“NRP Supports AI/ML/DS Researchers and Students at UC San Diego”

Presentation to the HDSI Faculty Council
December 5, 2024

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

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

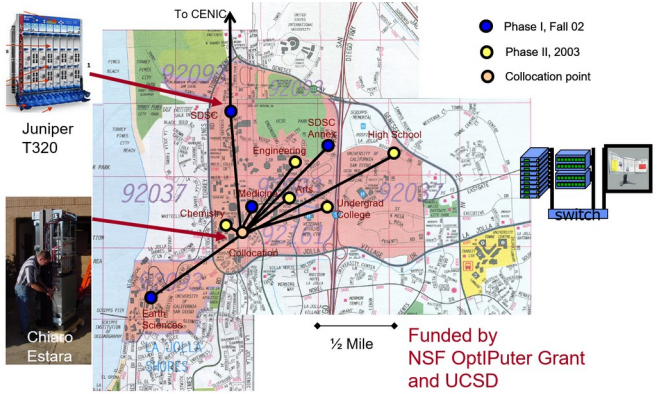
From the OptIPuter to the National Research Platform: 21 Years of Continuous NSF Cyberinfrastructure Grants to UC San Diego

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

2002-2015: The Pacific Research Platform (PRP) Was Built on 3 Prior NSF Awards: OptIPuter, Quartzite, & Prism

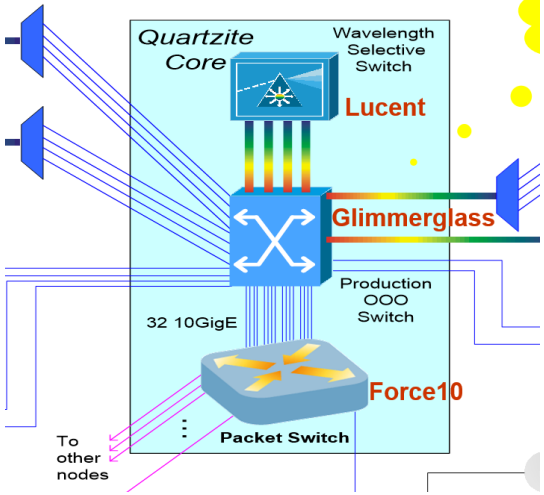
Precursors to DOE Defining DMZ in 2010



PI Smarr, 2002-2009

OptIPuter

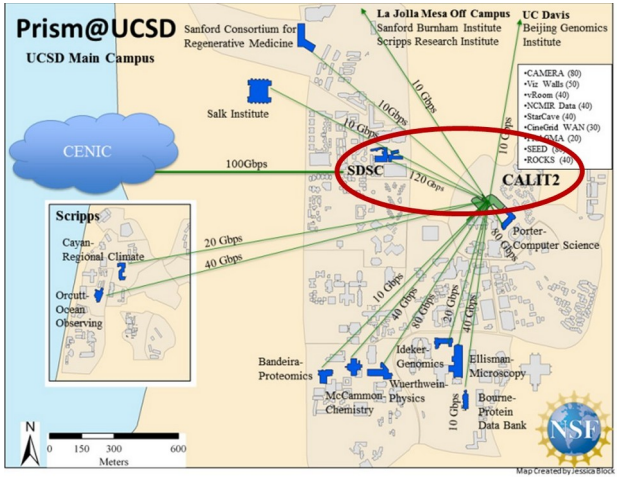
Quartzite Communications Core Year 3



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

NRP NATIONAL RESEARCH PLATFORM

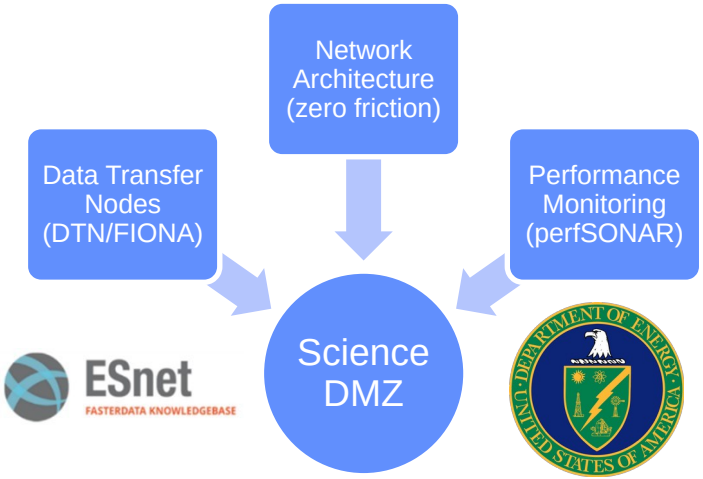
QUALCOMM INSTITUTE 20 YEARS

SDSC

NRP Nautilus

2010-2022:
NSF Adopted a DOE High-Performance Networking Model

ScienceDMZ Coined in 2010 by ESnet



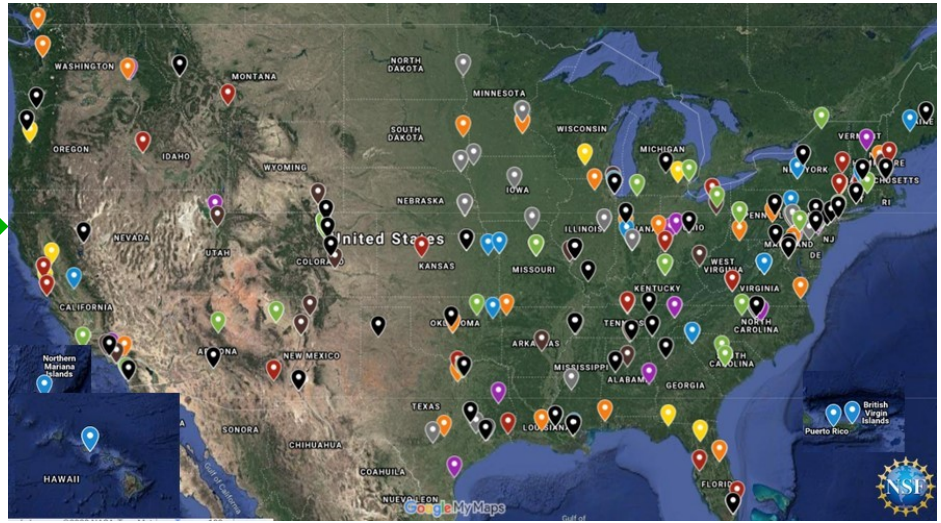
http://fasterdata.es.net/science-dmz/

Slide Adapted From Inder Monga, ESnet

DOE

NSF

NSF Campus Cyberinfrastructure Program 2012-2022



Has Made Over 340 Awards:
Across 50 States and Territories

Slide Adapted From Kevin Thompson, NSF

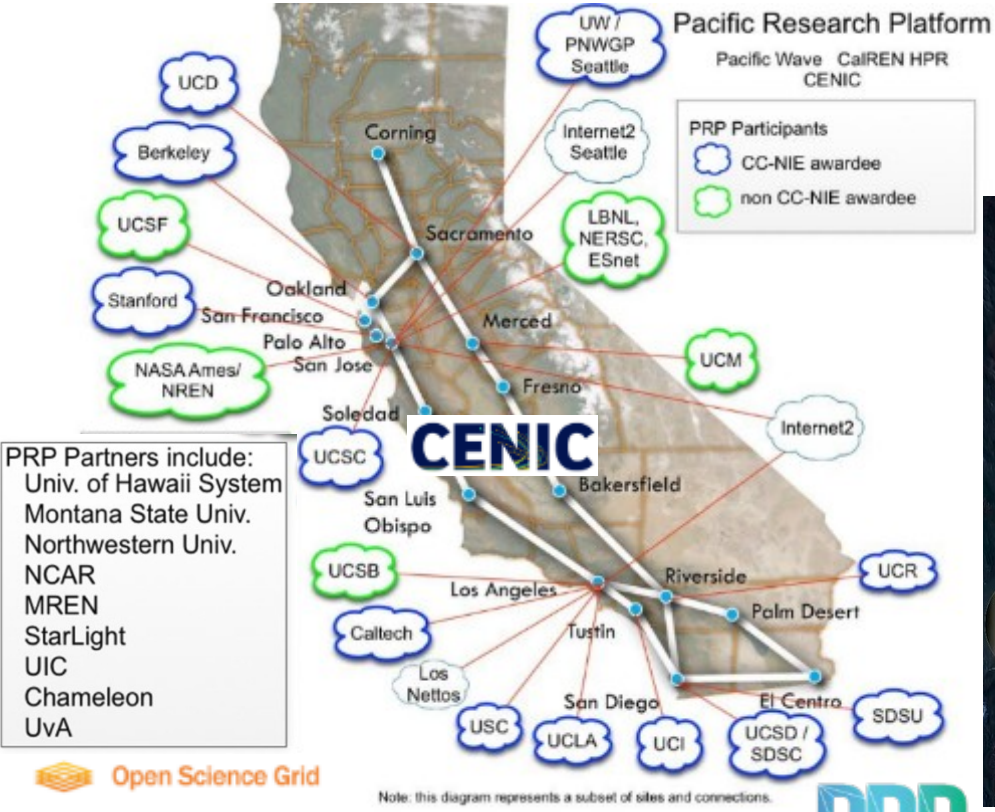
Quartzite

Prism

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

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
Extended – Ended Year 7 in Oct 2022



Source: John Hess, CENIC

NRP NATIONAL RESEARCH PLATFORM

PRP PACIFIC RESEARCH PLATFORM

NRP Nautilus

2015-2022: UCSD Designs and Deploys Rack-Mounted PRP Optical Fiber Endpoints on Campuses

UCSD Designed Data Transfer Nodes (DTNs), called Flash I/O Network Appliances (FIONAs), Solved the 10-400 Gbps Disk-to-Disk Data Transfer Problem

QUALCOMM INSTITUTE

SDSC

FIONA

Software Specifications

Installation Instructions

Hardware Specifications

2023 – 2021 – 2020 – 2019 – 2018 – 2017 – 2016 – 2015

https://nationalresearchplatform.org/fiona/



Add Up to 8 Nvidia GPUs Per 2U FIONA
To Add Machine Learning Capability
Up to 240TB Storage

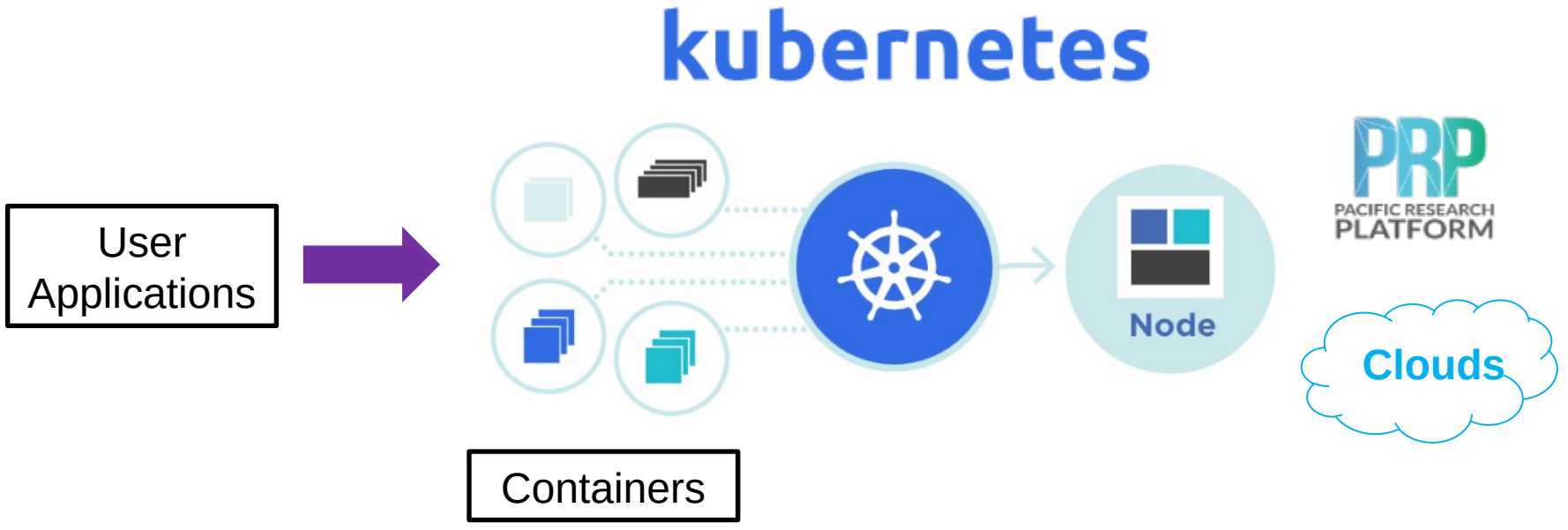
NRP NATIONAL RESEARCH PLATFORM

NSF

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

NRP Nautilus

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

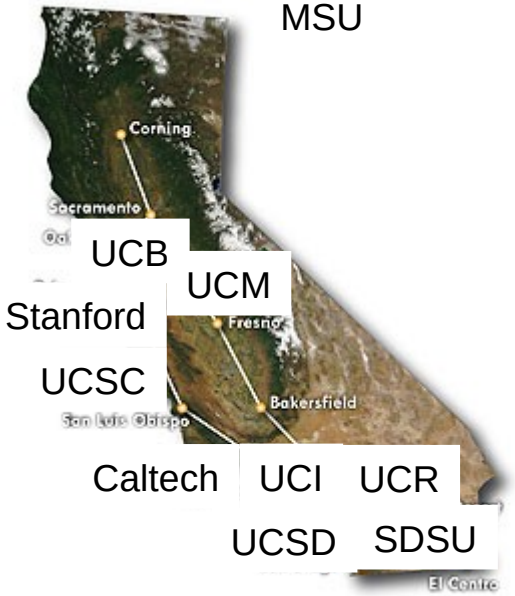


NRP NATIONAL RESEARCH PLATFORM

CLOUD NATIVE COMPUTING FOUNDATION

NRP Nautilus

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



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

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
Co-PIs:

- Tajana Rosing
- Ken Kreutz-Delgado
- Ilkay Altintas
- Tom DeFanti

JACOBS SCHOOL OF ENGINEERING

SDSC UC San Diego

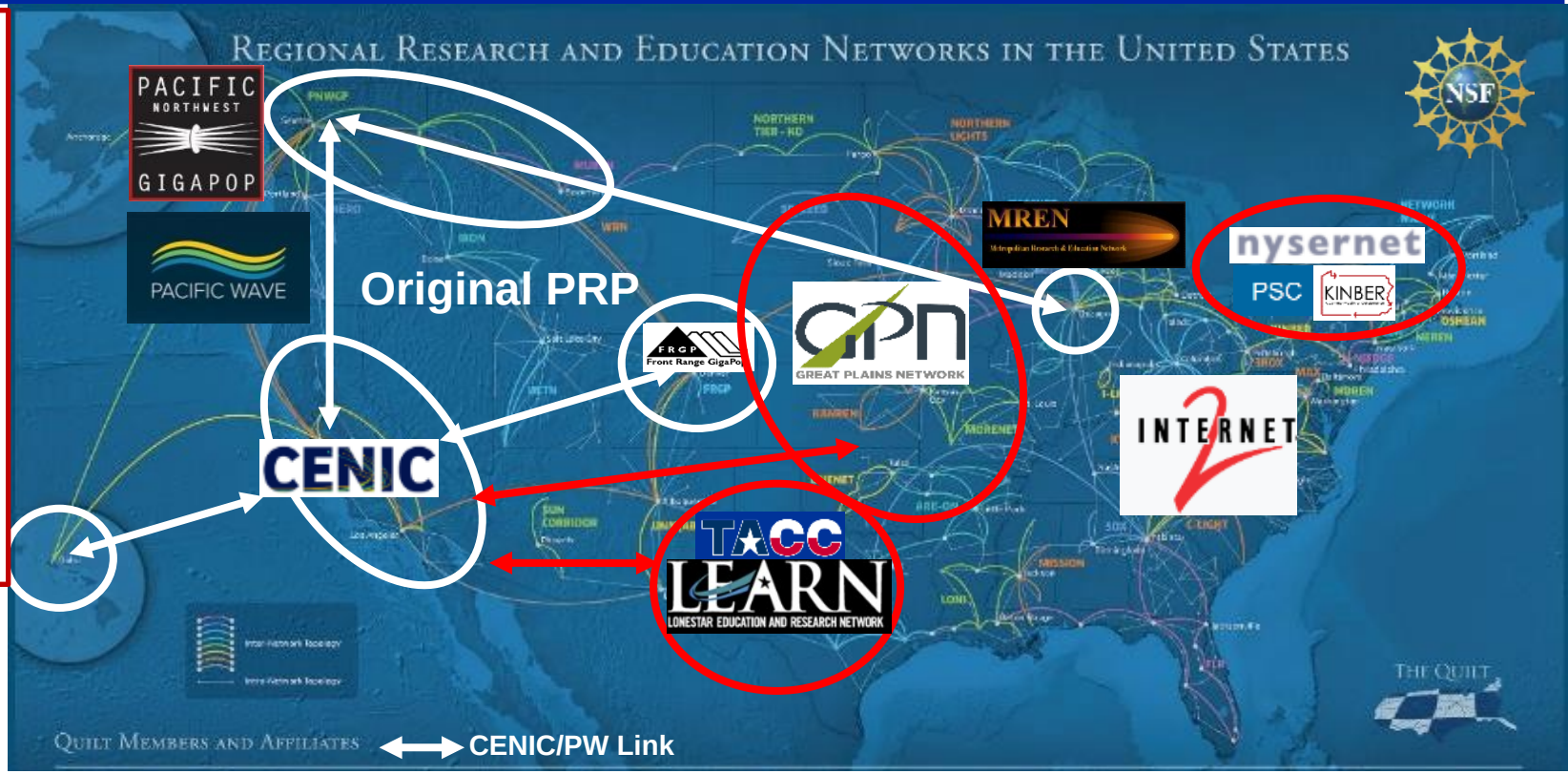
NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

2018-2021: Toward the National Research Platform (TNRP) - Using CENIC & Internet2 to Connect PRP to 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

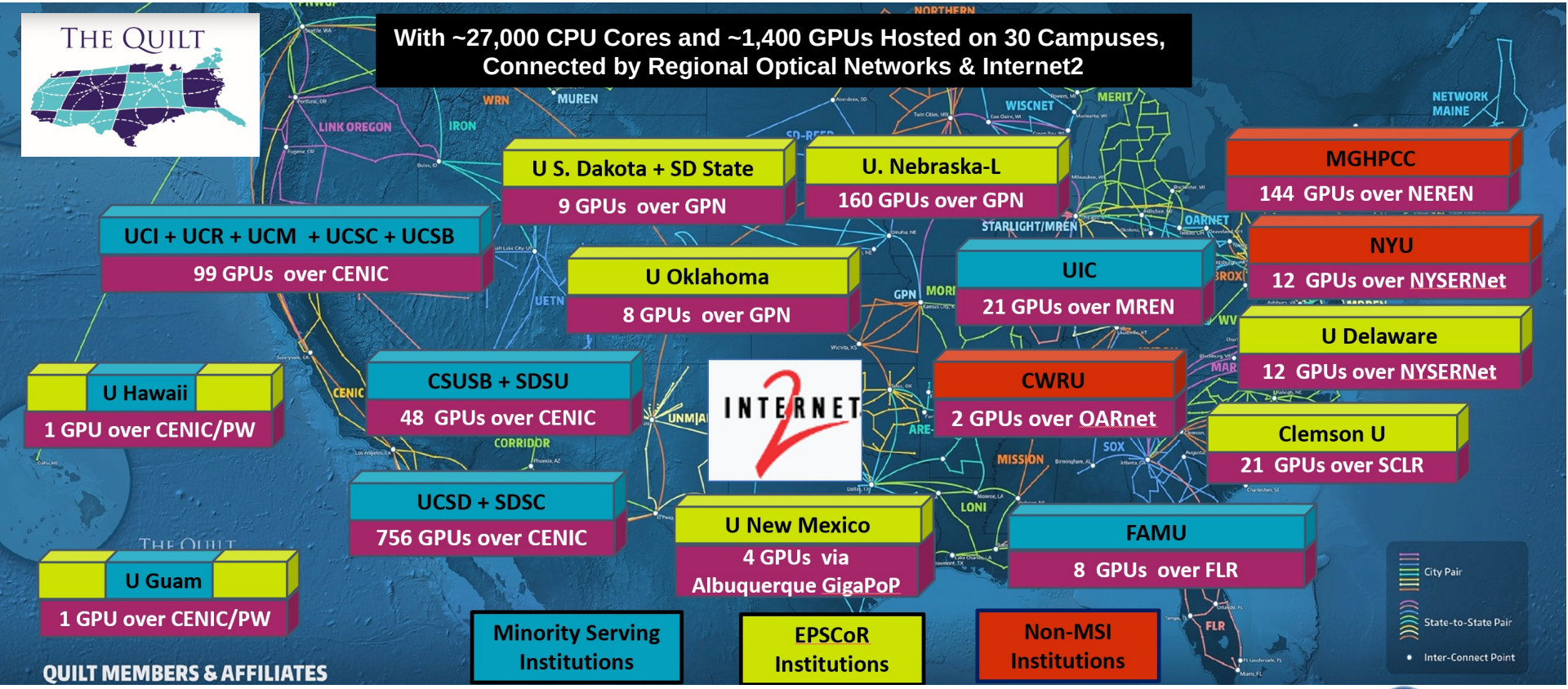


it²
SDSC

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

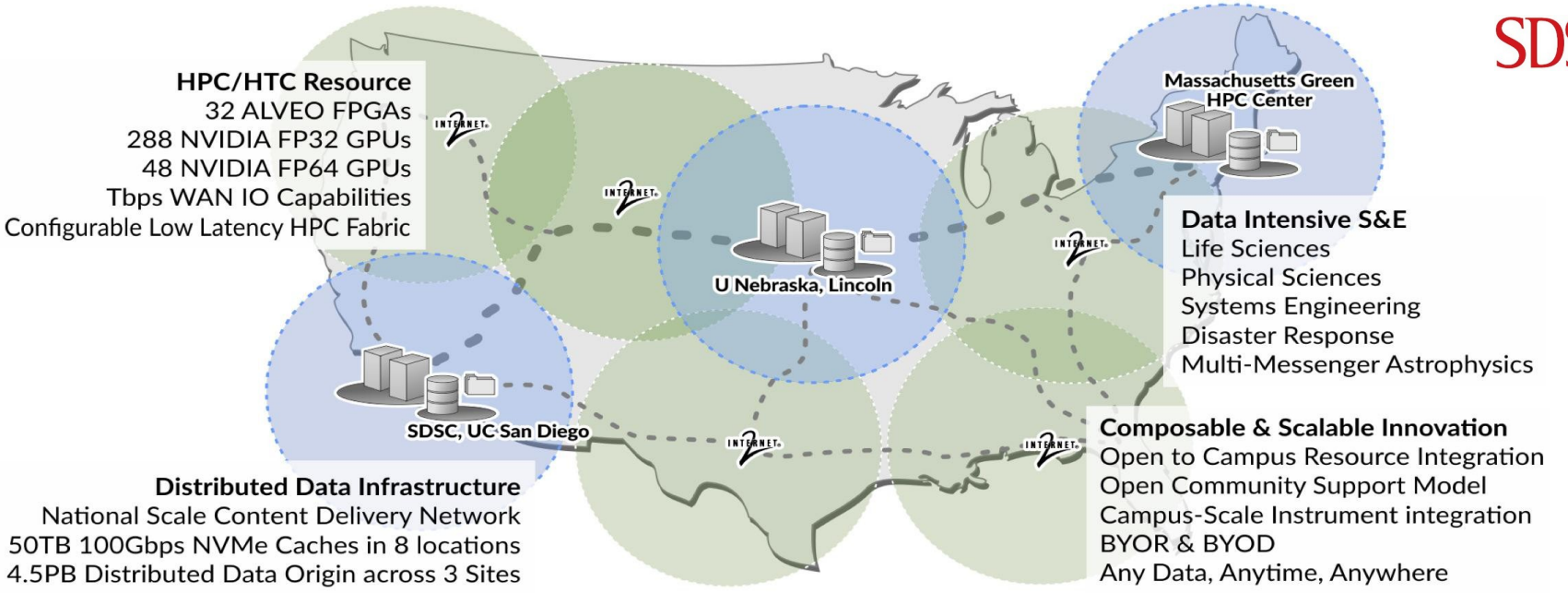
Nautilus is NRP's Multi-Institution Hypercluster Which Creates a Community Owned and Operated "AI Resource"



Slide Courtesy of Tom DeFanti, UCSD; PRP co-PI

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

SDSC



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)

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

2023 - The National Research Platform Emerges As a Unification of Over 20 Years of NSF Cyberinfrastructure Grants

NRP NATIONAL RESEARCH PLATFORM

NEWS GRANTS TECHNOLOGY COMMUNITY JOIN / CONTACT



Professor Frank Würthwein

- Director - SDSC
- Distributed High-Throughput Computing Lead - SDSC

Designed for Growth and Inclusion

The National Research Platform (NRP) is a partnership of more than 50 institutions, led by researchers and cyberinfrastructure professionals at UC San Diego, supported in part by awards from the National Science Foundation.

https://nationalresearchplatform.org/

The PRP→NRP Has Emphasized Expanding Diversity and Inclusion

The Expansion from PRP in 2015 to the NRP in 2023:

- 6 States ➡ Now 45 States
- 19 Campuses ➡ Now 135 Campuses
- 9 Minority Serving Institutions ➡ Now 25 MSIs
- 2 NSF EPCoR States ➡ Now 20 EPSCoR States, 2 Territories, and Wash DC

The NRP Web Site Provides Widely-Used Open-Source Services For How to Join, Application Research, Development, and Collaboration

Get Quick Access



Information



Getting Started with NRP



Matrix Chat

Over 800 Members

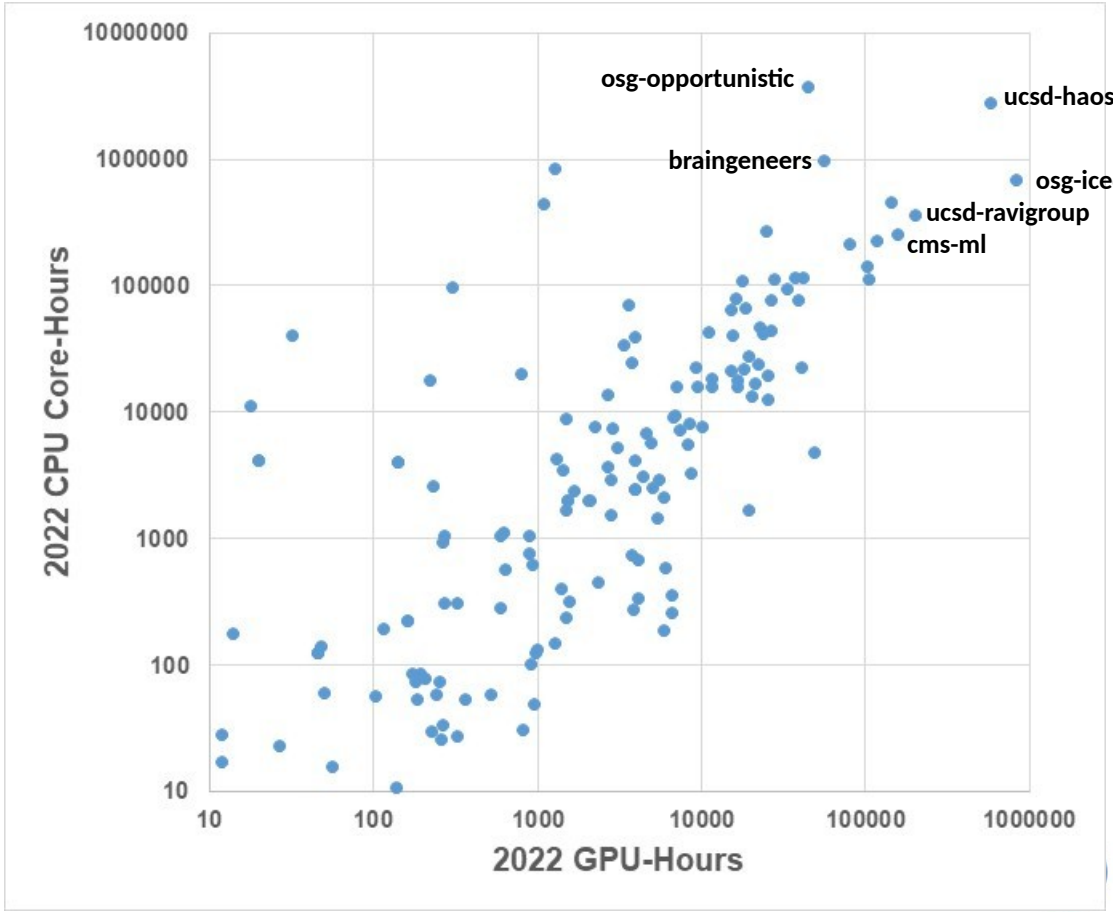
https://nationalresearchplatform.org/

https://nationalresearchplatform.org/nautilus/

NRP Nautilus

Nautilus Supports Namespace Users Over a Range of One Million-Fold Usage/Year

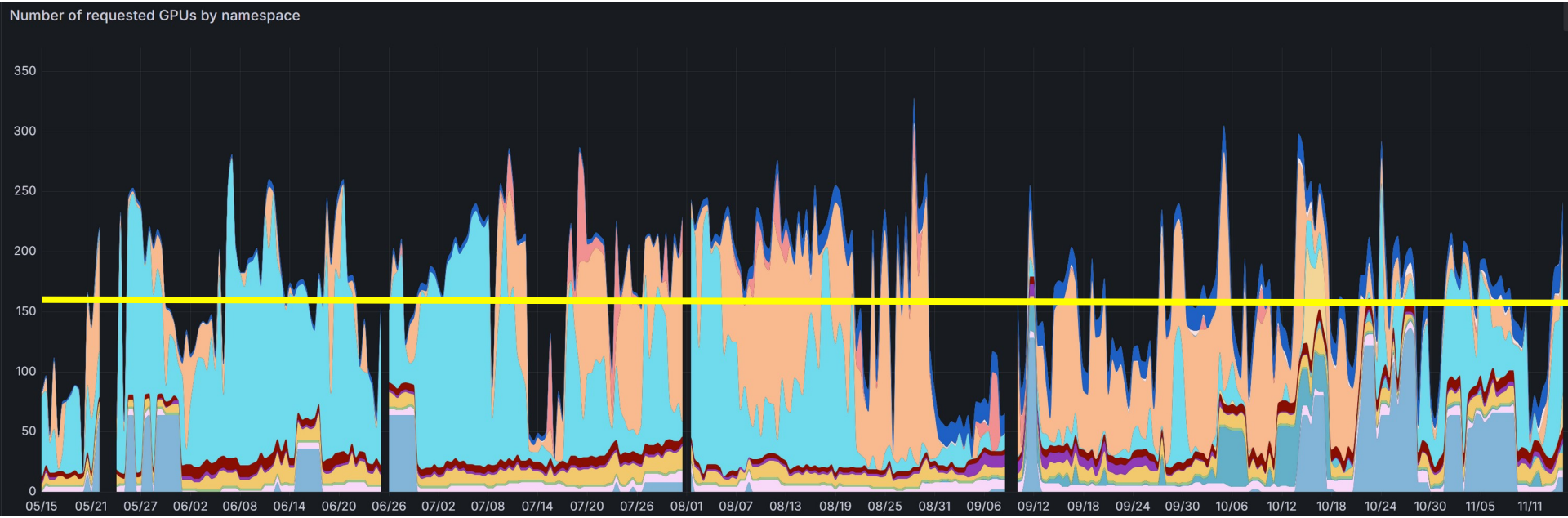
Nautilus Namespaces
Using >10 GPU-hrs/year
Or >10 CPU-hrs/year



NRP Nautilus

NRP NATIONAL RESEARCH PLATFORM

Observatories, jupyter, and Services Namespaces Use ~200 GPUs/Day



Last 6 Months:
5/15/2024 to
11/15/2024

Observatories:
- osg-icecube
- osg-ligo
- bbhnet

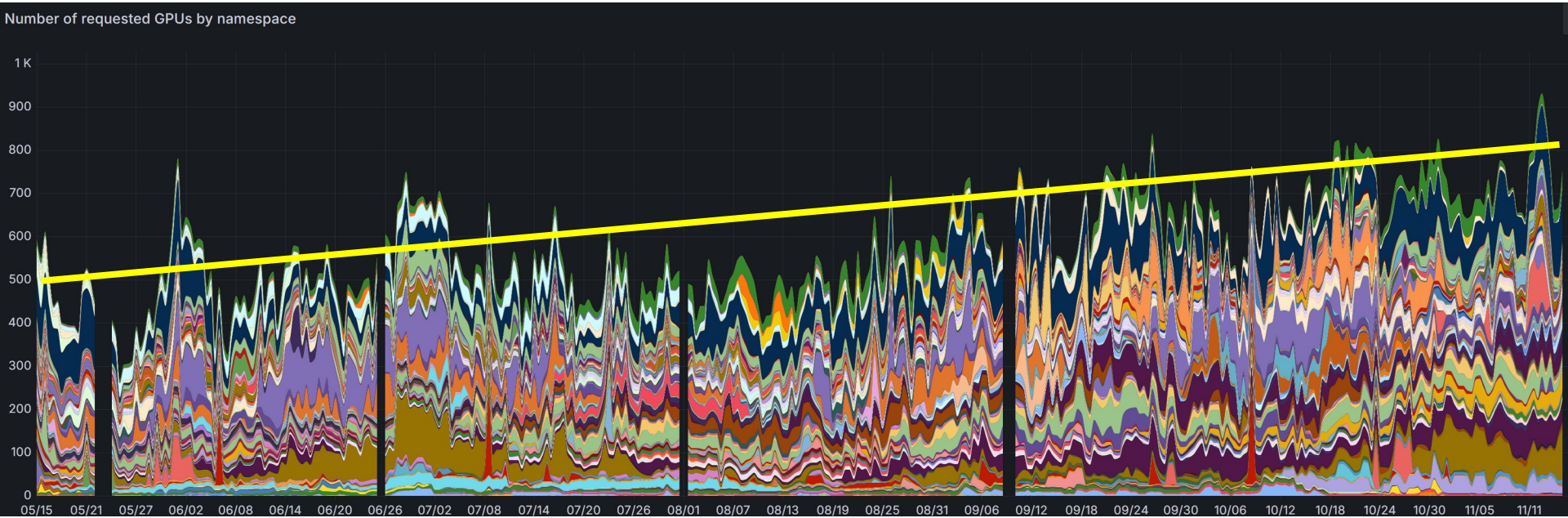
Jupyter:
- jupyterlab
- csu-tide-jupyterlab
- sdsu-rci-jh

Services:
- nrp-llm
- ndp
- juice

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

Individual Researcher Namespaces GPU Usage Has Grown from 500 to 800 GPUs Per Day in the Last 6 Months

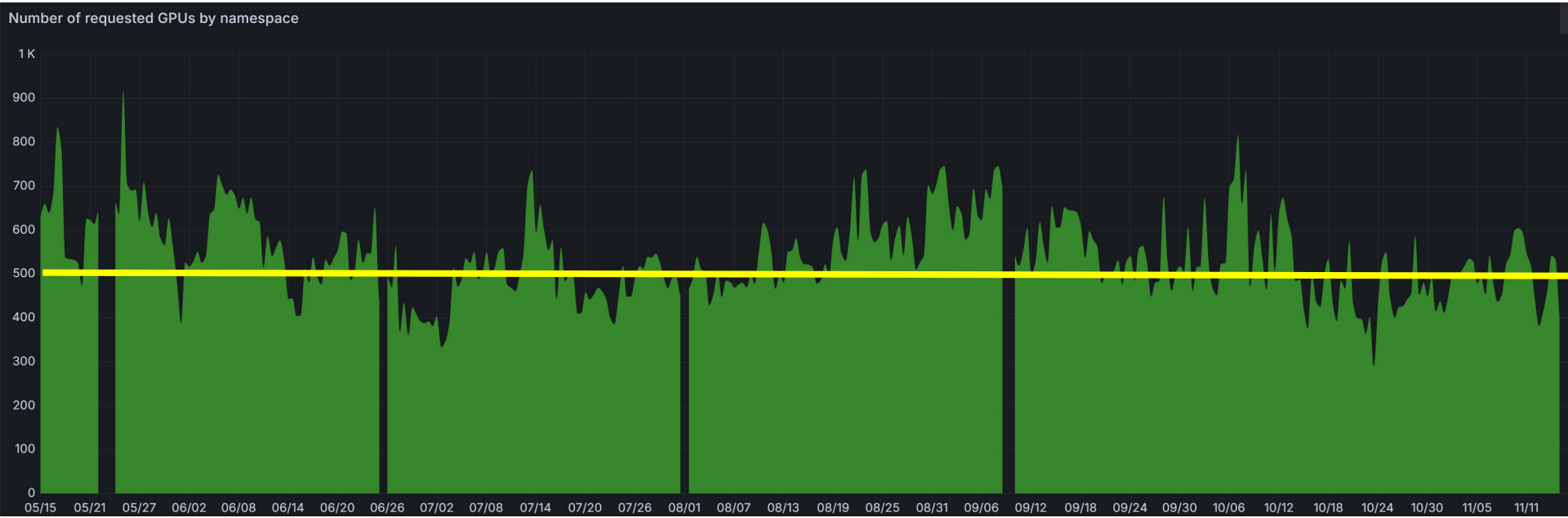


Last 6 Months:
5/15/2024 to
11/15/2024

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

~500 GPUs Are Available on Nautilus Every Day



Last 6 Months:
5/15/2024 to
11/15/2024

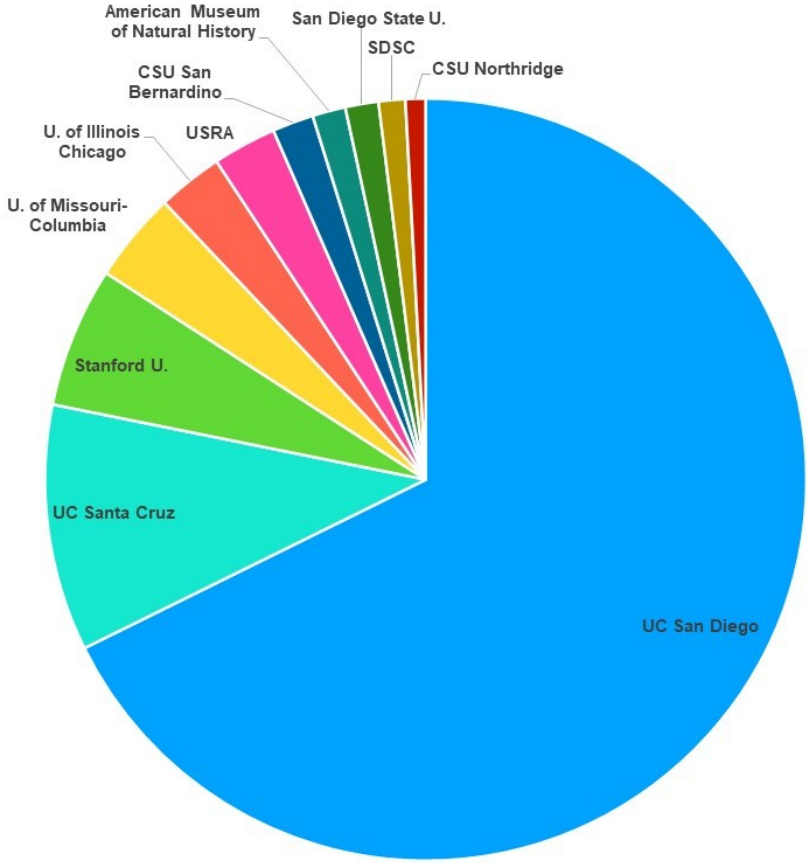
NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

UC San Diego
Dominates University Researcher NRP GPU Usage

2.4 Million
GPU-Hours
Total

Last 6 Months:
5/15/2024 to
11/15/2024



UCSD
1.5 Million
GPU-Hours

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

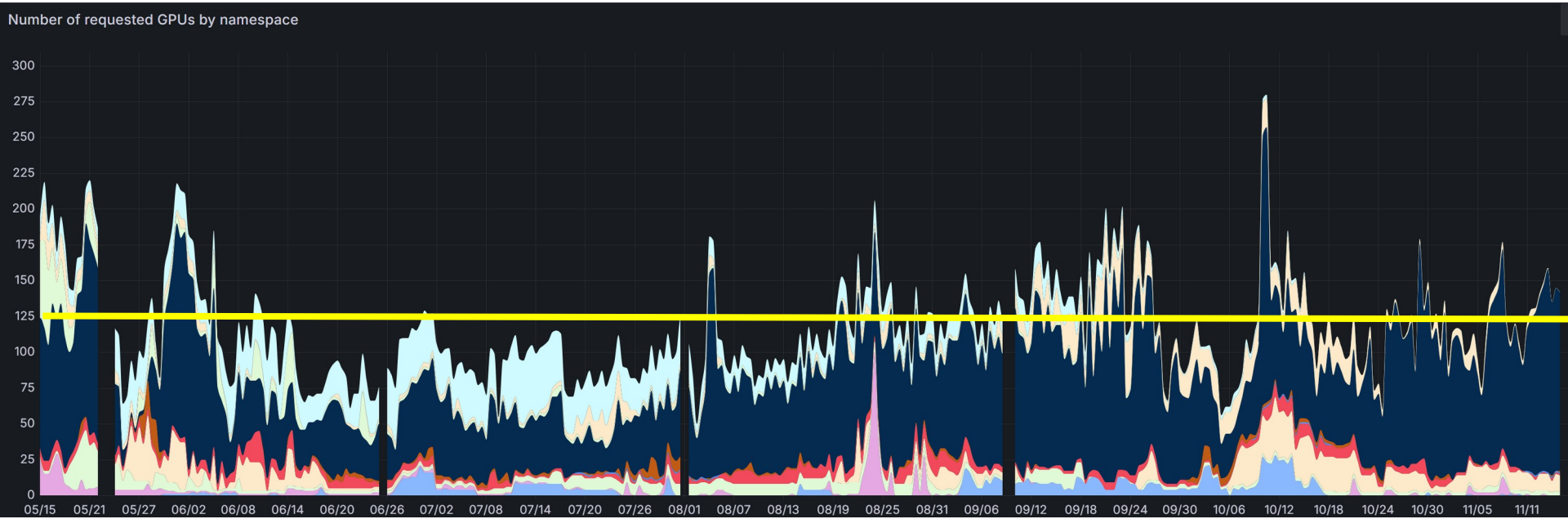
Last 6 Month Analysis of UCSD Nautilus Namespace Usage For CSE/ECE/HDSI Faculty

- **6 Months = 4,380 Hours**
- **Looking at Large Usage Namespaces:**
  - **>4,380 GPU-Hours [1 GPU Running 24x7]**
  - **>52,560 CPU Core-Hours [12 CPU Cores Running 24x7]**

To Avoid Double Counting, if a UCSD Faculty Member is on the HDSI Faculty list (https://datascience.ucsd.edu/faculty/
I count that Faculty as HDSI, even if they might also be a Faculty in CSE or ECE as well.

I Only Cover These 3 UCSD Departments Here

UCSD CSE Faculty (Not in HDSI) Use an Average of ~125 Nautilus GPUs Per Day



Last 6 Months:
5/15/2024 to
11/15/2024

Gary Cottrell
*Manmohan Chandraker
*Kamalika Chaudhuri
*Sanjoy Dasgupta

*Hadi Esmaeilzadeh
Ravi Ramamoorthi
*Tajana Simunic Rosing
Hao Su

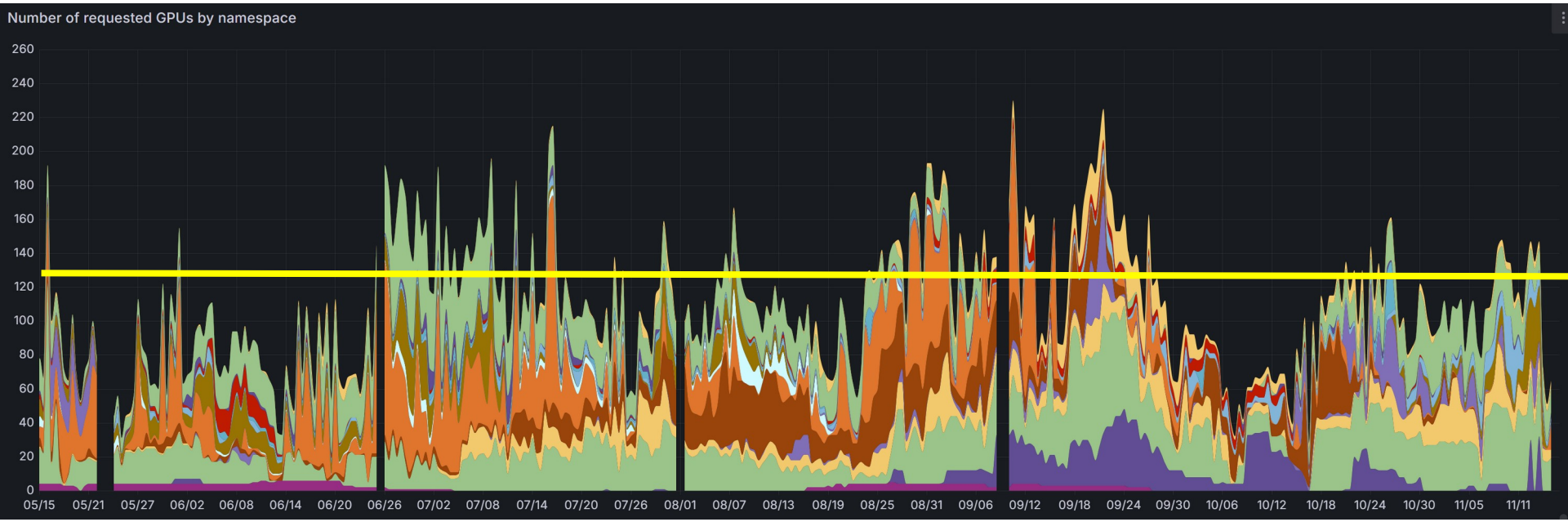
Also Running on Nautilus:
*Ryan Kastner
George Porter

NRP NATIONAL RESEARCH PLATFORM

* HDSI Faculty Affiliate

NRP Nautilus

UCSD ECE Faculty Use an Average of ~125 Nautilus GPUs Per Day



Last 6 Months:
5/15/2024 to
11/15/2024

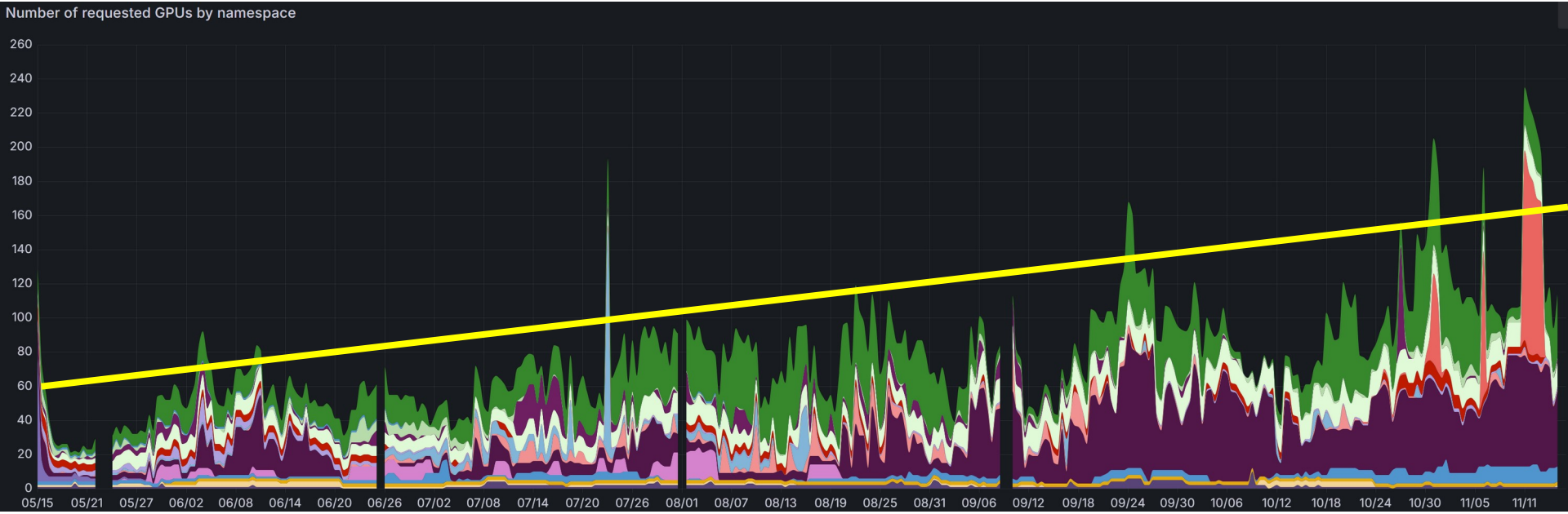
Nikolay Atanasov
*Dinesh Bharadia
Pamela Cosman
Sujit Dey
Xiaolong Wang
*Pengtao Xie

* HDSI Faculty Affiliate

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

UCSD HDSI Facuty Have Tripled Their Usage to ~180 Nautilus GPUs Per Day



Last 6 Months:
5/15/2024 to
11/15/2024

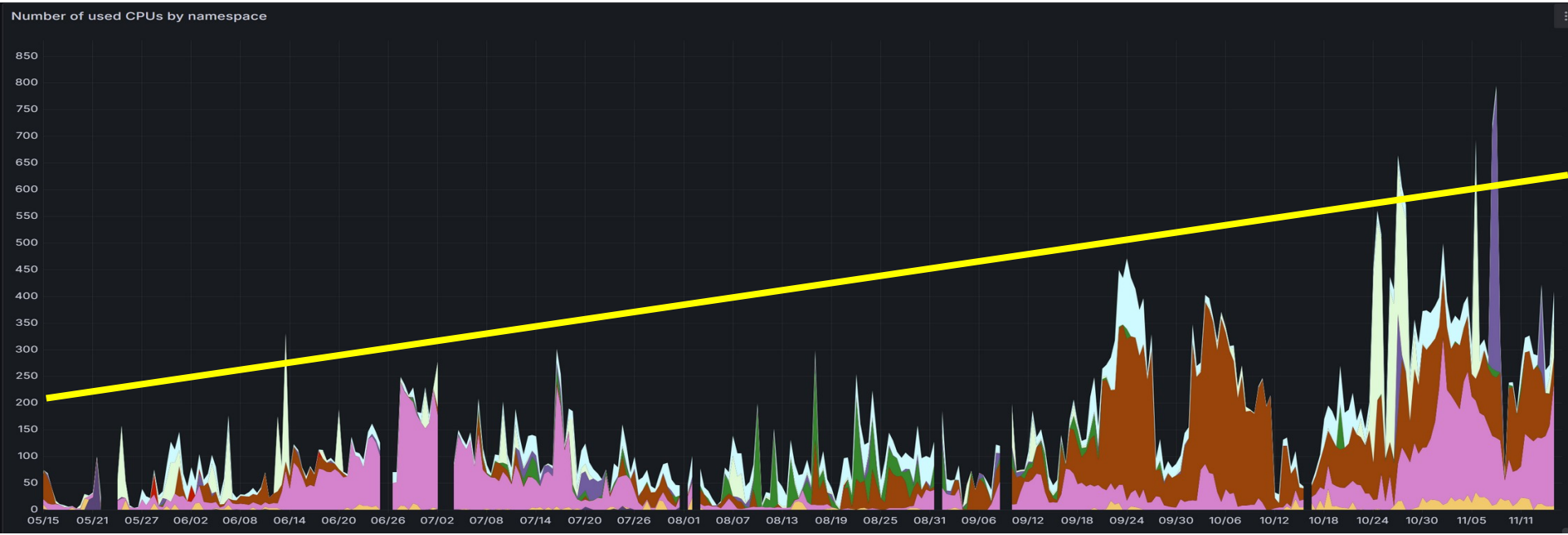
Ilkay Altintas
Mikio Aoi
Vineet Bafna+
Henrik Christensen+
Alex Cloninger
Rob Knight+
Arya Mazumdar
Tauhidur Rahman
Duncan Watson-Parris
Frank Wuerthwein
Qi (Rose) Yu+

+ CSE Faculty

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

UCSD HDSI Faculty Have Also Roughly Tripled Their Usage to ~600 Nautilus CPU Cores Per Day



Ilkay Altintas
Henrik Christensen+
Arya Mazumdar

Tauhidur Rahman
Ben Smarr
Frank Wuerthwein
Qi (Rose) Yu+

+ CSE Faculty

Also Running on Nautilus:
Tara Javidi
Bradley Voytek
Alex Warstadt
Lily Weng

NRP NATIONAL RESEARCH PLATFORM

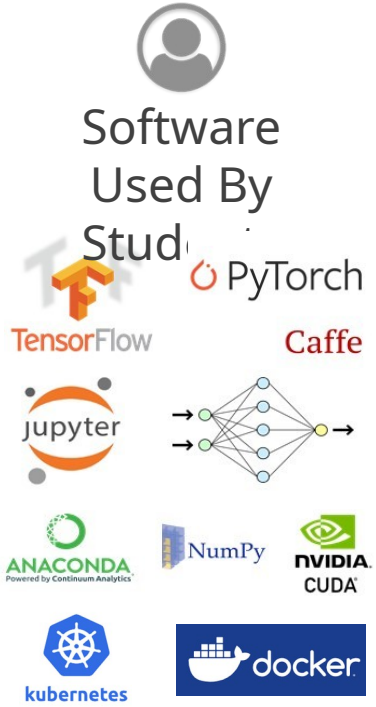
NRP Nautilus

UCSD's Information Technology Services Has Adapted NRP FIONA8s To Support Students in Data Science (DS) & Machine Learning (ML) Courses



SDSC Racked FIONAs:

- 132 32-bit GPUs
- 1024 CPU-cores
- 10/100G Networking
- Currently Not Federated With NRP



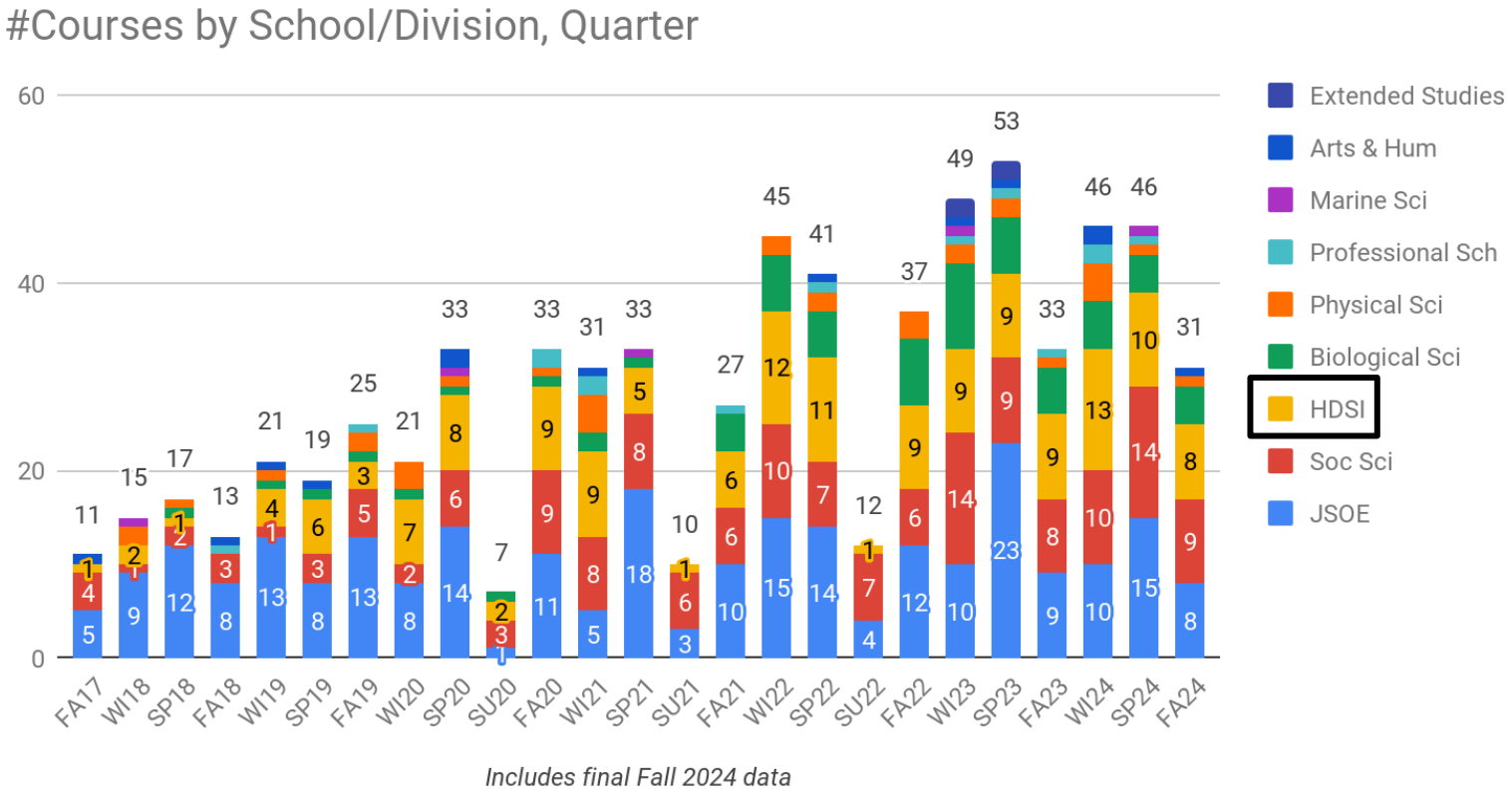
- **Student-Focused Platform For:**
  - Undergraduate & Graduate Coursework
  - For-Credit Independent Study
  - Thesis/Dissertation Research
  - Capstones & Projects
- **Research-Driven Architecture**
- **Managed by UCSD IT Services**
- **Used in Over 60 Courses**
- **By More Than 6000 Students**

Source: Adam Tilghman, UCSD ITS

NRP NATIONAL RESEARCH PLATFORM

IT SERVICES
UC San Diego
NRP Nautilus

DSMLP Courses by School/Division, Term



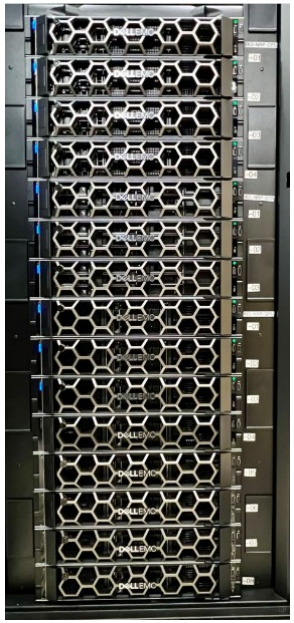
Includes final Fall 2024 data

HDSI Courses
Using ITS DSMLP FA22 to WI24

DSC 10	Principles of Data Science	7	2623	Rampure, Suraj (SP23); Tiefenbruck, Janine LoBue (FA22,FA23,FA24,SP24,WI23,WI24)
DSC 20	Prgrmng/DataStruc for Data Sci	1	178	Langlois, Marina (SP24)
DSC 40A	Theor Fndtns of Data Sci I	6	1049	Li, Aobo (WI24); Mishne, Gal (FA24,WI23); Rampure, Suraj (FA22,SP24); Tiefenbruck, Janine LoBue (SP23)
DSC 80	Practice of Data Science	5	711	Eldridge, Justin Matthew (FA22); Lau, Samuel Ethan (FA23,SP24); Rahman, Tauhidur (SP23); Rampure, Suraj (WI23)
DSC 96	Workshop in Data Science	3	91	Jemmott, Colin William (FA23); Quer, Giorgio (WI23,WI24)
DSC 102	Systems for Scalable Analytics	7	776	Albuyeh, Rod (SP23); Bellur, Umesh (WI24); Jin, Haojian (FA23,FA24,SP24); Kolovos, Alexandros (WI23); Kumar, Arun Kumar (FA22)
DSC 106	Intro to Data Visualization	1	199	Lau, Samuel Ethan (WI24)
DSC 120	Signal Processing/Data Analys	1	31	Mishne, Gal (FA23)
DSC 140A	Probabilistic Modeling and ML	1	248	Eldridge, Justin Matthew (WI24)
DSC 155 MATH 182	Hidden Data in Random Matrices	1	60	Dumitriu, Ioana (WI24)
DSC 170	Spatial Data Science and Appl	2	98	Zaslavsky, Ilya (WI23,WI24)
DSC 180A	Data Science Undergraduate Capstone I	3	676	Bellur, Umesh (FA24); Rampure, Suraj (FA22,FA23)
DSC 180B	Data Science Undergraduate Capstone II	1	235	Rampure, Suraj (WI24)
DSC 190	Topics in Data Science	4	60	Mishne, Gal (FA22); Silberman, Jack (FA22,WI23,WI24)
DSC 190 DSC 291	Topics in Data Science	1	71	Ustun, Berk (SP24)
DSC 200	Data Science Programming	1	34	Liao, Soohyun Nam (FA22)
DSC 204A	Scalable Data Systems	2	126	Jin, Haojian (SP23); Zhang, Hao (WI24)
DSC 205	Geometry of Data	1	17	Mishne, Gal (FA23)
DSC 208R	Data Management for Analytics	2	89	Salimi, Babak (WI23,WI24)
DSC 242	High-Dimensional Prob/Stats	1	33	Bradic, Jelena (FA24)
DSC 245	Intro to Causal Inference	3	180	Huang, Biwei (FA24,SP23,SP24)
DSC 256R	Data Mining on the Web	2	80	McAuley, Julian John (FA23,FA24)
DSC 261	Responsible Data Science	1	38	Salimi, Babak (WI24)
DSC 291	Topics in Data Science	3	131	Hu, Zhiting (SP24,WI23); Mishne, Gal (FA22)

SDSU's VERNE Expands UCSD's Instructional Model by Federating with NRP: Visionary, Education, Research, Network, Ecosystem

SDSU Information Technology Division



CENIC

NVIDIA. docker kubernetes jupyter PyTorch TensorFlow MATLAB RStudio

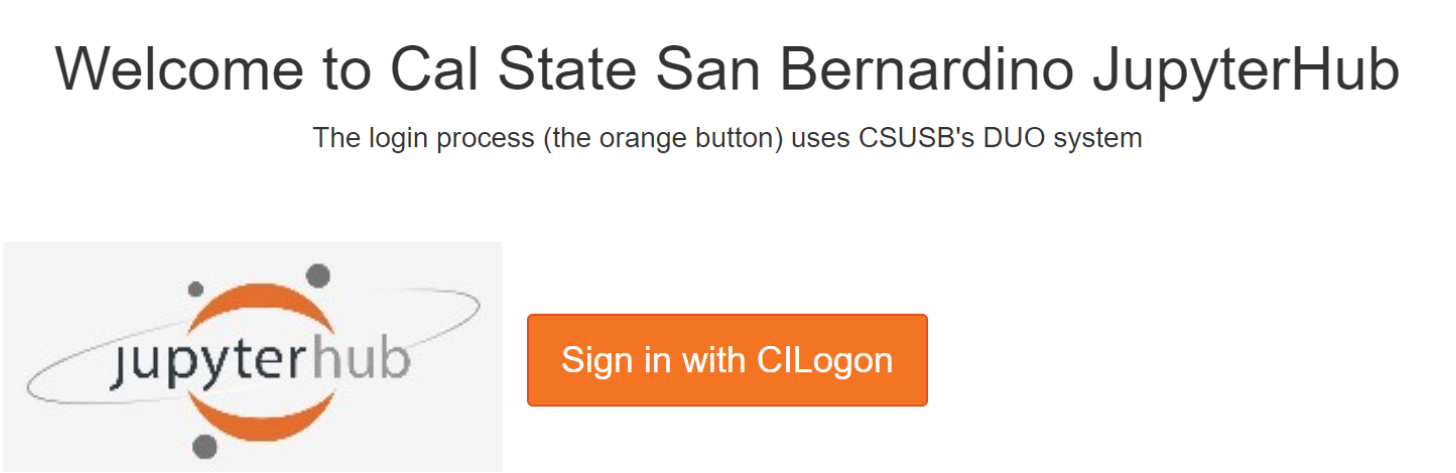
- **Dell PowerEdge Cluster for Instructional Use:**
  - 15 Nodes with:
    - 32 A100 GPUs (can bemanaged as 232 independent GPU
    - 768 CPU cores
    - 240 TB Persistent Storage
  - A Learning Resource that Surges for Research and Instruction
  - System Administration as a Service from the NRP
- **Growth in Both SDSU Courses and Students Using VERNE**
  - Spring Quarter 2024
    - 14 Courses
    - 300 Students

Slide courtesy of Jerry Sheehan and Mike Farley

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

A Key Reason CSUSB Has The Largest CSU Nautilus Usage: They Installed and Publicized the JupyterHub “Easy Button”



https://csusb-jupyter.nrp-nautilus.io/hub/login

Slide Adapted from Prof. Youngsu Kim

Department of Mathematics Faculty & Staff

Statistics From Dung Vu

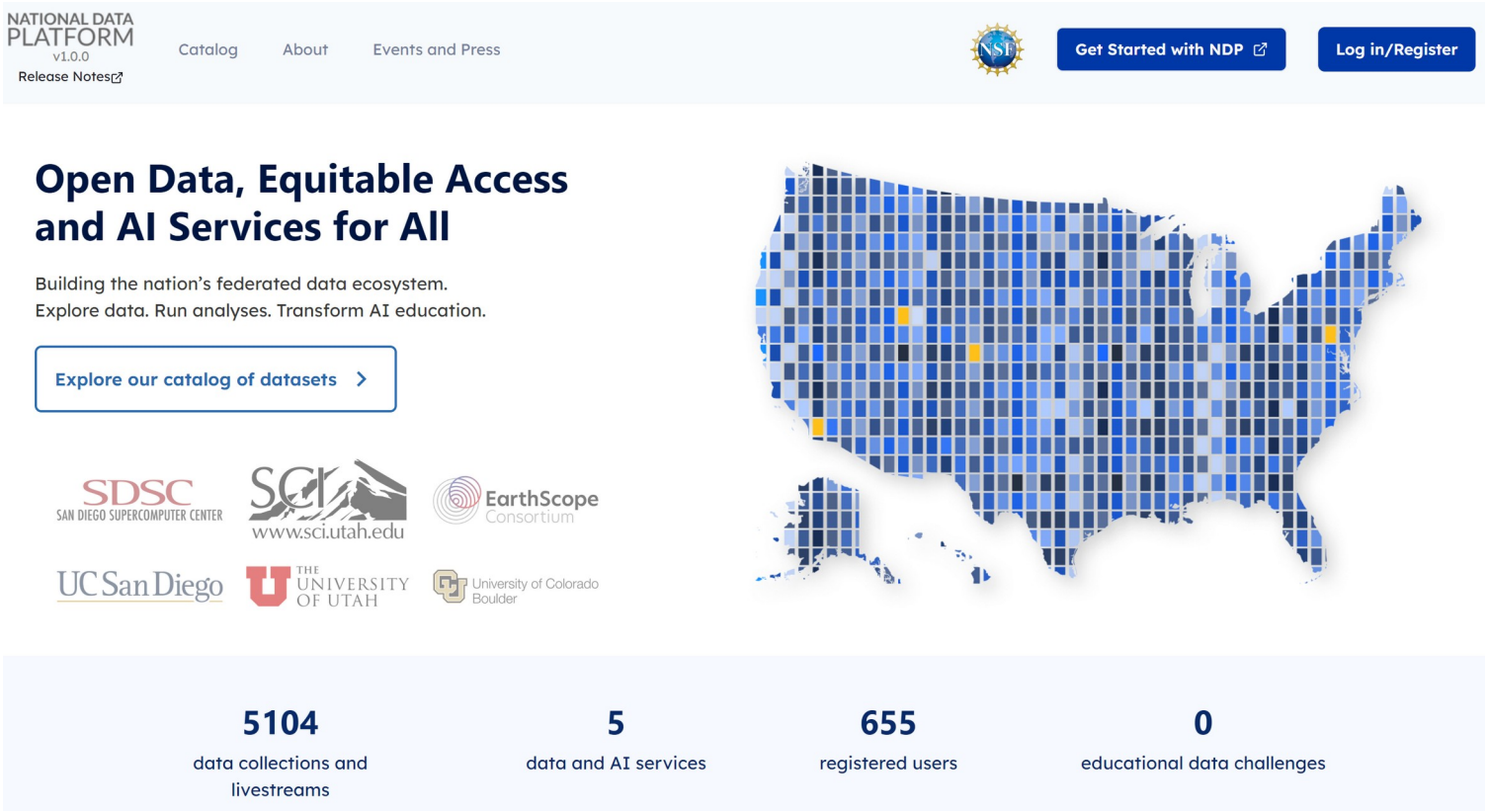
Over 650 Total Users
Tripled in the Last Six Months!

NRP NATIONAL RESEARCH PLATFORM

CSUSB WE DEFINE THE Future

NRP Nautilus

The National Data Platform (Ilkay Altintas, PI) is Being Integrated with the NRP



https://www.nationaldataplatform.org/

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

Attend the 6NRP to Learn More

Sixth National Research Platform (6NRP) Workshop

Tuesday, January 28, 2025 - Thursday, January 30, 2025

NRP NATIONAL RESEARCH PLATFORM

www.sdsc.edu/event_items/202501_6nrp.html

NRP Nautilus