“The CENIC-AI Resource”

Los Angeles Community College District (LACCD)
Technology Policy and Planning Committee (TPPC)
December 5, 2024

Dr. Larry Smarr
Distinguished Professor Emeritus, Dept. of Computer Science and Engineering
Jacobs School of Engineering, UC San Diego

Website: https://lsmarr.net/

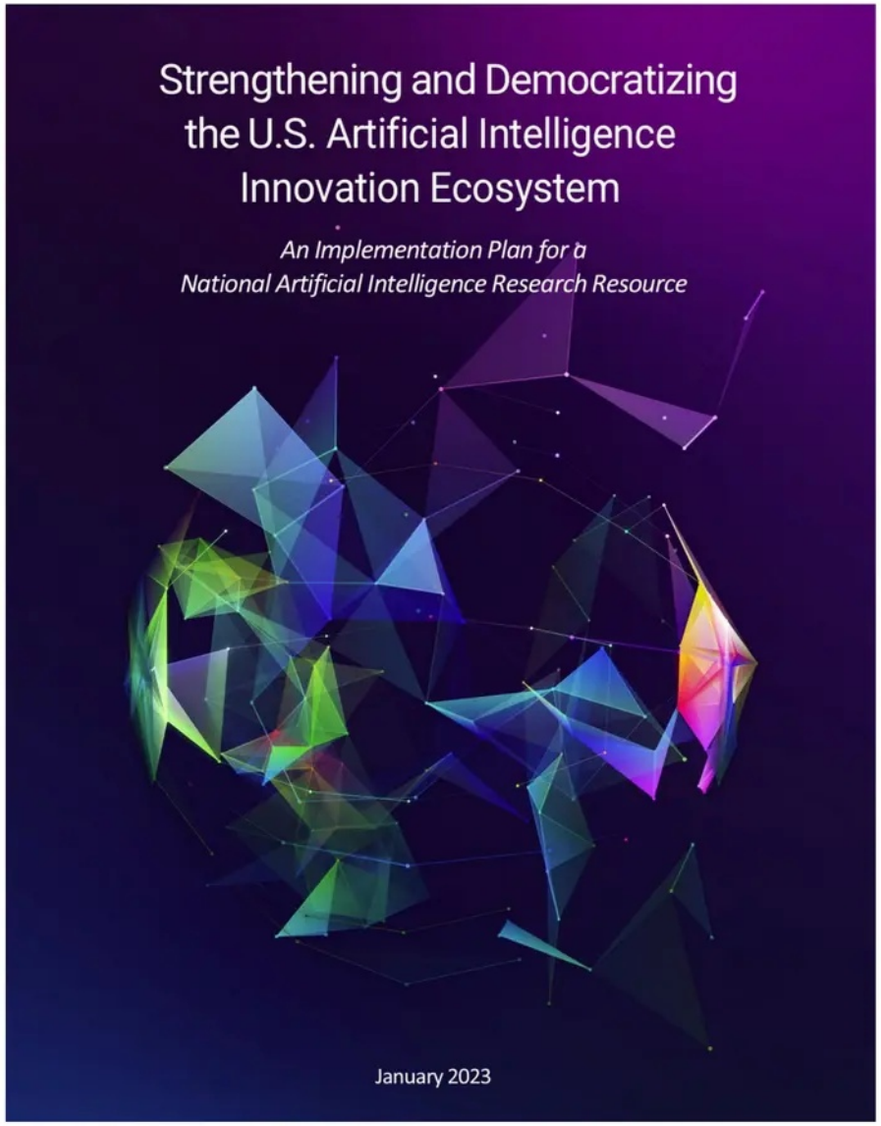
NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

The Congress Has Asked NSF to Develop a National AI Research Resource

A widely accessible AI research cyberinfrastructure that brings together computational resources, data, testbeds, algorithms, software, services, networks, and expertise, as described in this report would help to democratize the AI R&D landscape in the United States for the benefit of all.



In order to achieve its vision and goals, the Task Force estimates the budget for the NAIRR as $2.6 billion over an initial six-year period.

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

First Step: the NAIRR Pilot

U.S. National Science Foundation

NSF News

Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot

January 24, 2024

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

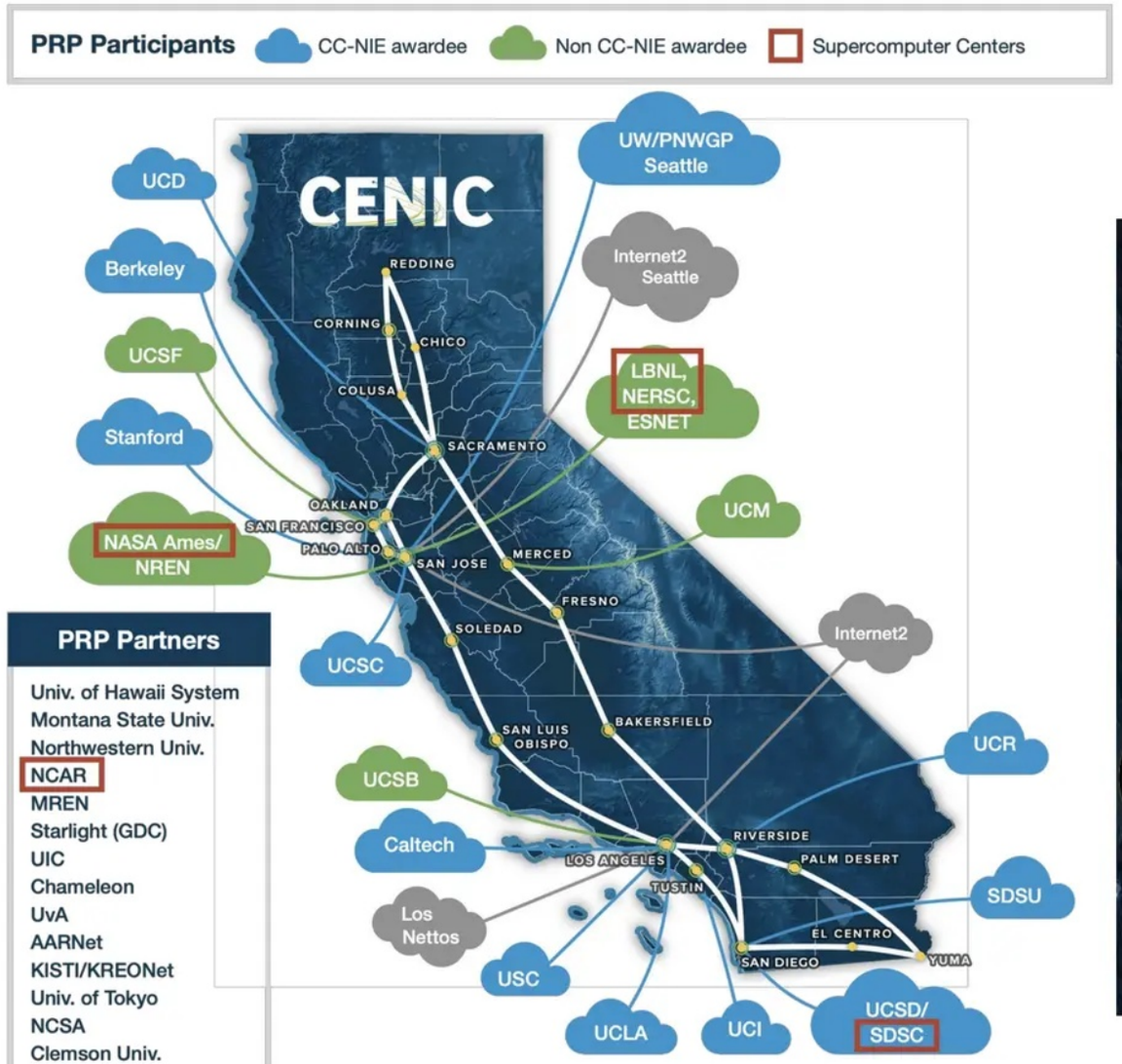
California is On The Path to Provide Leadership for NAIRR

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCE

NRP Nautilus

2015 Vision: The Pacific Research Platform Will Build on CENIC to Connect Science DMZs Creating 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

CENICAIR
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

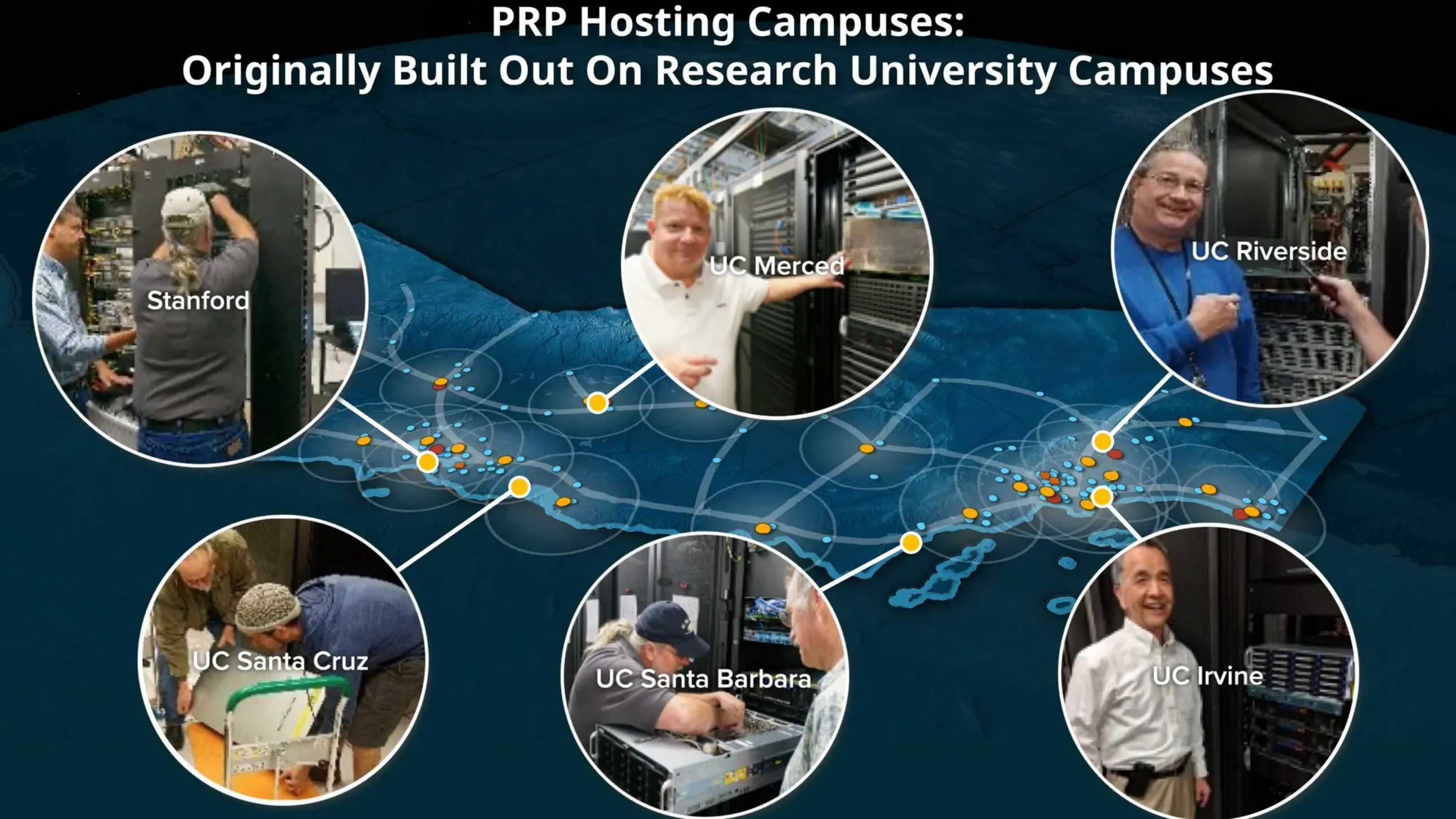
FIONAs Designed by UCSD's Phil Papadopoulos, John Graham, Joe Keefe, and Tom DeFanti

NRP NATIONAL RESEARCH PLATFORM

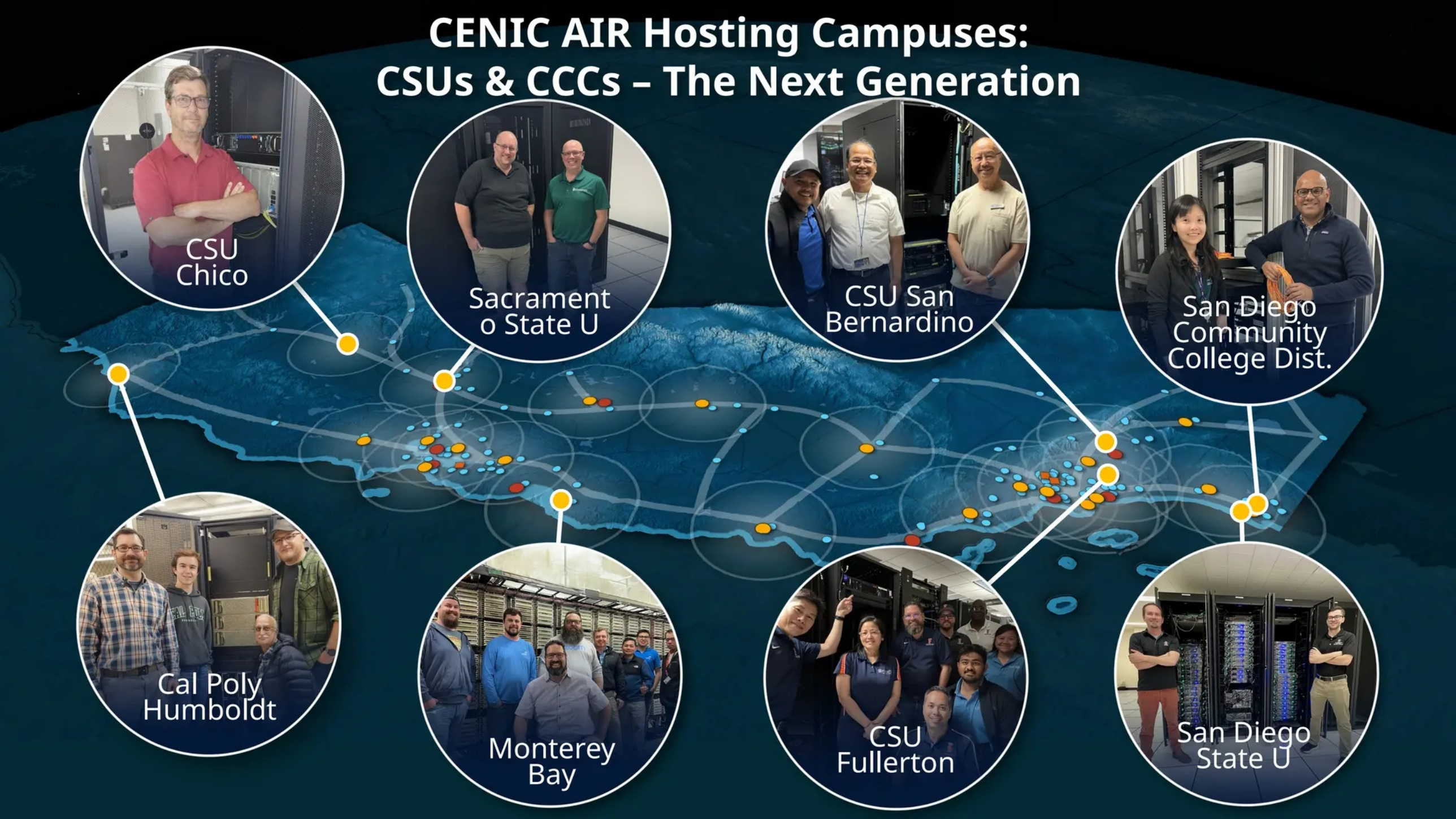
CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

PRP Hosting Campuses:
Originally Built Out On Research University Campuses



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



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/

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

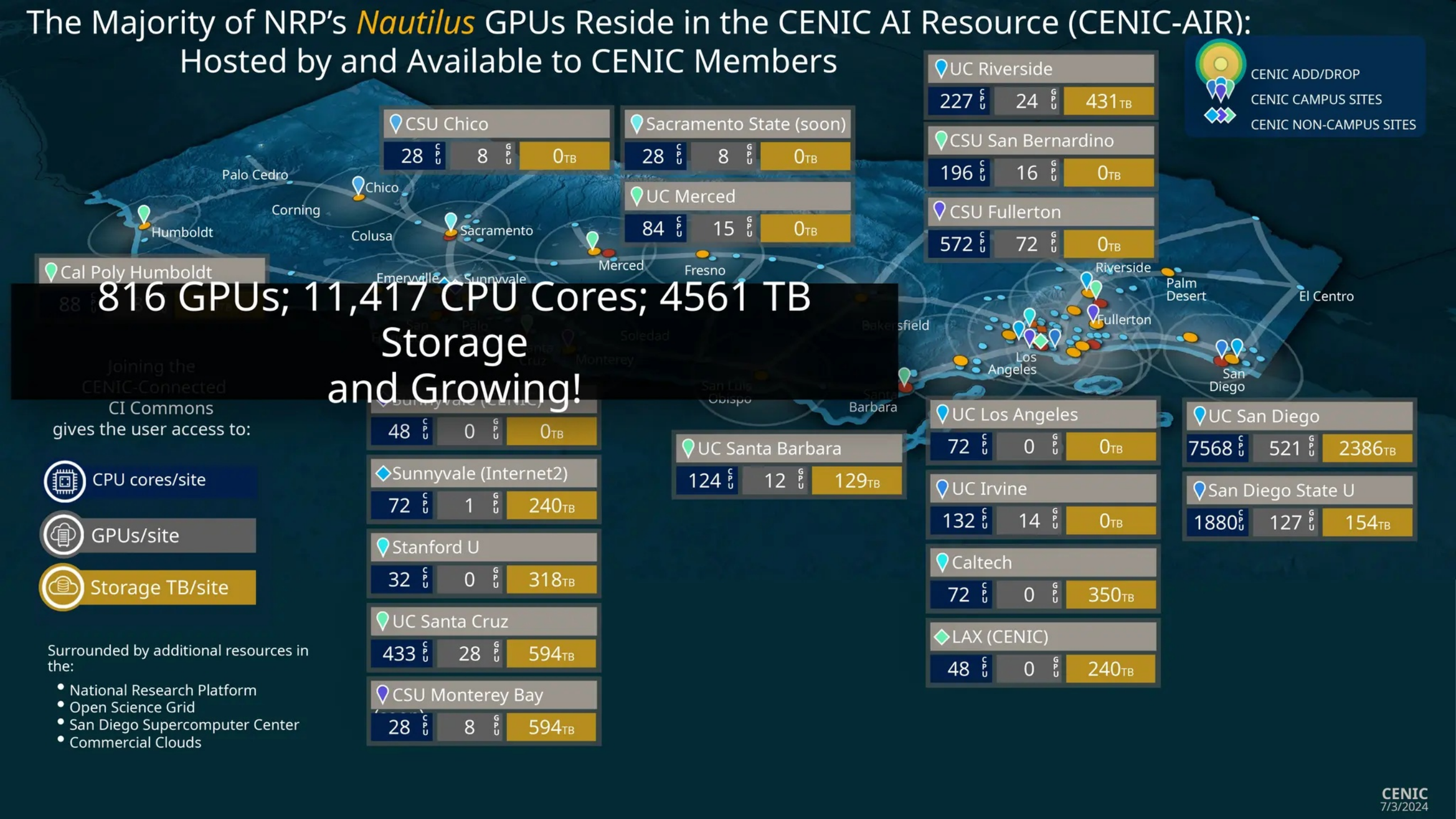
INTERNET2®

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

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



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

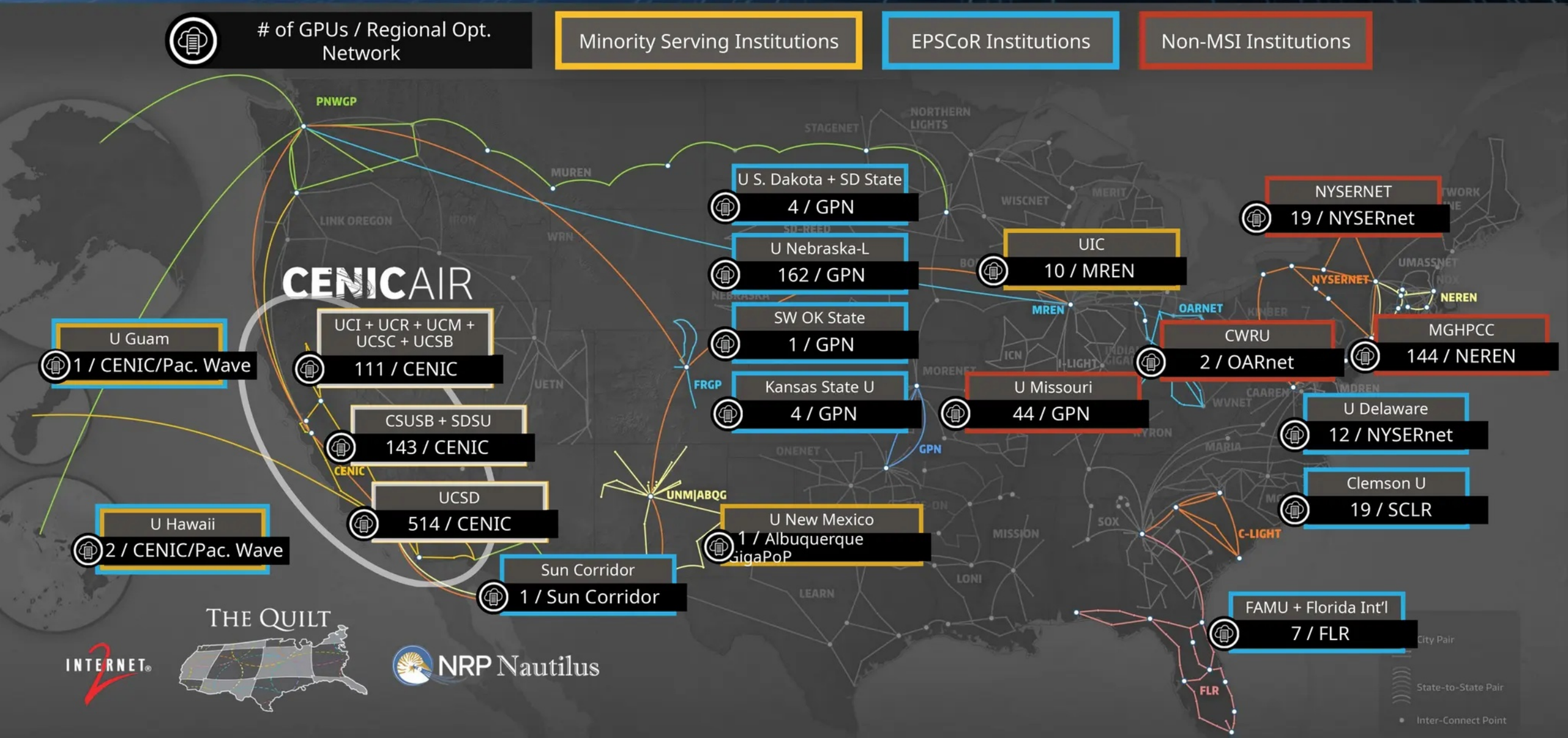
Surrounded by additional resources in the:

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

	CPU	GPU	Storage
Cal Poly Humboldt	88		
CSU Chico	28	8	0TB
Sacramento State (soon)	28	8	0TB
UC Merced	84	15	0TB
UC Riverside	227	24	431TB
CSU San Bernardino	196	16	0TB
CSU Fullerton	572	72	0TB
Sunnyvale (CENIC)	48	0	0TB
Sunnyvale (Internet2)	72	1	240TB
Stanford U	32	0	318TB
UC Santa Cruz	433	28	594TB
CSU Monterey Bay	28	8	594TB
UC Santa Barbara	124	12	129TB
UC Los Angeles	72	0	0TB
UC Irvine	132	14	0TB
Caltech	72	0	350TB
LAX (CENIC)	48	0	240TB
UC San Diego	7568	521	2386TB
San Diego State U	1880	127	154TB

CENIC
7/3/2024

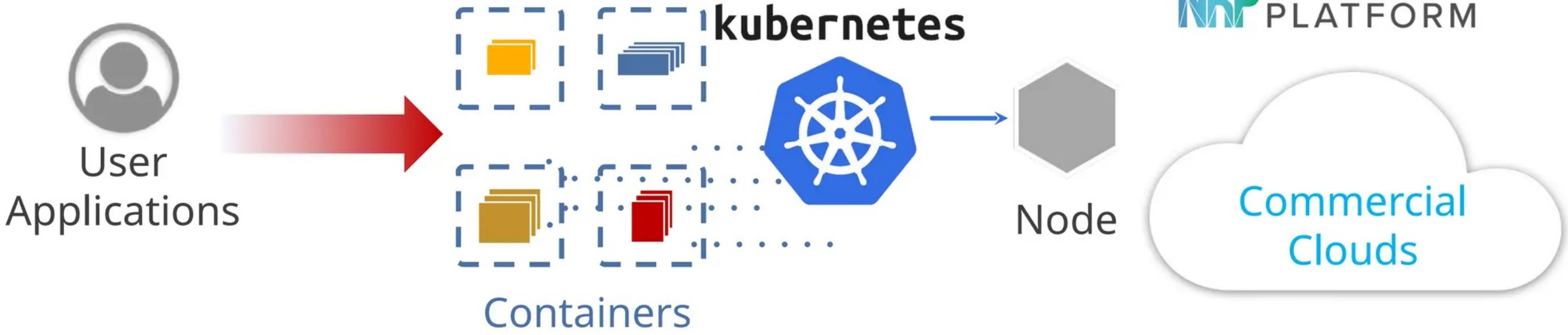
The Users of the CENIC-AIR Can Burst into NRP's Nautilus Hypercluster Outside of California



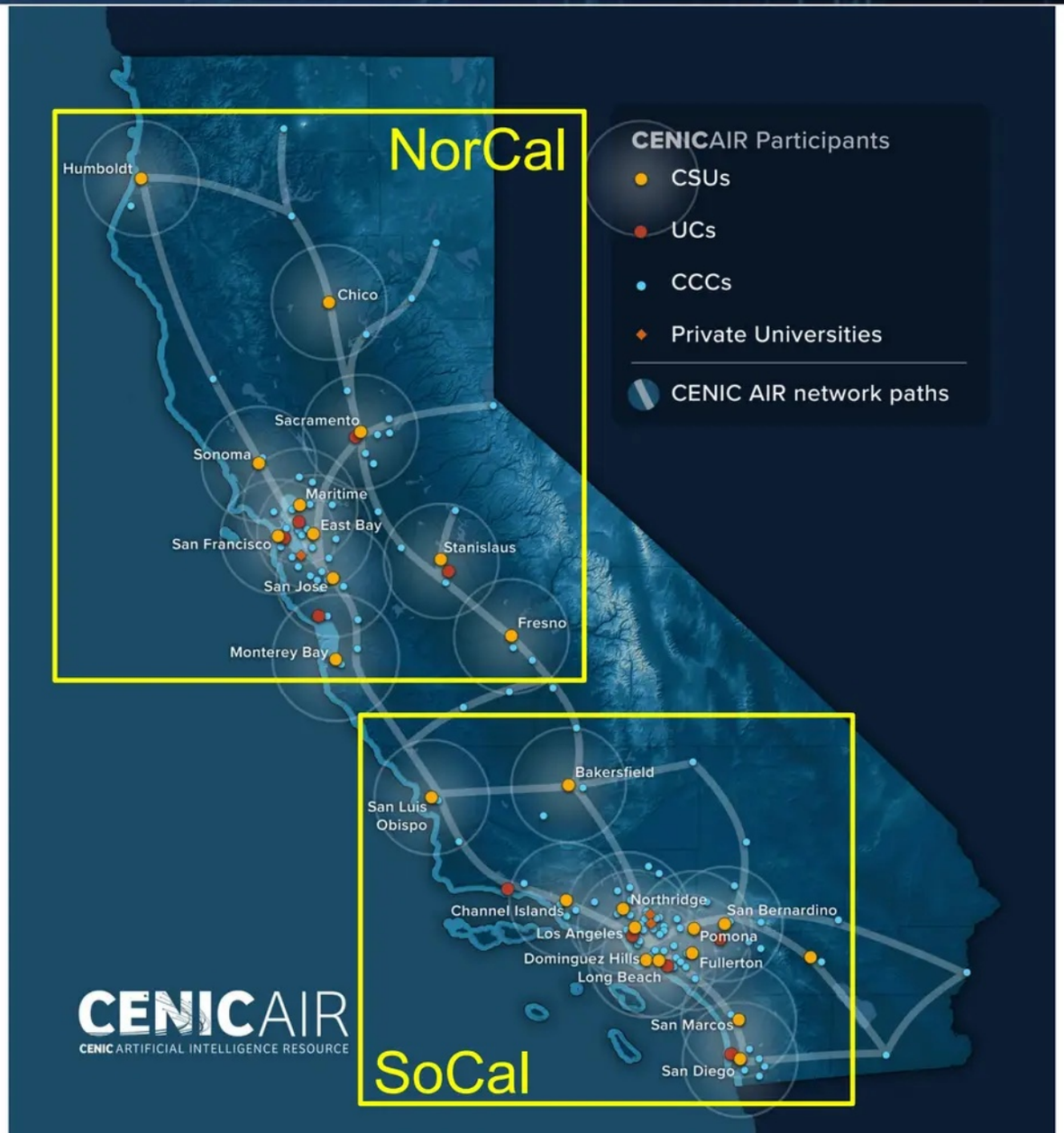
Users Can Execute Their Containerized Applications in the NRP or in Commercial Clouds

Nautilus Containerized Applications Are "Cloud Ready"

Nautilus



California's Research & Education Campuses are Regionally Organized



NRP NATIONAL RESEARCH PLATFORM

NRP Nautilus

SDSU Is the Lead Campus for the New CENIC-Connected Technology Infrastructure for Data Exploration (TIDE) NSF Grant

SDSU San Diego State University | Cal Poly Humboldt. | CSUN California State University Northridge | SDSC | CENIC | California State University San Bernardino | California State University San Marcos | Channel Islands California State University | California State University Stanislaus

- **A CSU Resource for Research & Education:**
  - **NSF CC* Regional Computing Award for ~$1M**
    - **$800 for a New Computational Core**
    - **$200 for Students to Help Onboard and Support Users**
  - **Federated with NRP Nautilus**
  - **First New NSF Resource in CENIC-AIR**

CENIC 2024 Innovations in Networking Award



616 CPU Cores | 73 GPUs | 14.592 TB RAM

NRP National Research Platform | CENIC AIR CENIC Artificial Intelligence Resources

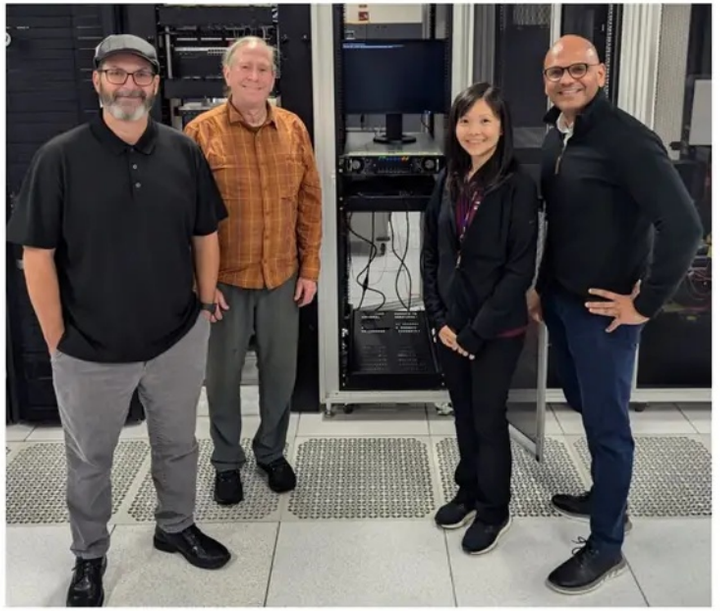
Source: Jerry Sheehan, TIDE PI

NRP Nautilus

By Collaborating with UCSD, SDSU, and CENIC, SDCCD Became the 1st California Community College to Host a CENIC AIR Node



Peter Maharaj, Rose Parnsoonthorn,
In the SDCCD Machine Room





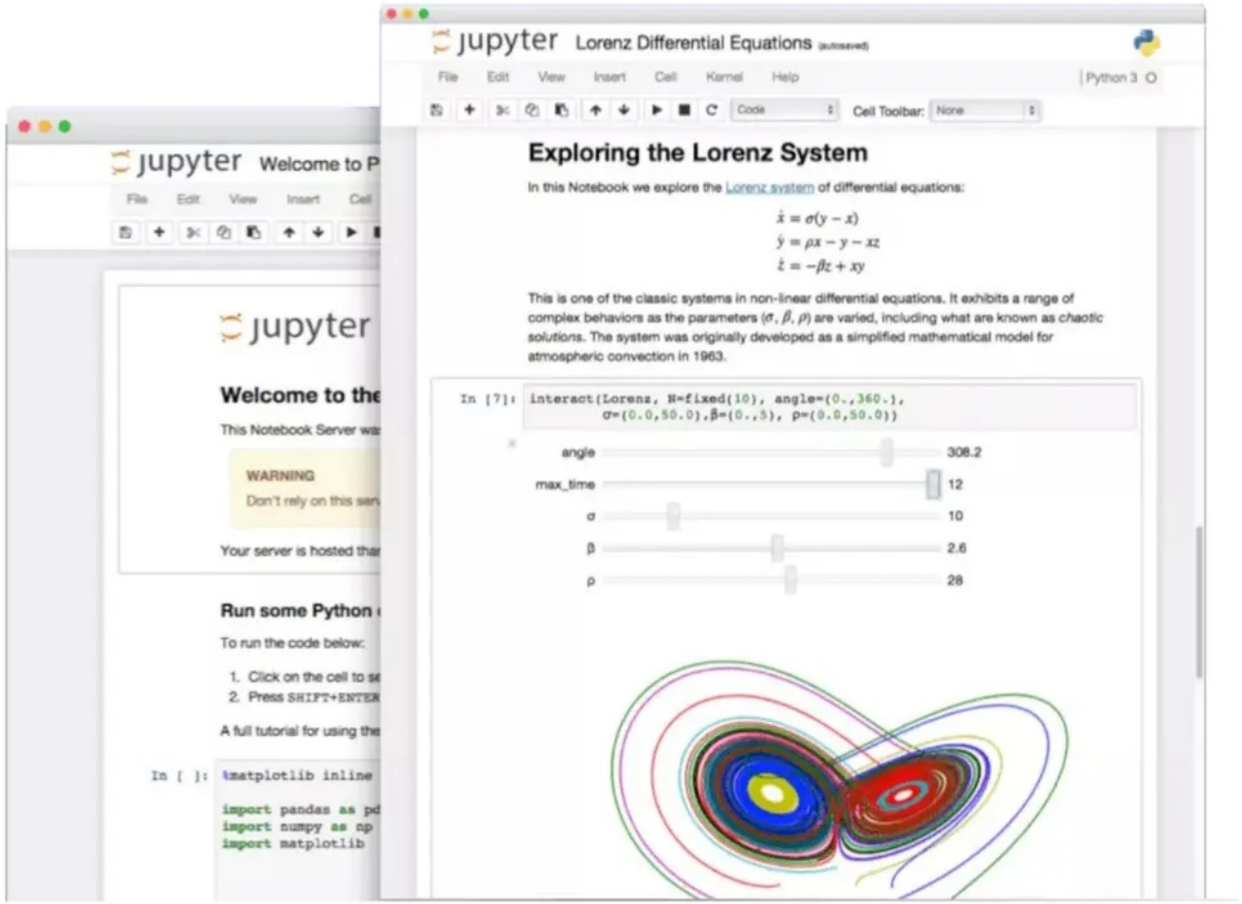
First Nautilus Node Installed In a CCC Machine Room August 30, 2024

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

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

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

jupyterhub is the Multi-User Version of the jupyter Notebook



A multi-user version of the notebook designed for companies, classrooms and research labs

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
- 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

CENIC AIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

IT SERVICES
UC San Diego

NRP Nautilus

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

SDSU Information Technology Division



CENIC

- **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

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 California State San Bernadino Namespaces With Active JupyterHubs

Updated Statistics From CSUSB's Dung Vu

namespace	gpu	cpu
csusb-cousins-lab	100	192,026
csusb-hpc	2,126	101,377
csusb-jupyterhub	1,096	94,686
csusb-khan	9,651	60,432
csusb-qchen	8,399	80,173
csusb-salloum	226	8,901
csusb-ykim	0	22,031
csusb-zhang	59	12,244

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

6 Month Total of
21,676 GPU-Hours
571,869 CPU Core-Hours

SDCCD Faculty Partnered with ITS Leadership to Use CENIC AIR for SDCCD Courses

To: Peter Maharaj, SDCCD Associate Vice Chancellor for ITS

June 26, 2024

I have briefly used the NRP, and it looks awesome. Great work.

I can use a Linux terminal on the cloud for the CISC 211 course.

This platform can deliver Data Science/Analytics and ML courses.

I have a plan to create these courses, and I am happy to hear that the platform will be available.

Again. Good work. I will continue to use this platform this summer.

Is it possible to have this platform available for students this Fall?

Regards,
Danish Khan
Assistant Professor, Computer Science, San Diego Miramar College

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

Step Two: SDCCD Began Building Up CENIC AIR Enabled Course Work

SAN DIEGO Community College District

NewsCenter Home | Articles | Videos | Photo Galleries | Media Resources

SDCCD Information Technology Services is bringing supercomputing to the classroom

September 23, 2024 | San Diego Community College District



The San Diego Community College District has taken a major step toward the future of supercomputing by joining a worldwide collaboration of research institutions seeking groundbreaking scientific discoveries.

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

The San Diego Community College District Set Up Two Namespaces With Active JupyterHubs

namespace	gpu	cpu
sdccd-jupyterhub-dev	804	6,071
sdccd-jupyterhub-prod	2,016	12,162

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

6 Month Total of
2,820 GPU-Hours
18,233 CPU Core-Hours

Namespaces PI Areeluck Parnsoonthorn

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

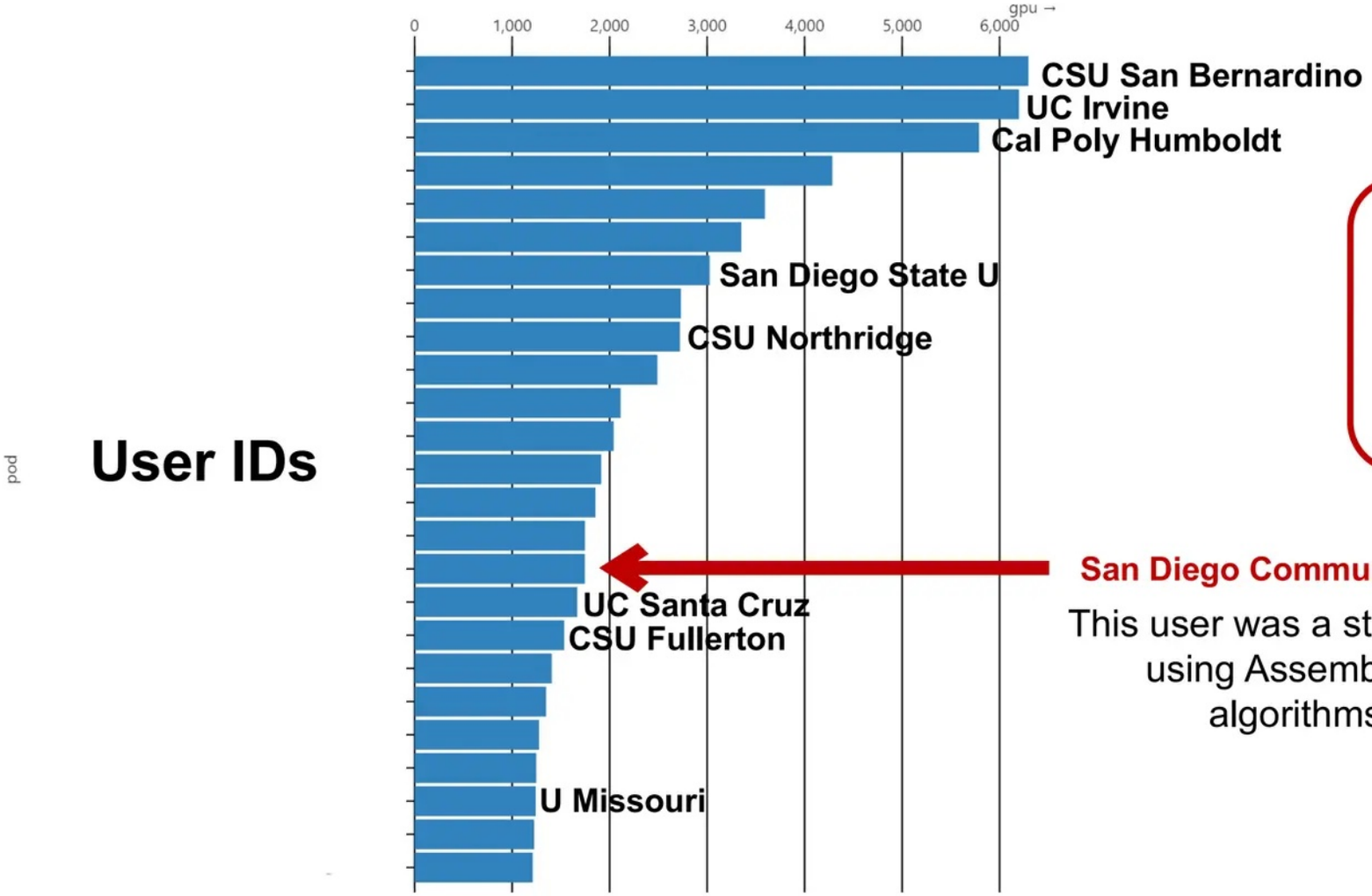
In the Last 6 Months, 88 NRP Namespaces Contained Active JupyterHubs: SDCCD Has the User With the 16th Largest GPU Allocation Out of 363!

Generic GPU Hours Allocated

Top 25 Users out of 363

https://observablehq.com/

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



Only 2 UCs and 4 CSUs Had Users With More JupyterHub GPU-Hours

San Diego Community College District [1,746 GPU-Hours]

This user was a student in an SDCCD College course, using Assembly Language and trying different algorithms to generate prime numbers

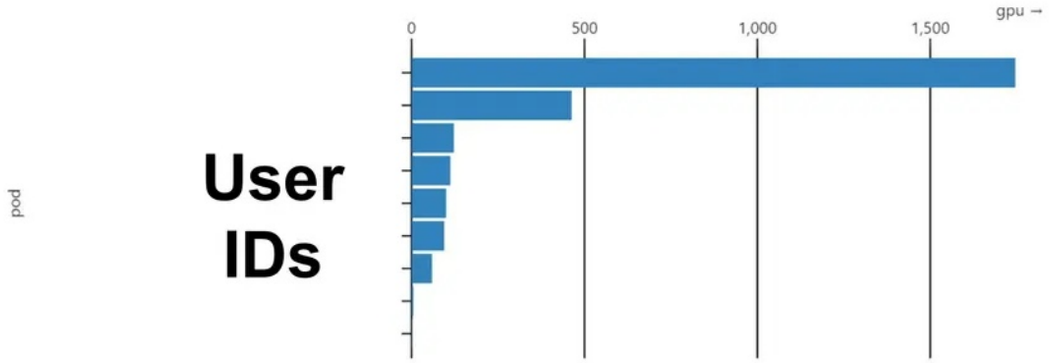
In the Last 6 Months the San Diego Community College District's Two Namespaces with JupyterHubs Had 35 Users!

https://observablehq.com/

Students were in the San Diego Miramar College (SDCCD) Course *Assembly Language and Python*, Taught by Professor Danish Kahn

Generic GPU Hours Allocated

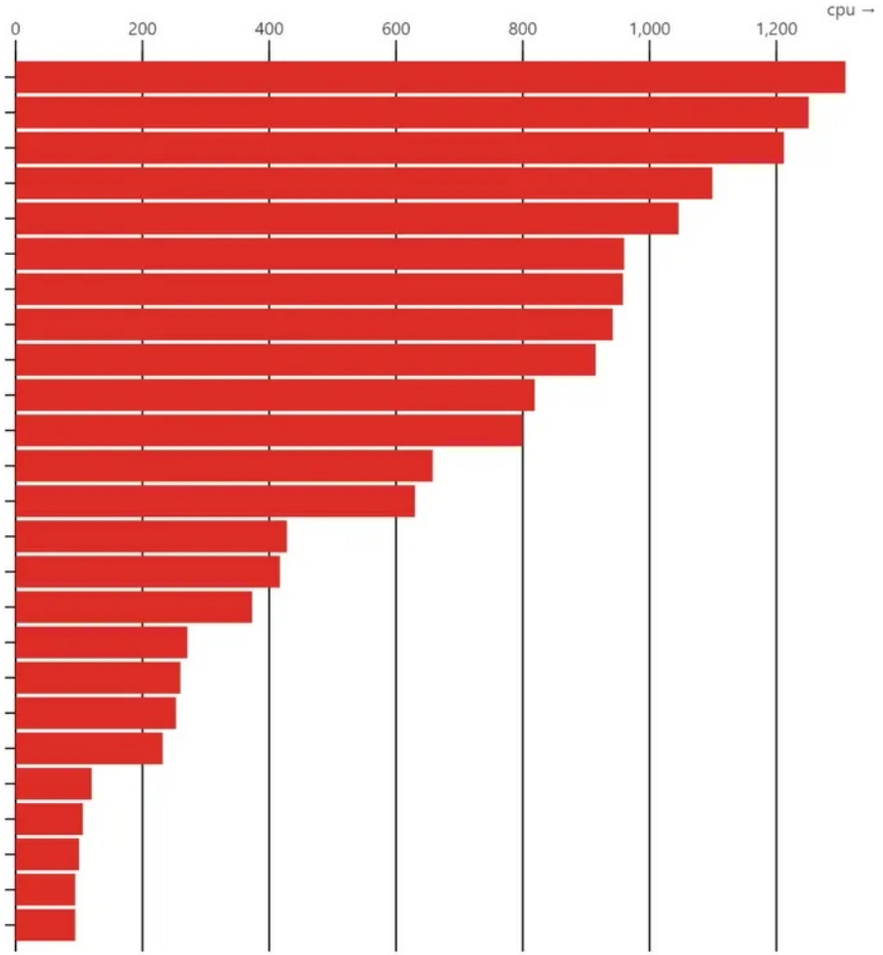
Top 25 Users out of 9



User IDs

CPU Hours Allocated

Top 25 Users out of 35



User IDs

- [x] sdccd-jupyterhub-dev
- [x] sdccd-jupyterhub-prod

Namespaces PI Areeluck Parnsoonthorn

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

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**

SDCCD Submitted an NSF Regional Computing Proposal
October 15, 2024

NSF 24-530-Campus Cyberinfrastructure (CC*) Area 2: Regional Computing San Diego Community College District - Data, Infrastructure, Science, Computing, Opportunities, Virtual, Exploration and Research (DISCOVER)

(PI) Dr. Peter Maharaj (SDCCD);
Co-PIs Dr. Michelle Fischthal (SDCCD) and Areeluck Parnsoonthorn (SDCCD).

Senior Personnel, Carmen Lidz,
Vice Chancellor and Chief Information Officer
for the Los Angeles Community Colleges District

"SDCCD proposes to lead the development of the first dedicated computational core facility for undergraduate research within the California Community College system."

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus

Join in Using the CENIC-AIR

NRP NATIONAL RESEARCH PLATFORM

CENICAIR
CENIC ARTIFICIAL INTELLIGENCE RESOURCES

NRP Nautilus