“The CENIC AI Resource CENIC AIR”

CENIC Retreat 2024:
July 16, 2024

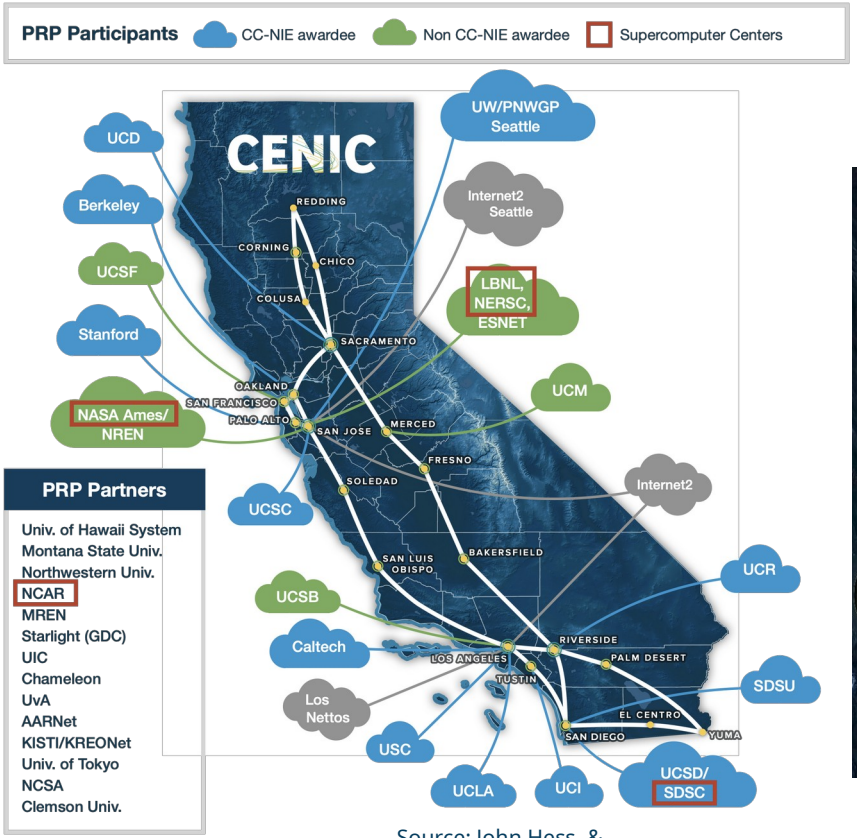
Dr. Larry Smarr
Founding Director Emeritus, California Institute for Telecommunications and Information Technology;
Distinguished Professor Emeritus, Dept. of Computer Science and Engineering
Jacobs School of Engineering, UC San Diego

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

2015 Vision: The Pacific Research Platform Will Connect Research Campuses Using CENIC to Create a Regional Community Cyberinfrastructure



Source: John Hess, & Hunter Hadaway, CENIC

NSF CC*DNI Grant
$6.3M 10/2015-10/2020
Extended – Ended Year 7 in Oct 2022



NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

Data Transfer Nodes (DTNs) Terminate CENIC's Fiber Optics on Campuses: 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



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



FIONAs Are Rack Mounted

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

NSF

QUALCOMM INSTITUTE

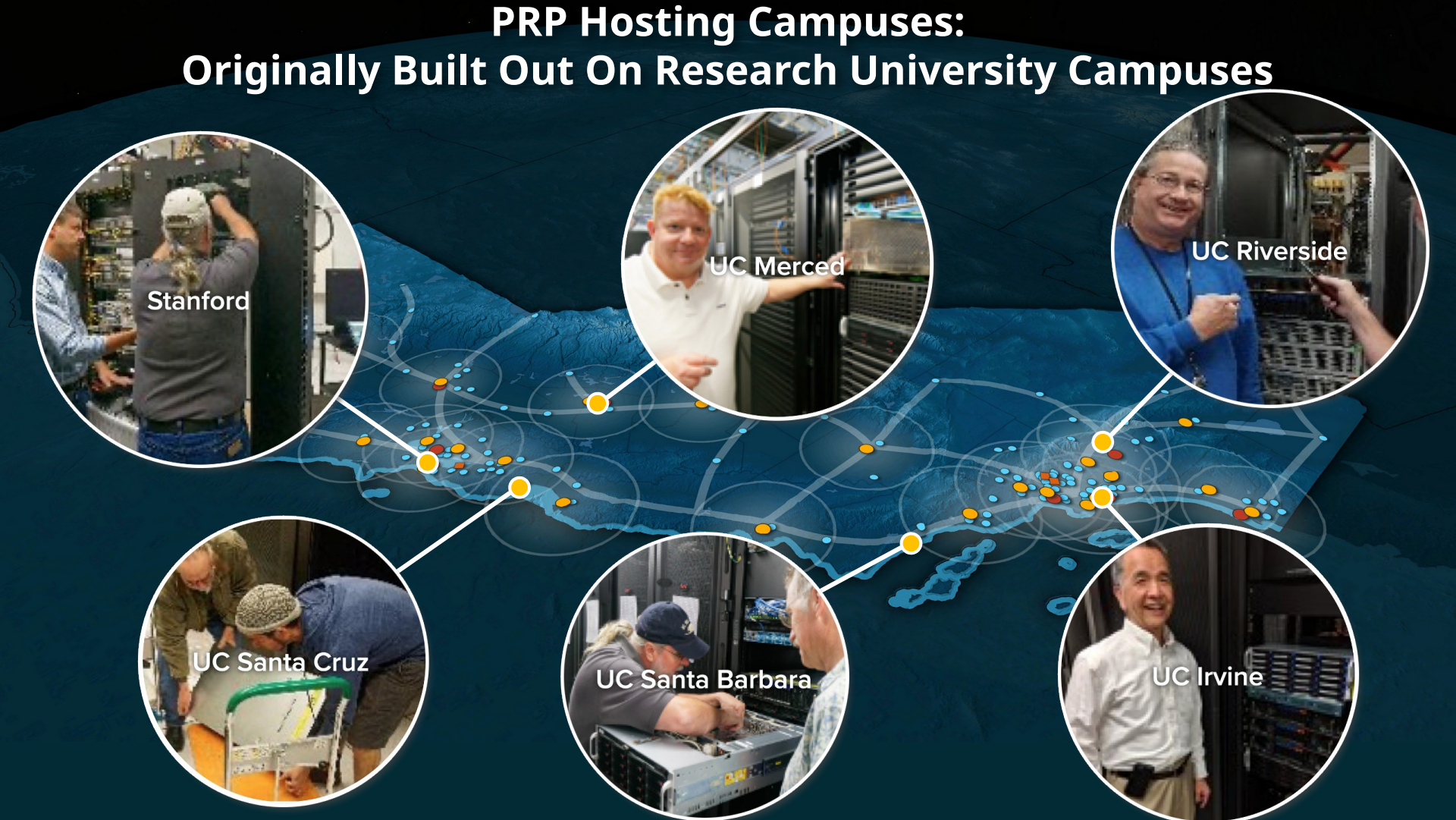
SDSC

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

PRP Hosting Campuses:
Originally Built Out On Research University Campuses



2023 - The National Research Platform Emerges From the Pacific Research Platform

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

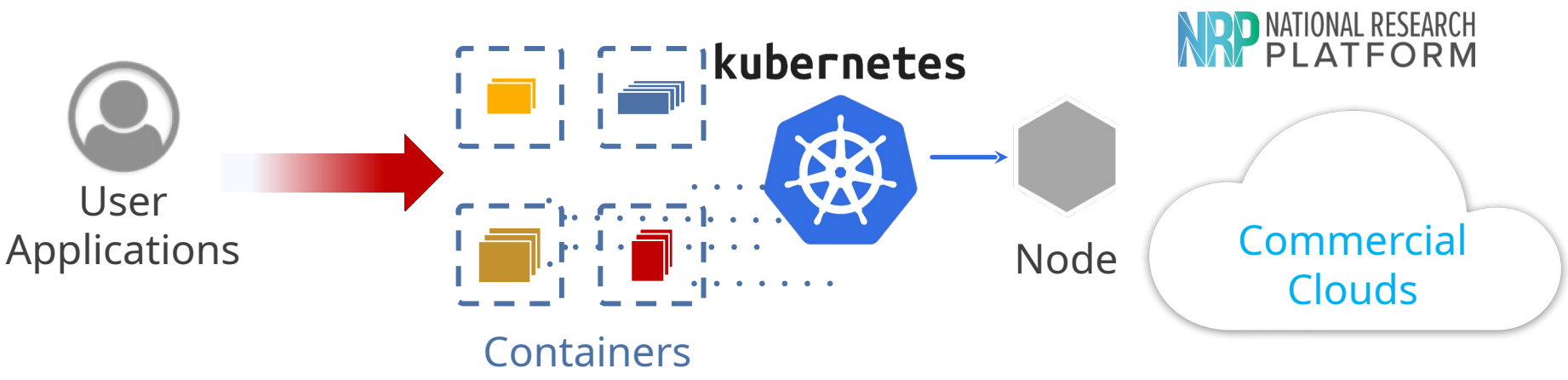
The National Research Platform (NRP) is a partnership of more than 50 institutions, led by researchers and cyberinfrastructure professionals at UC San Diego, supported in part by awards from the National Science Foundation.

https://nationalresearchplatform.org/

Users Can Execute Their Containerized Applications in the NRP or in Commercial Clouds

Nautilus Containerized Applications Are “Cloud Ready”

Nautilus



NRP NATIONAL RESEARCH PLATFORM

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

May 9, 2024

Nautilus NSF

~200 FIONAs on 27 Partner Campuses
Networked Together at 10-100Gbps

Installed GPUs	Installed CPU Cores	Rotating Storage
1210	23855	7000 TB

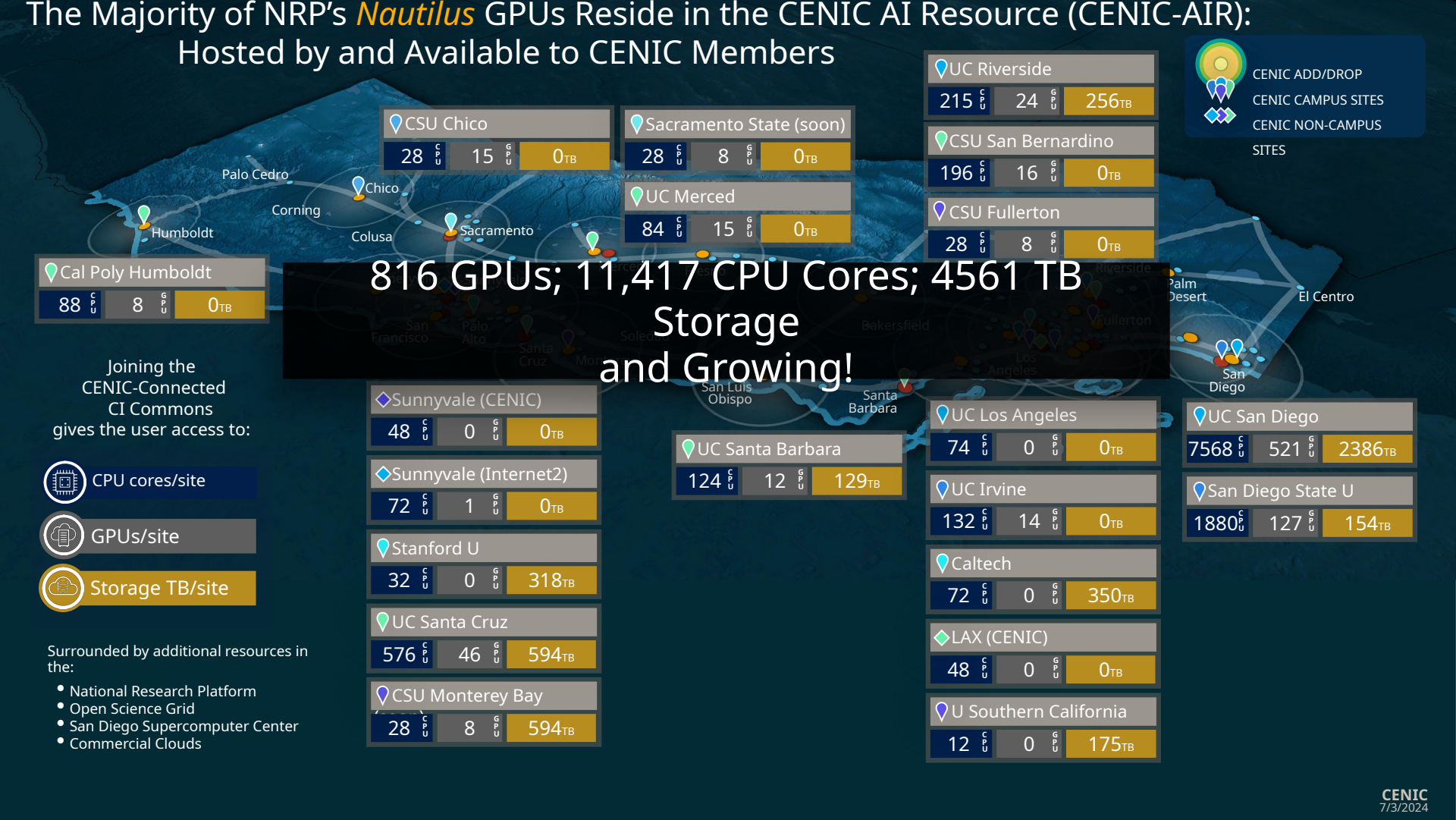
CENIC THE QUILT INTERNET2

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

The Majority of NRP's *Nautilus* GPUs Reside in the CENIC AI Resource (CENIC-AIR): Hosted by and Available to CENIC Members



CSU Chico: 28 CPU | 15 GPU | 0TB

Sacramento State (soon): 28 CPU | 8 GPU | 0TB

UC Merced: 84 CPU | 15 GPU | 0TB

UC Riverside: 215 CPU | 24 GPU | 256TB

CSU San Bernardino: 196 CPU | 16 GPU | 0TB

CSU Fullerton: 28 CPU | 8 GPU | 0TB

Cal Poly Humboldt: 88 CPU | 8 GPU | 0TB

816 GPUs; 11,417 CPU Cores; 4561 TB Storage and Growing!

Joining the CENIC-Connected CI Commons gives the user access to:

- CPU cores/site
- GPUs/site
- Storage TB/site

Sunnyvale (CENIC): 48 CPU | 0 GPU | 0TB

Sunnyvale (Internet2): 72 CPU | 1 GPU | 0TB

Stanford U: 32 CPU | 0 GPU | 318TB

UC Santa Cruz: 576 CPU | 46 GPU | 594TB

CSU Monterey Bay: 28 CPU | 8 GPU | 594TB

UC Santa Barbara: 124 CPU | 12 GPU | 129TB

UC Los Angeles: 74 CPU | 0 GPU | 0TB

UC Irvine: 132 CPU | 14 GPU | 0TB

Caltech: 72 CPU | 0 GPU | 350TB

LAX (CENIC): 48 CPU | 0 GPU | 0TB

U Southern California: 12 CPU | 0 GPU | 175TB

UC San Diego: 7568 CPU | 521 GPU | 2386TB

San Diego State U: 1880 CPU | 127 GPU | 154TB

Surrounded by additional resources in the:

- National Research Platform
- Open Science Grid
- San Diego Supercomputer Center
- Commercial Clouds

CENIC
7/3/2024

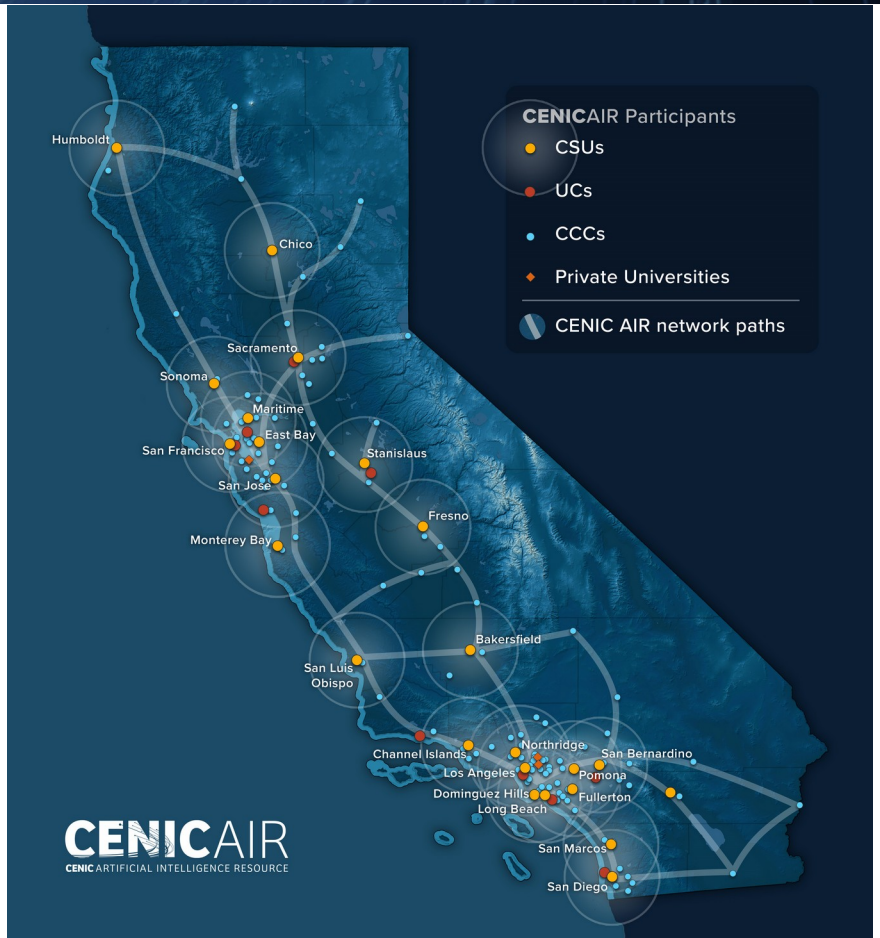
Accelerating the Development of California's AI/ML Workforce By Using CENIC-AIR for Campus Courses

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

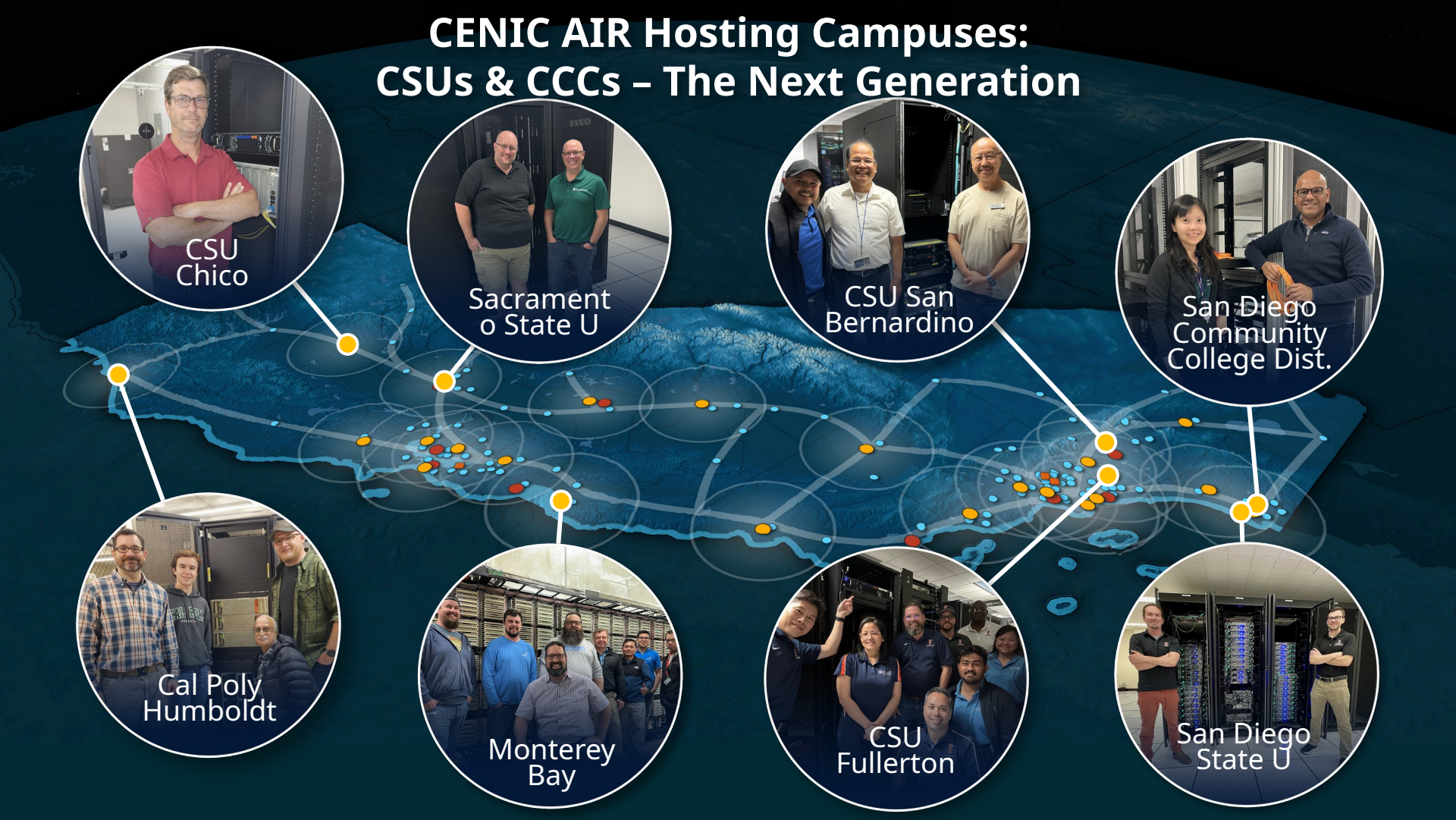
California's Research & Education Campuses are Regionally Organized



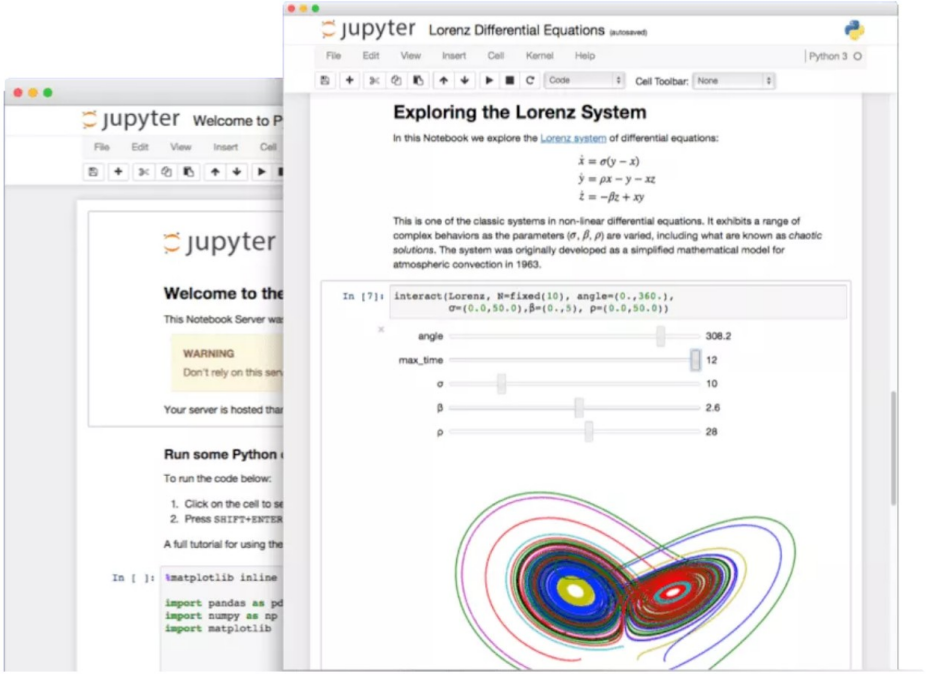
NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

CENIC AIR Hosting Campuses:
CSUs & CCCs – The Next Generation



Jupyter Notebooks
Have Become The Popular Method of Sharing Computational Documents



Jupyter Notebook: The Classic Notebook Interface

The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

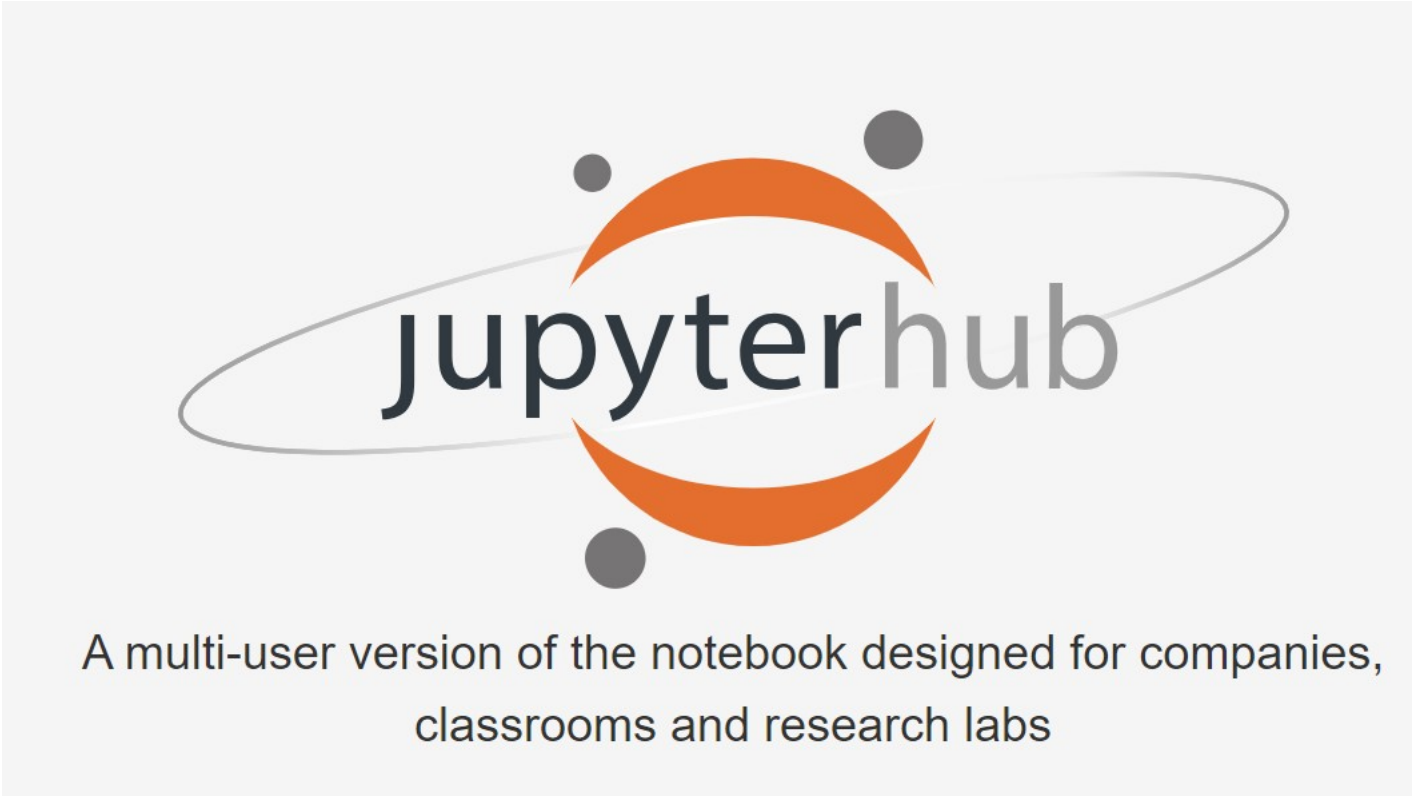
Try it in your browser | Install the Notebook

NRP NATIONAL RESEARCH PLATFORM

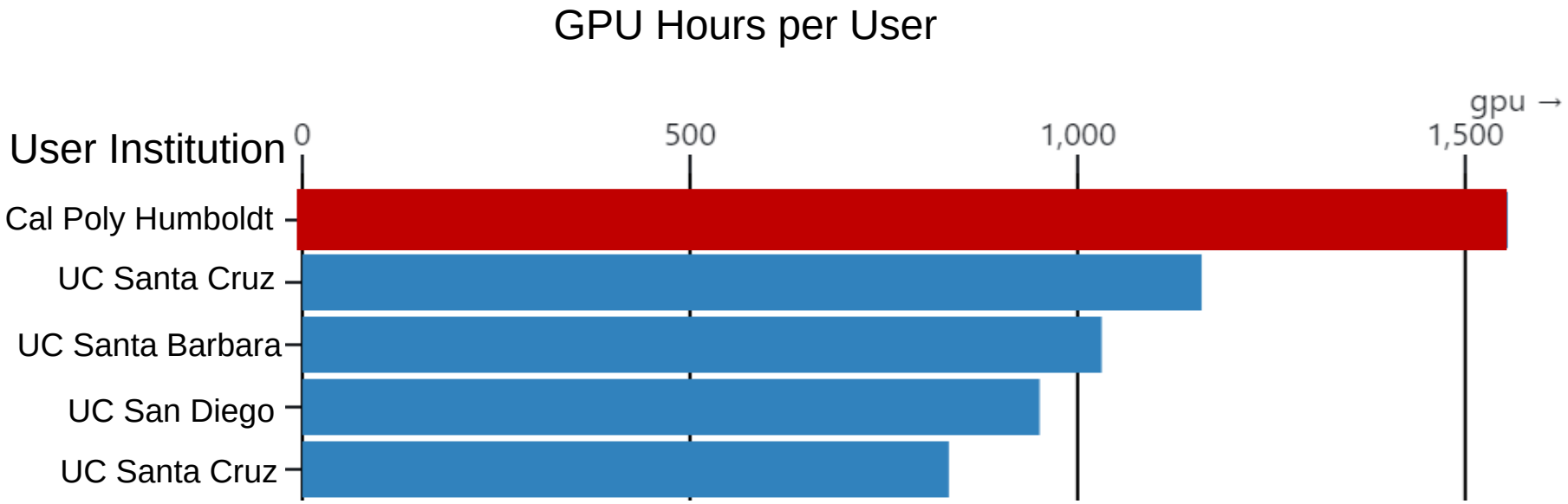
CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

jupyterhub is the Multi-User Version of the Jupyter Notebook



Nautilus Namespace Jupyterlab Has an Active jupyterhub: Jupyterlab's Top 5 GPU-Hour Users Over Last 4 Months



Users Are Identified Only by Email Addresses

Tom DeFanti Emailed the Largest GPU User of Namespace Jupyterlab, Asking "Who Are You and What is Your Research Project?"

We Received This Reply:

– **"Wow! Who'd have thought a botanist would be using the most GPU resources 😂"**

- **Alana Chin**
- **Assistant Professor of Plant Physiology**
- **Cal Poly Humboldt**



I am running some big Fully Convolutional Network (FCN) models to segment some of the highest resolution CT scans ever made on cores of wood collected from redwood trunks across their full height, up to 100 m above ground.

We are segmenting out tissue types for 3D measurements and brightness histograms to understand wood density and hydraulic parameters."

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 (10% NRP)
- 1024 CPU-cores
- 10/100G Networking
- Not Federated With NRP PUATFORM

Software Used By Stud

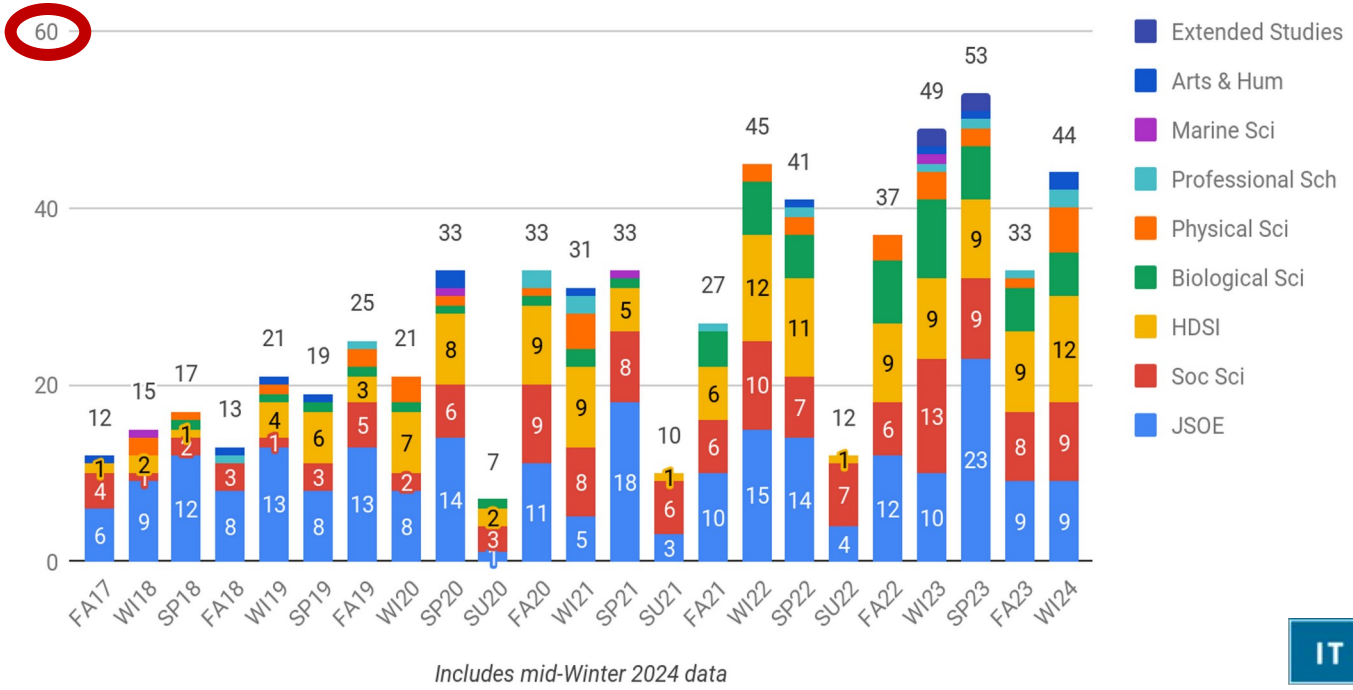
- **Student-Focused Platform For:**
  - Undergraduate & Graduate Coursework
  - For-Credit Independent Study
  - Thesis/Dissertation Research
  - Capstones & Projects
- **Research-Driven Architecture**
- **Managed by UCSD IT Services**

Source: Adam Tilghman, UCSD ITS
Data Science/ML Platform - 2/2024 Update

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

IT SERVICES
UC San Diego
NRP Nautilus

UC San Diego's DS/ML Platform Has Supported Up To Nearly 60 Courses Over the Last Six Years



Includes mid-Winter 2024 data

Source: Adam Tilghman, UCSD ITS
Data Science/ML Platform - 2/2024 Update

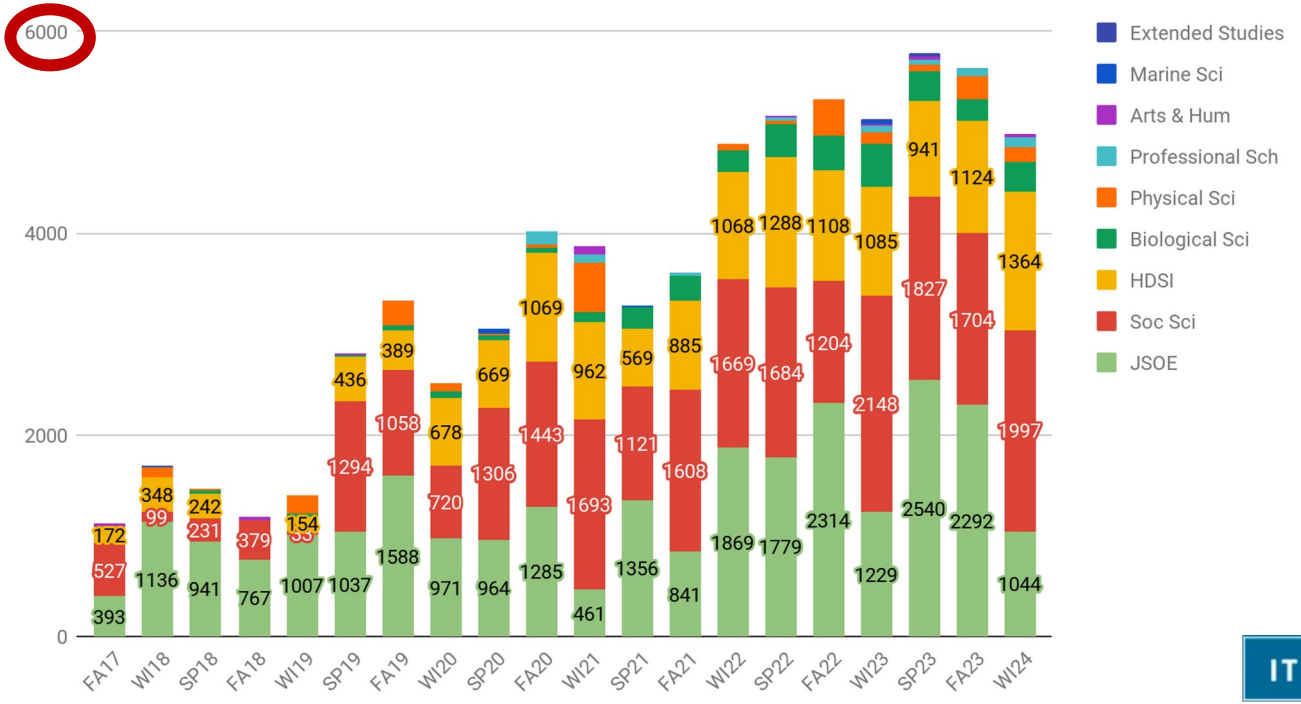
IT SERVICES
UC San Diego

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

UC San Diego's DS/ML Platform Has Supported Up To 6,000 Students Per Quarter Over the Last Six Years



Source: Adam Tilghman, UCSD ITS
Data Science/ML Platform - 2/2024 Update

IT SERVICES
UC San Diego

NRP NATIONAL RESEARCH PLATFORM

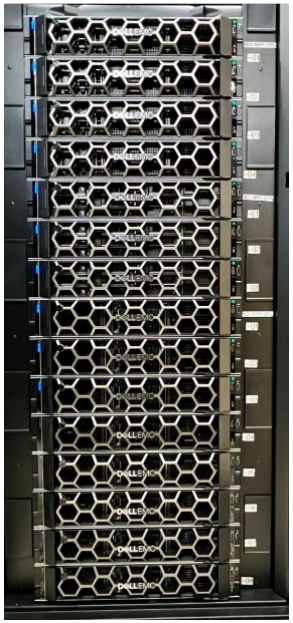
CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

Education and Workforce Development Using CENIC AIR: San Diego State University

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

CENIC AIR CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

SDSU AI survey



SDSU launched an AI student survey in fall 2023

Understand students' perception and use of AI

Inform institutional response and policymaking on AI



7,800+ student responses (21% response rate)

Largest known survey on AI in higher education

More responses than nationwide in Sweden (~5,900) and Australia (~1,100)



Representative of the SDSU* student body

Survey results largely mirror university-wide statistics

Results represent all fields, not just computer science and engineering

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

SDSU AI survey findings

Usage

59%

Report that they use AI in some capacity

45%

Use ChatGPT

41%

Use Grammarly

Career development

"AI will play a significant role in my future career"

College	Agree (%)
Arts & Letters	48%
Business	72%
Education	44%
Engineering	72%
Health & Human Services	43%
Professional Studies & Fine Arts	52%
Sciences	62%
Overall	*57%*

Resources

52%

Expressed interest in receiving formal university training on AI

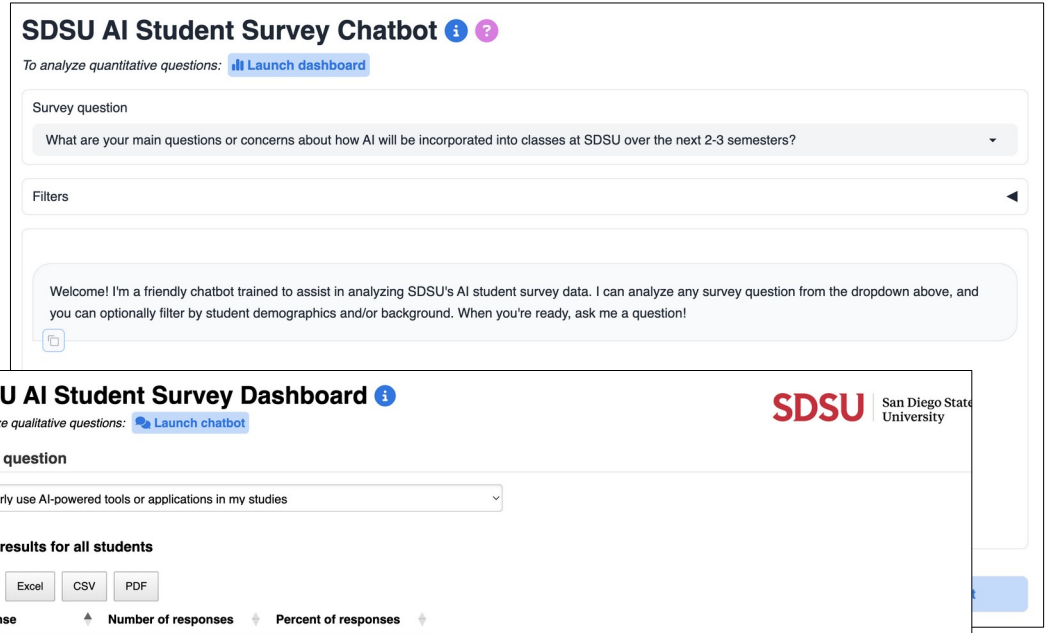
28%

Report that they are currently offered adequate AI training opportunities

SDSU AI survey findings

SDSU AI survey findings are available online through interactive dashboard and chatbot tools

aisurvey.sdsu.edu



NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

SDSU AI survey replications



- **SDSU's AI survey is being replicated across the world**
  - 17 known replication efforts
  - 5 known countries
- **The nexus of this effort is *regional***
  - In San Diego, CSU (SDSU), UC (UCSD), and CCC (SDCCD) are working together to replicate the survey in fall 2024
  - Additional efforts in California at CSU Chico, CSU Fullerton, Fresno State, San Francisco State, and UC San Francisco

San Diego Community College District
Collaborated with UCSD and SDSU

Can UCSD, SDSU, SDSC & NRP Work Together With SDCCD & SDCOE to Create a Common San Diego Instructional JupyterHub?



VERNE

SDSU

Launch JupyterHub



DSMLP

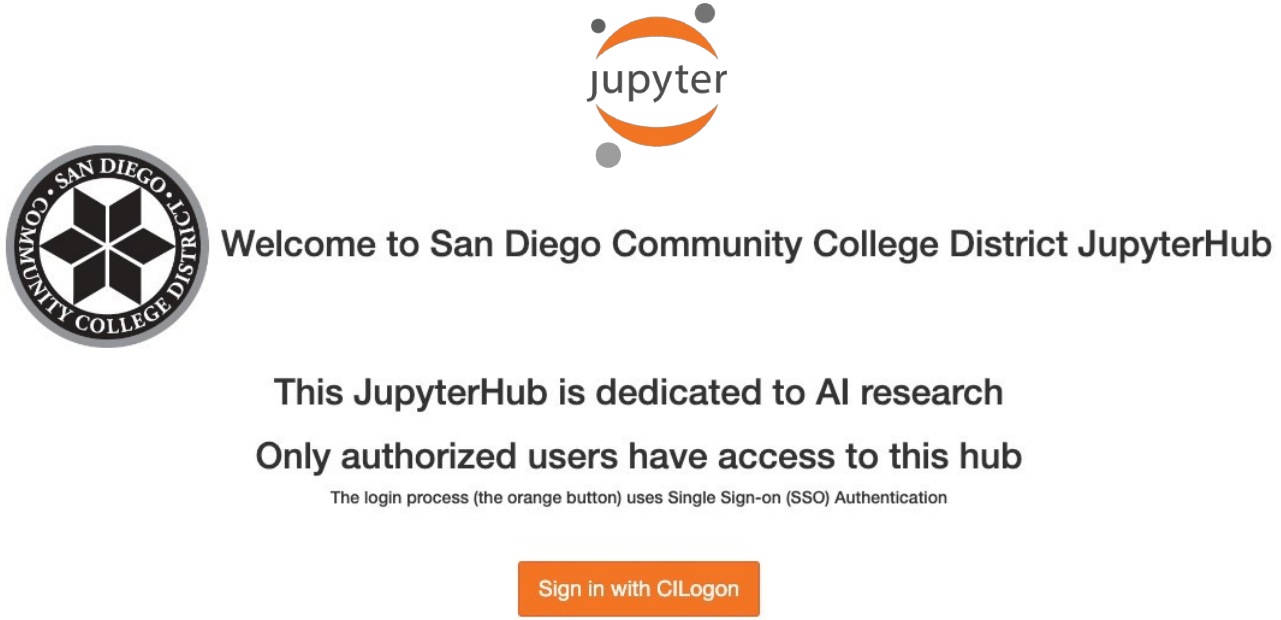
UC San Diego

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

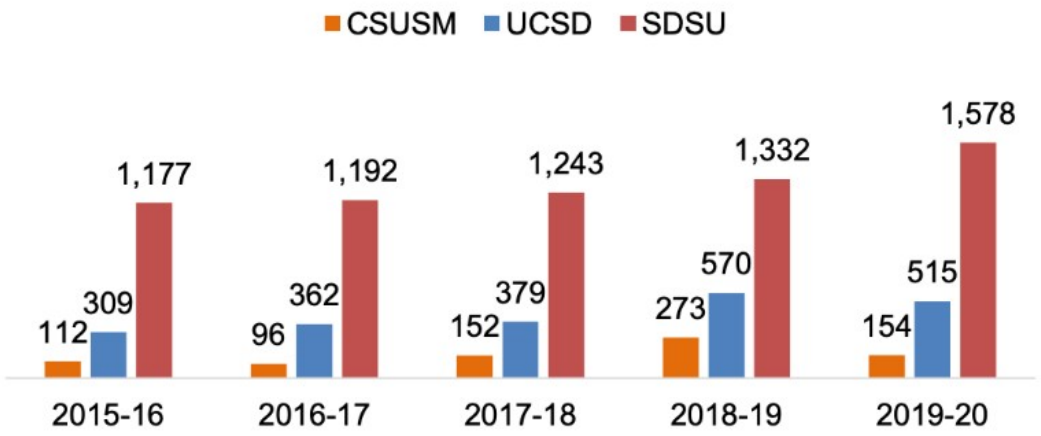
Potential JupyterHub
That Could be Created for the SDCCD



Concept and Image from Dung Vu, CSUSB

Toward a San Diego Region
UC/CSU/CCC Shared Use of CENIC-AIR

San Diego County Community College District (SDCCD)
Had 4,839 Students Transfer in 2019-20,
Almost Half (46%) to a Local Public University:
(SDSU: 33%, UCSD: 11%, CSUSM: 3%).



Local Public Universities Name Key

CSUSM: California State University, San Marcos

UCSD: University of California, San Diego

SDSU: San Diego State University

www.sdccd.edu/docs/Research/Student%20Outcomes/Transfer/Transfer%20Report_2019-20_Final_v2.pdf

Step One:
SDCCD Establishes InCommon Identity Management

San Diego Community College District — San Diego, California

Federation | Certificates | Eduroam

Organization

Service:	IdP
Org Name:	San Diego Community College District
Display Name:	San Diego Community College District
URL:	https://www.sdccd.edu/

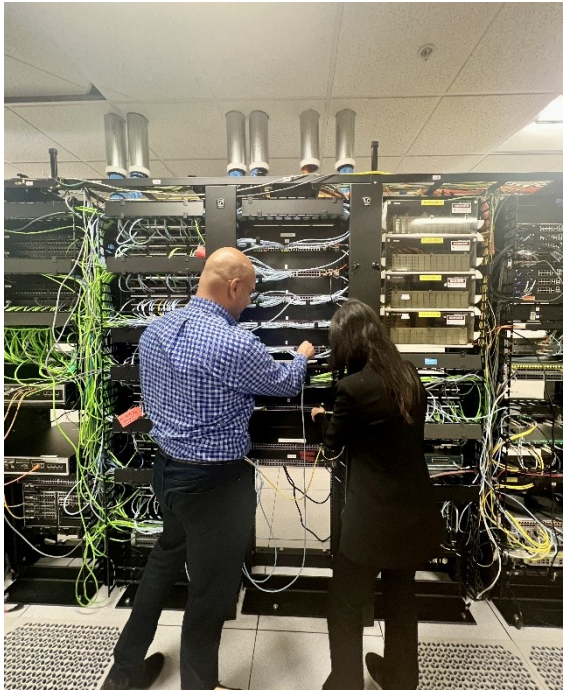
Federation

Identity Providers

https://sso.sdccd.edu/idp Info

SDCCD Can Host
CENIC AIR Computational Resources





Peter Maharaj, Rose Parnsoonthorn

NRP NATIONAL RESEARCH PLATFORM

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

SDCCD:
Next Steps for 2024

- **Create Automated Access for Users**
- **Pursue NSF Funding**
- **Host FIONAs and Upgrade CENIC Network**
- **Strengthen Collaboration With Our Partners: SDSU, UCSD, SDCOE, CCCCO, LACCD, Los Rios, CENIC**
- **Pursue Alignment of AI/ML/DS Curricula with SDSU and UCSD**