San Diego	Rajesh Gupta	"Jason Koh" <jhkoh@ucsd.edu>
jupyterlab	155.47	536,527	University of California, San Diego	Larry Smarr	"John Graham" <jjgraham@ucsd.edu>
chrs	150.16	518,202	University of California, Irvine	Soroosh Sorooshian	"Eric Jay Shearer" <eshearer@uci.edu>
connect	136.84	472,235	University of California, San Diego	Scott L. Sellars	"Scott Sellars" <scottlsellars@gmail.com>
satellite-unicef	115.57	398,832	University of California, San Diego	Ilkay Altintas	"Jiaxin Li" <jil160@ucsd.edu>

CSE ECE Struc. Eng

UC San Diego

it²

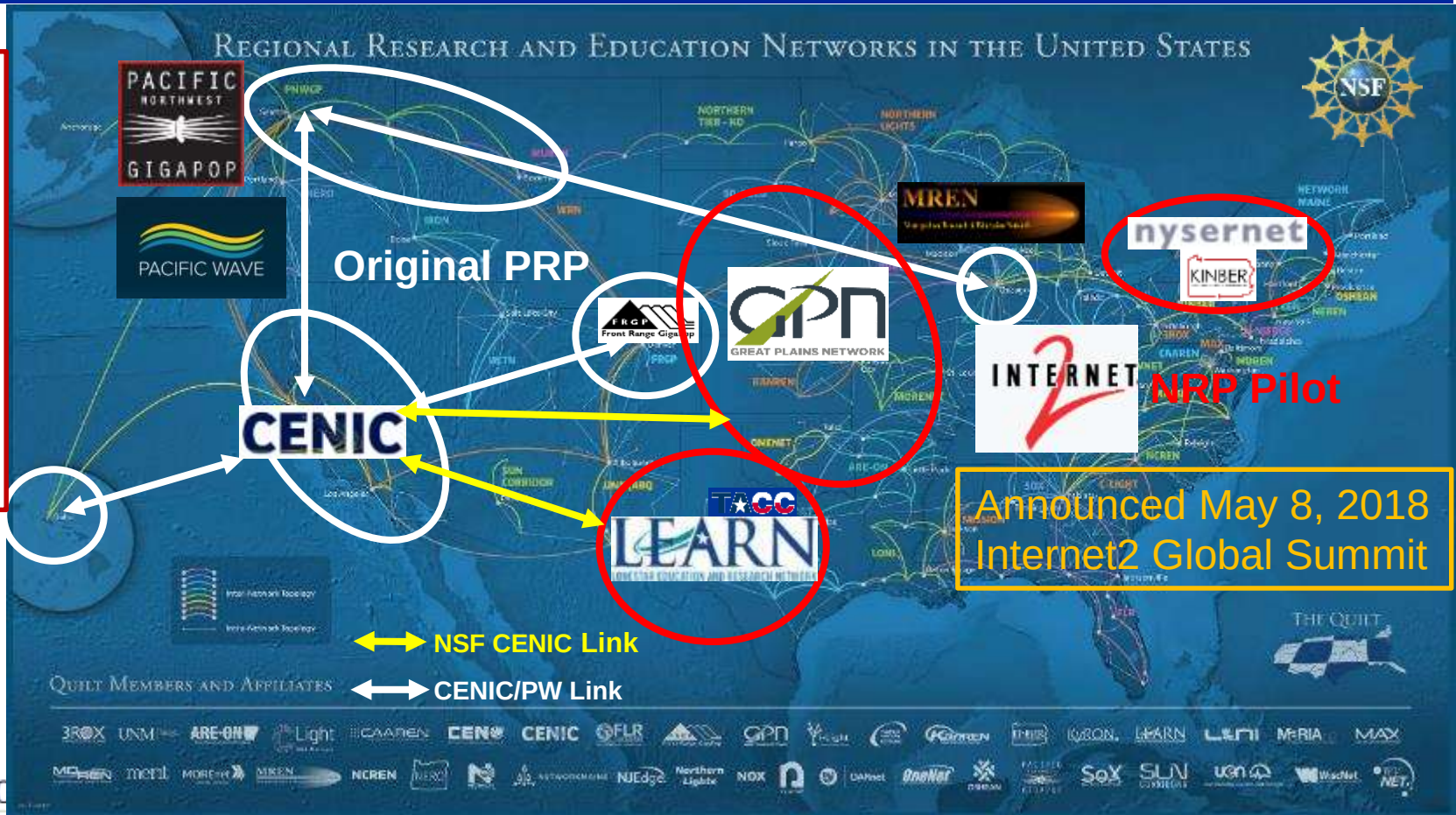
Extra slides

2018-2019: National-Scale Pilot - Using CENIC & Internet2 to Connect Quilt Regional R&E Networks

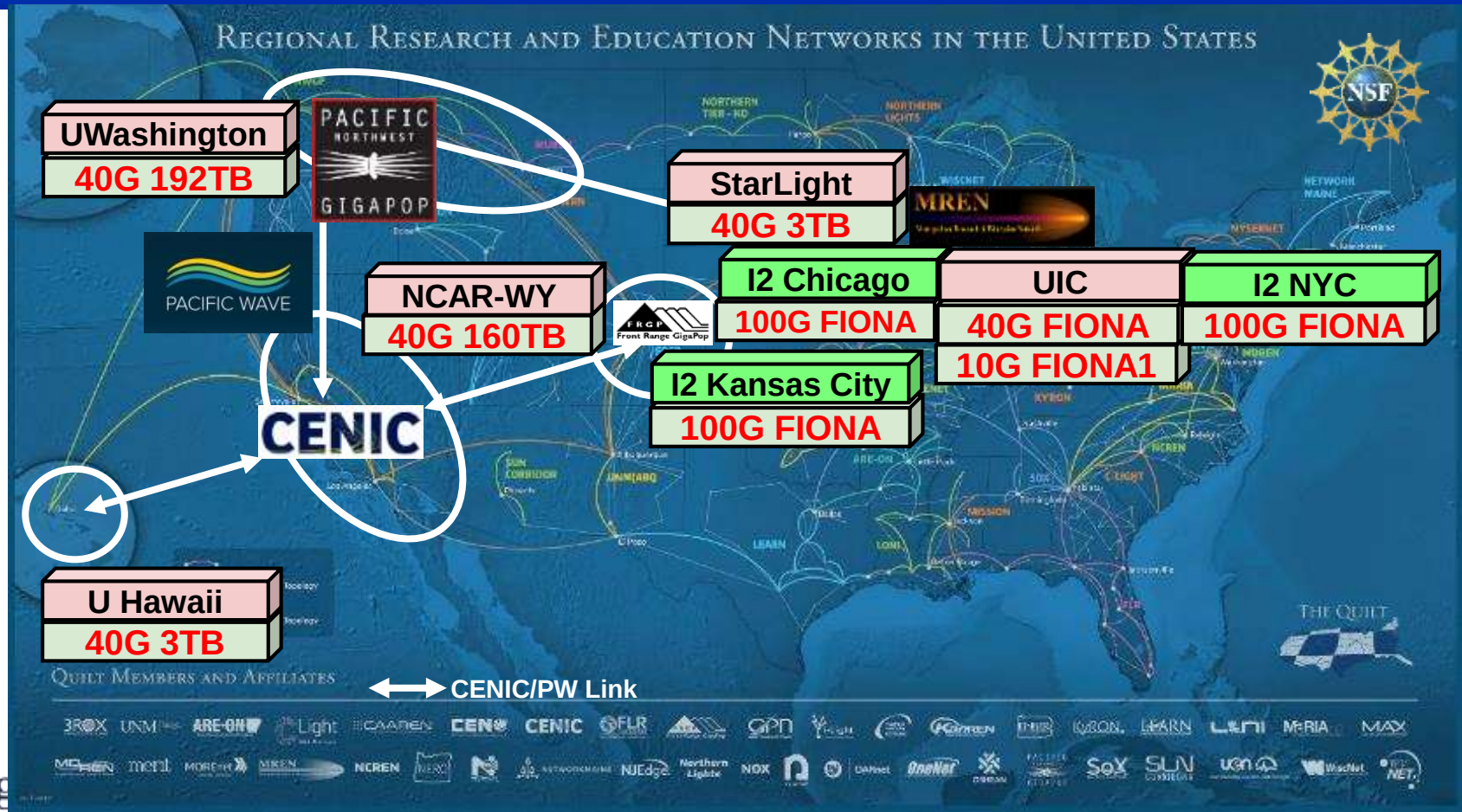
"Towards The NRP"
3-Year Grant Funded by NSF
$2.5M
OAC-1826967
PI Smarr
Co-PIs Altintas Papadopoulos Wuerthwein Rosing
Mgr: DeFanti

SDSC

UC San Diego

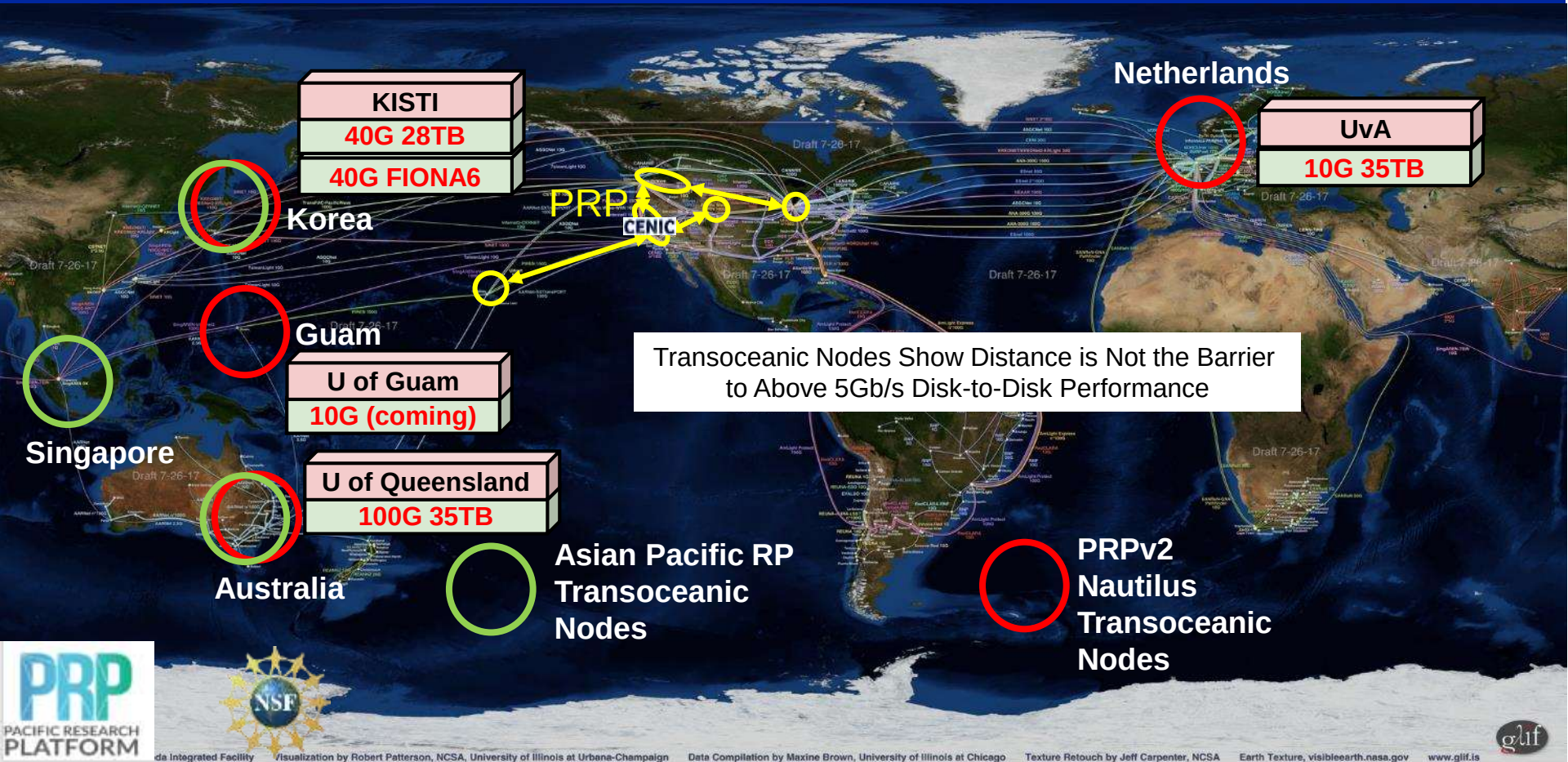


United States PRP Nautilus Hypercluster FIONAs
We Now Connect 3 More Regionals and 3 Internet2 sites



it²
ESnet
CENIC
SDSC
CITRIS BANATAO INSTITUTE
UC San Dieg

Global PRP Nautilus Hypercluster Is Rapidly Increasing Partners Beyond Our Original Partner in Amsterdam—May 2019



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



CMS

Particle Physics

Biomedical ‘omics

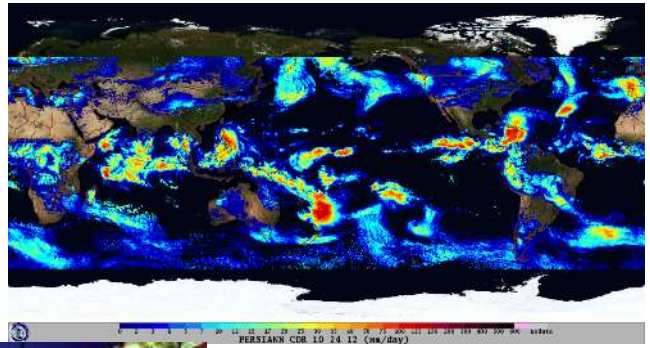


NSF



Telescope Surveys

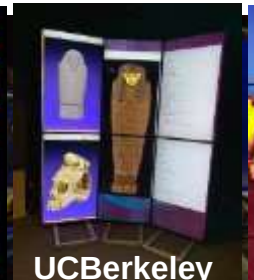
Earth Sciences



Visualization, Virtual Reality, Collaboration



UC San Diego



UCBerkeley



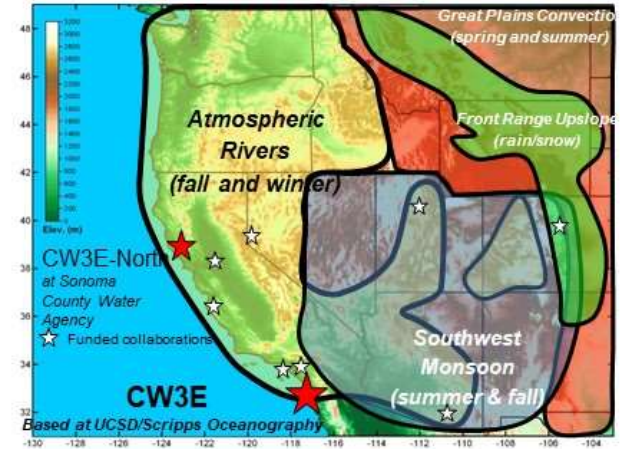
UC Merced

CITRIS THE BANATAO INSTITUTE

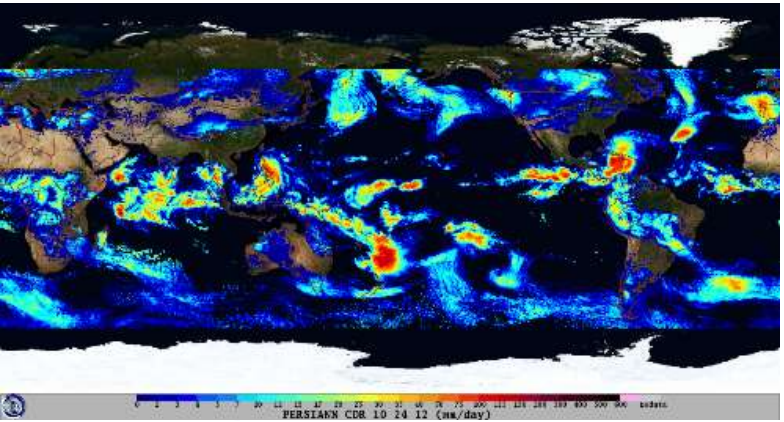
SDSC

it²

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

PRP Shortened Scott Sellar's 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, US State Dept.

UC, San Diego

GPUs



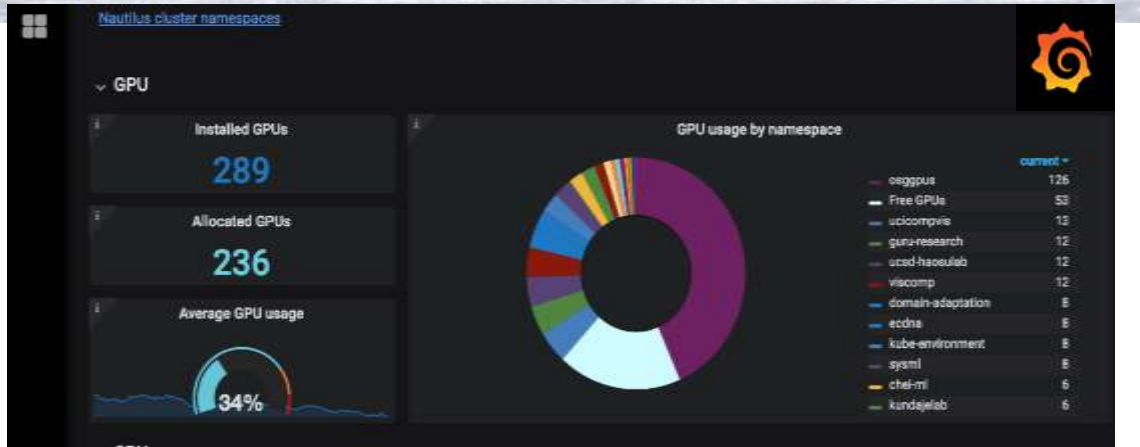
Calit2's FIONA

OSG IceCube Usage on PRP (Purple Segment) 3/9/19: Using 126 GPUs + 142 CPUs + 49 GB RAM

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



IceCube



UC San Diego

PRP Actively Develops Diversity

- **Grants**
  - **3 Female co-PIs**
  - **1 Hispanic co-PI**





- **Campuses**
  - **8 Minority-Serving Institutions in PRP/Nautilus**
- **Workshops**
  - **NRP'18 Workshop Program Committee 80% Female**
  - **Multiple MSI, EPSCoR Focused Workshops**


Jackson State University





PRP MSI Workshop

Presenting FIONettes

UC San Diego

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



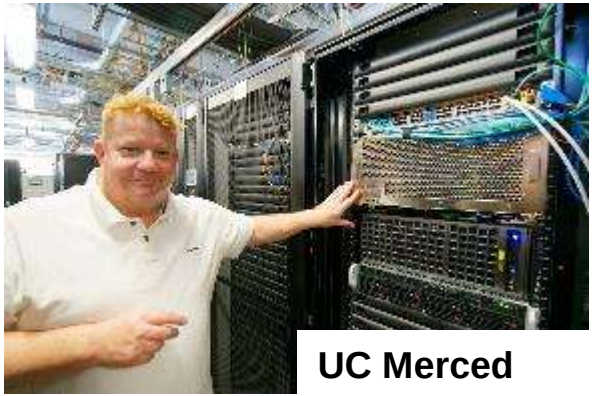
UC Irvine



Stanford



UC Santa Barbara



UC Merced



UC Santa Cruz



UC Riverside

UC San Diego

it²