# “The CENIC-Connected Cyberinfrastructure Commons: Enabling AI for Research and Education”

**Annual CENIC Retreat**

**AI Panel**

**July 18, 2023**

**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

# The United States is Facing an AI Workforce Shortage










NRP NATIONAL RESEARCH PLATFORM

# The White House Convened a Top Level Task Force



JANUARY 24, 2023

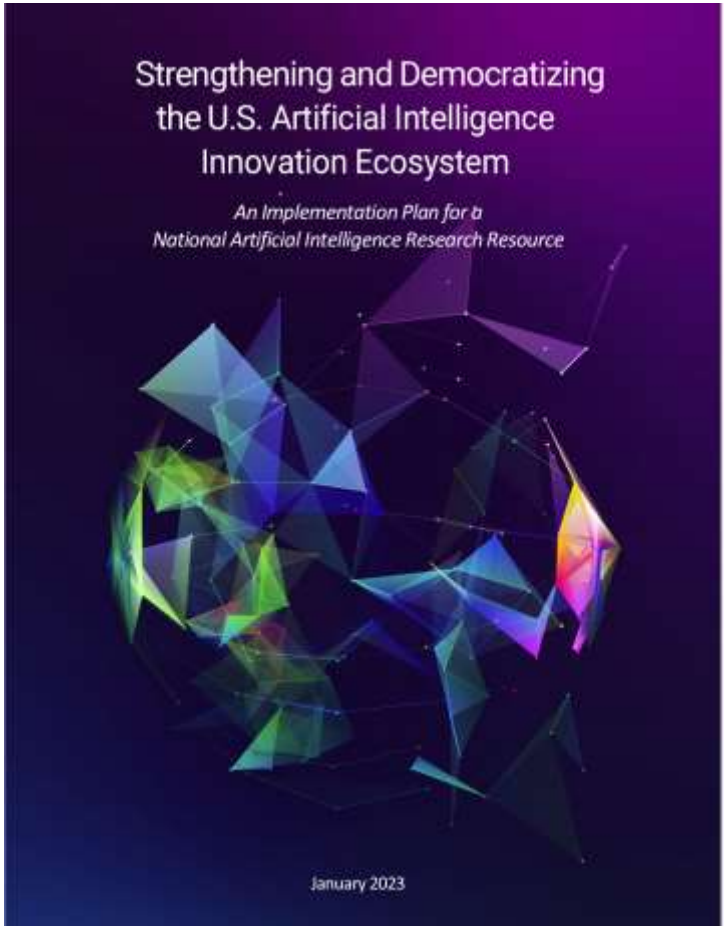
National Artificial Intelligence Research Resource Task Force Releases Final Report

While AI research and development (R&D)
in the United States is advancing rapidly, opportunities to pursue cutting-edge
AI research and new AI applications are often inaccessible to researchers.

NRP NATIONAL RESEARCH PLATFORM

# 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

# California is On The Path to Provide Leadership for NAIRR

NRP NATIONAL RESEARCH PLATFORM

# 2015-2022: The Pacific Research Platform NSF Grants Were Built on CENIC

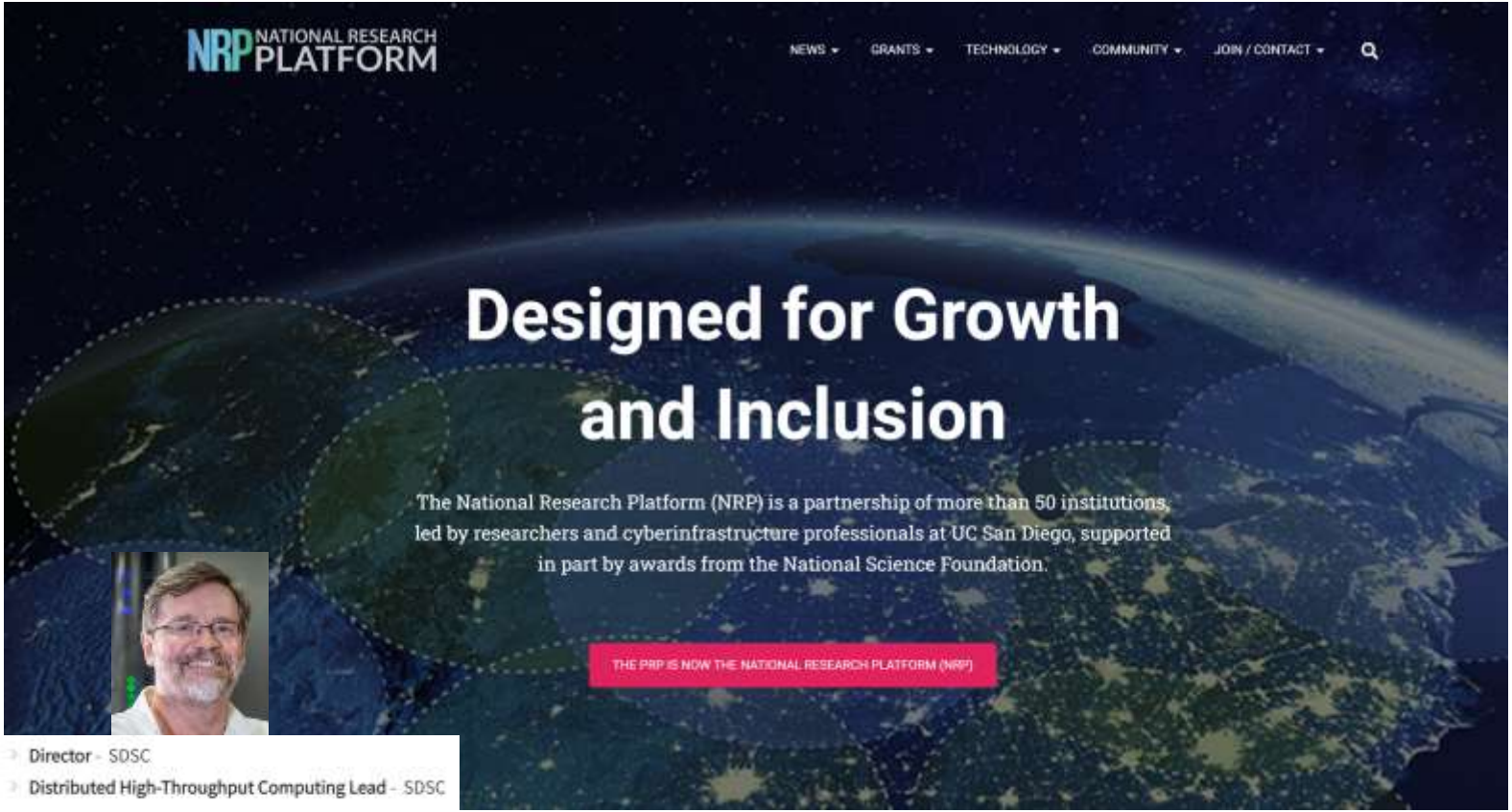



**NSF CC*DNI Grant**
**$7.3M 10/2015-10/2020**
**Extended 2 Years – Ended 10/2022**




Source: John Hess, CENIC

# 2023 - The National Research Platform Unifies The Previous 8 Years of NSF Grants




NRP NATIONAL RESEARCH PLATFORM

https://nationalresearchplatform.org/

# 2023: NRP's *Nautilus* is a Multi-Institution Hypercluster Which Creates a "Cyberinfrastructure Commons"

Nautilus

~200 FIONAs on 30 Partner Campuses
Networked Together at 10-100Gbps

NSF

| Installed CPU Cores | Installed GPUs | Rotating Storage |
|---|---|---|
| 19210 | 1229 | 10,000 TB |

CENIC

THE QUILT

INTERNET 2

NRP NATIONAL RESEARCH PLATFORM

July 14, 2023

# PRP's Nautilus Hypercluster Adopted Open-Source Kubernetes and Rook to Orchestrate Software Containers and Manage Distributed Storage







Production-Grade Container Orchestration

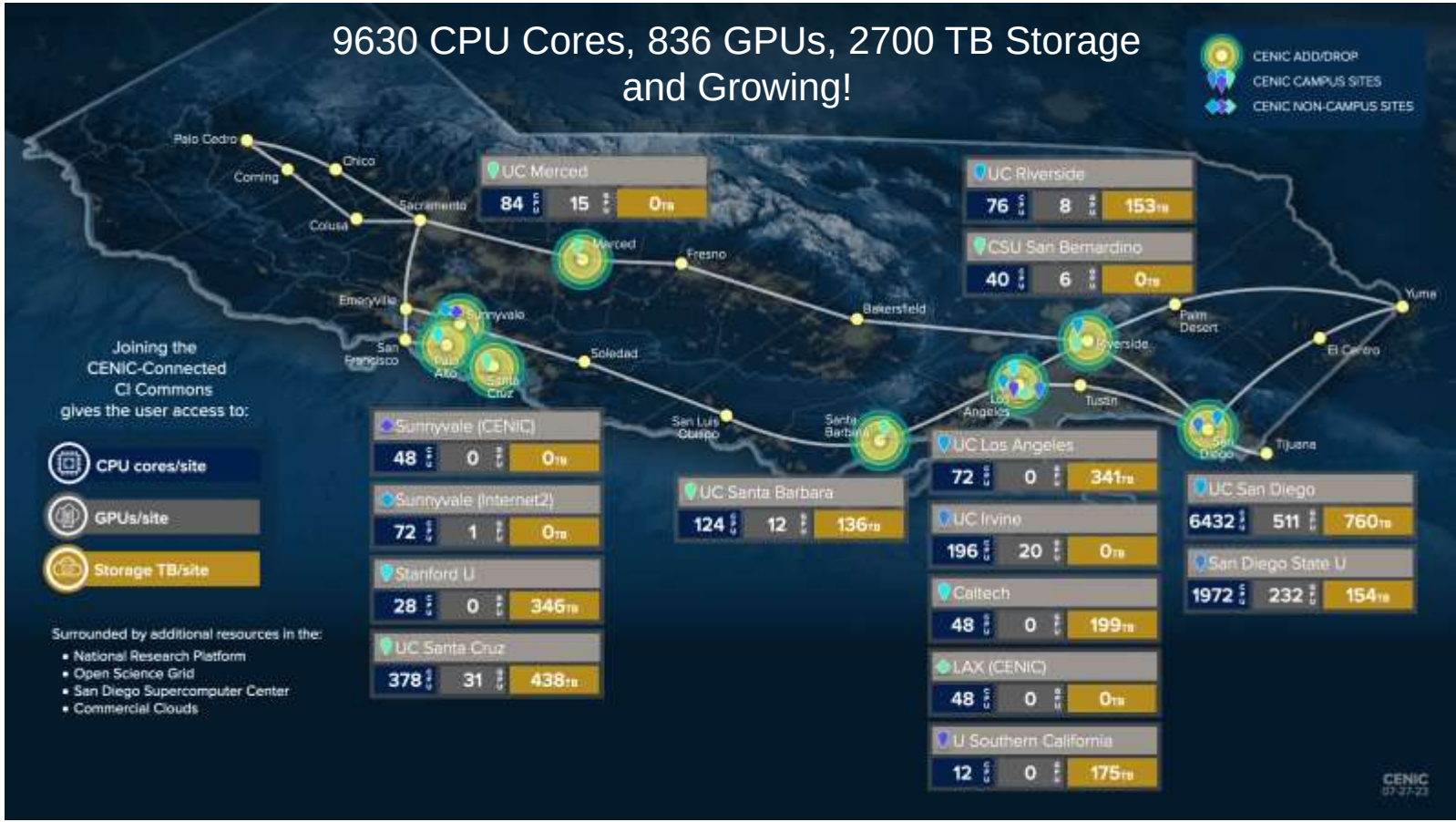



"Kubernetes with Rook/Ceph Allows Us to Manage Petabytes of Distributed Storage and GPUs for Data Science, While We Measure and Monitor Network Use."
--John Graham, UC San Diego

NRP NATIONAL RESEARCH PLATFORM

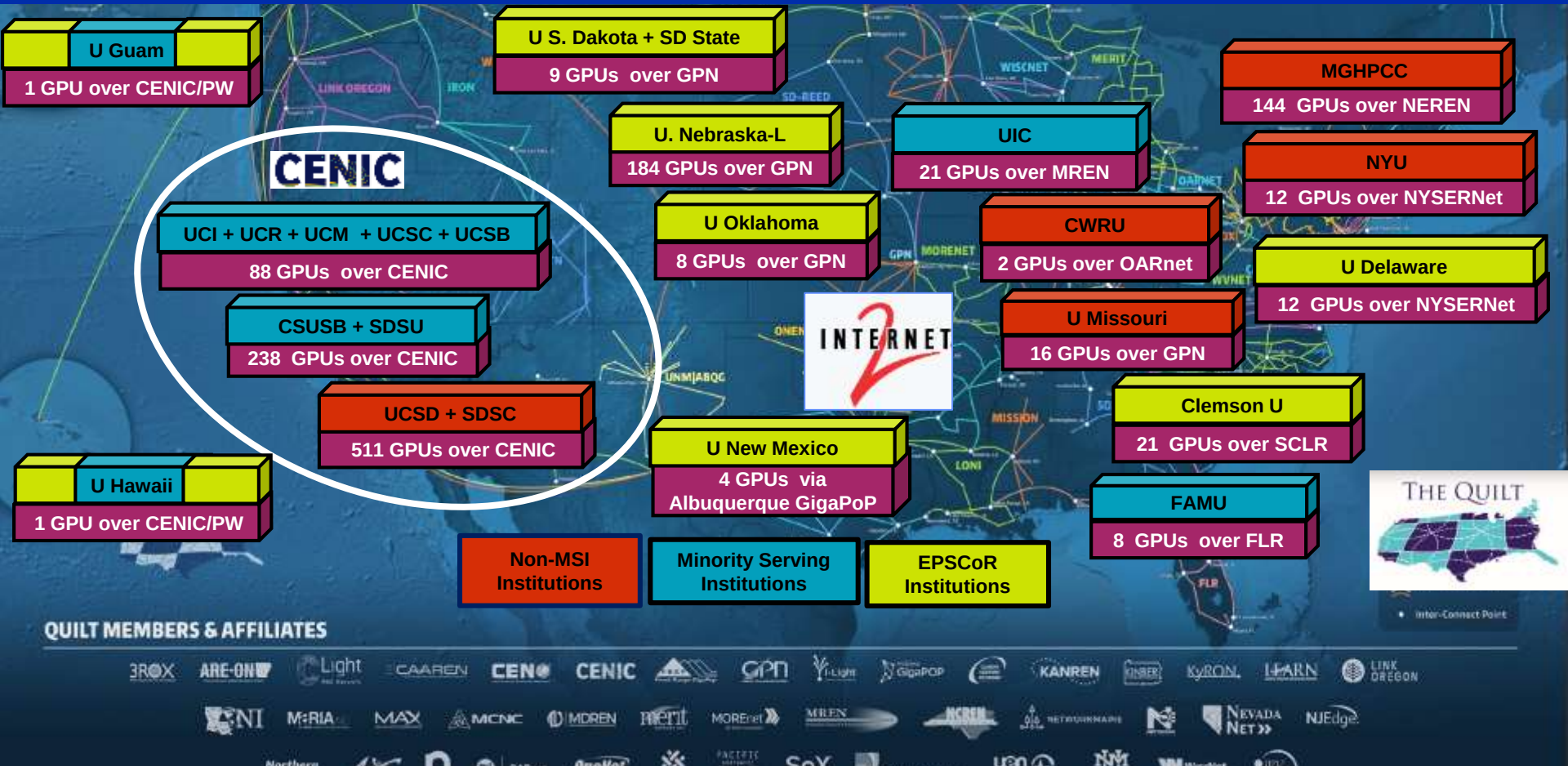
# The CENIC-Connected Cyberinfrastructure Commons: Hosted by and Available to CENIC Members

**The CENIC-Connected Subset of the NRP's Nautilus**




Graphics by Hunter Hadaway, CENIC; Data by Tom DeFanti, UCSD

# The Users of the CENIC-Connected CI Commons Can Burst into NRP's Nautilus Hypercluster Outside of California



# The Users of the CENIC-Connected CI Commons Can Burst into The Open Science Grid (OSG), Which Has Been Integrated With the NRP

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

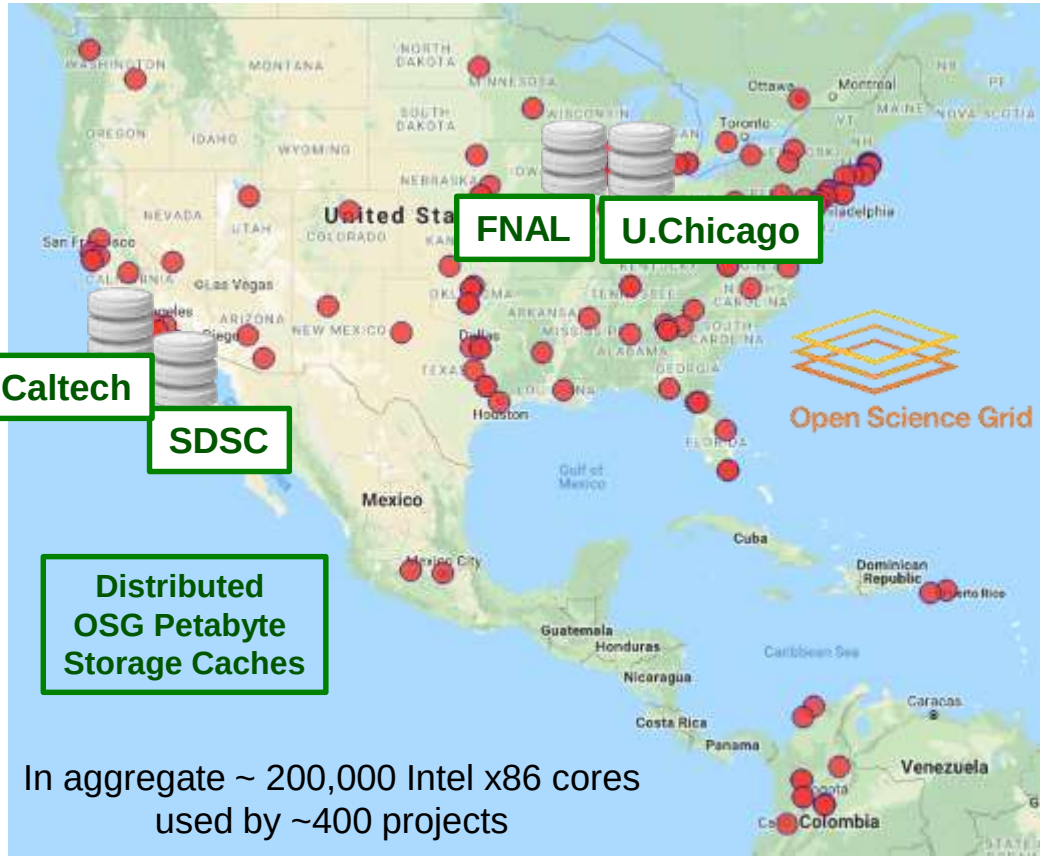
All OSG User Communities
Use HTCondor for Resource Orchestration

HTCondor
High Throughput Computing

NSF

NRP NATIONAL RESEARCH PLATFORM

## OSG Federates ~100 Clusters Worldwide



# The Users of the CENIC-Connected CI Commons Can Burst into SDSC's Supercomputing and Data Storage Resources

SDSC SAN DIEGO SUPERCOMPUTER CENTER
at the UNIVERSITY OF CALIFORNIA SAN DIEGO
COMPUTING WITHOUT BOUNDARIES



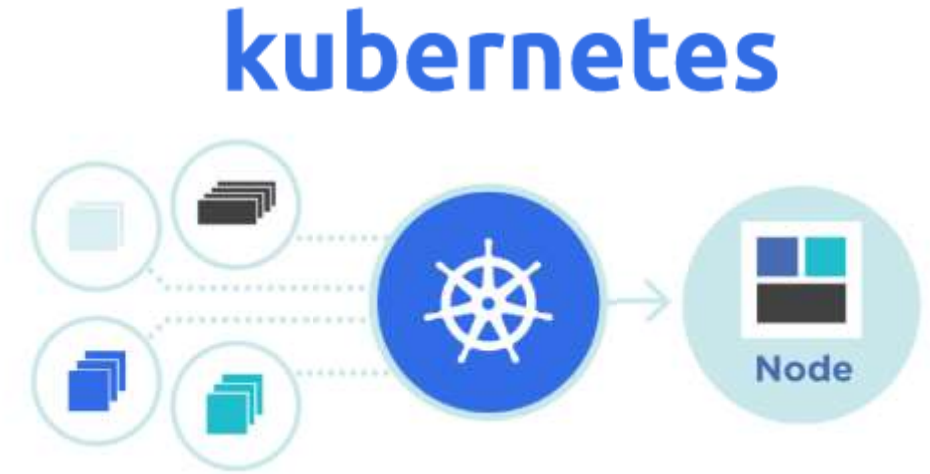

EXPANSE

Data Oasis

NRP NATIONAL RESEARCH PLATFORM

# The Users of the CENIC-Connected CI Commons Can Also Burst Into the Commerical Clouds

User Applications



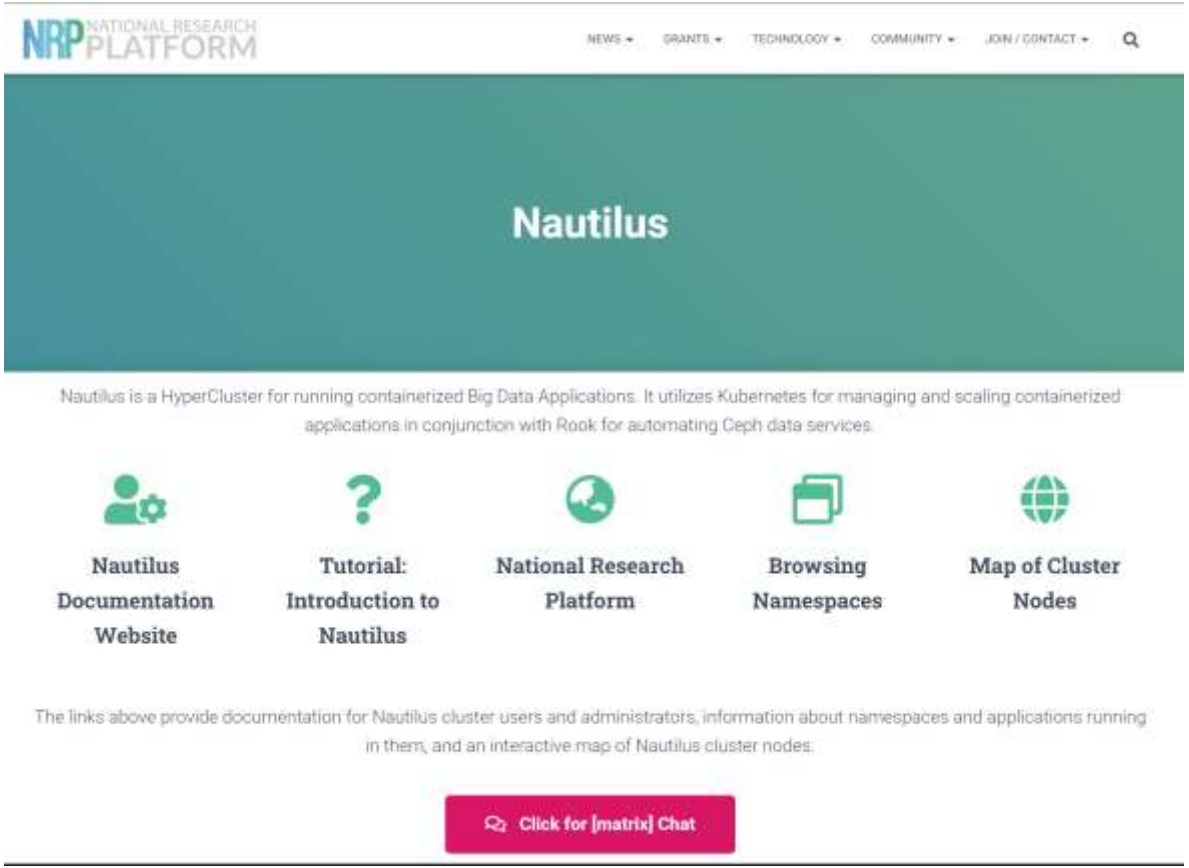

Containers

NRP NATIONAL RESEARCH PLATFORM

Clouds

NRP NATIONAL RESEARCH PLATFORM

# The NRP Web Site Provides Widely-Used Open-Source Services For How to Join and Get Expert Support




https://nationalresearchplatform.org/nautilus/

# California State University San Bernardino is an Excellent Example of How to Help Your Faculty and Students Use NRP



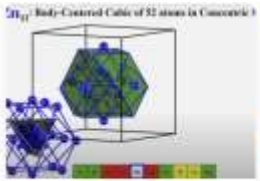

Their Campus HPC Program
Enabled CSUSB Faculty & Students
to Use More NRP GPU-Hours
In the Last 12 Months
Than 7 of the 10 UC Campuses!

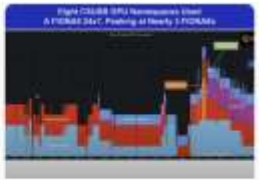


JupyterHub, NRP



ACCESS


Projects


2023 Faculty Showcase

www.csusb.edu/academic-technologies-innovation/xreal-lab-and-high-performance-computing/high-performance-computing

NRP NATIONAL RESEARCH PLATFORM

**Hear More: Dr. Samuel Sudhakar, CSUSB VP ITS**

# The Rapid Rise of Machine Learning, Artificial Intelligence, and Data Science Research Applications & Student Training on the NRP

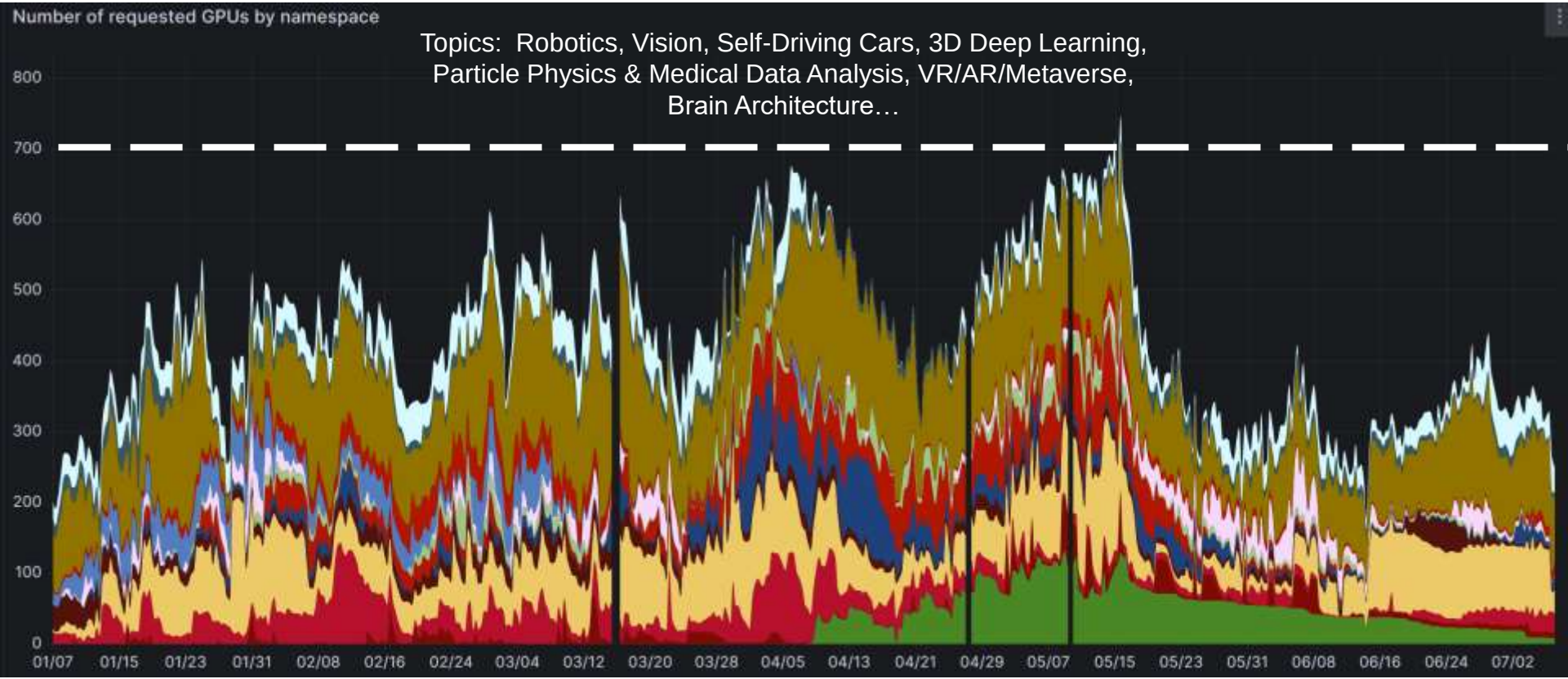
"Graphics processing units (GPUs), originally developed for accelerating graphics processing, can dramatically speed up computational processes for deep learning. They are an *essential* part of a modern artificial intelligence infrastructure, and new GPUs have been developed and optimized specifically for deep learning."

run:ai

www.run.ai/guides/gpu-deep-learning

NRP NATIONAL RESEARCH PLATFORM

# Top 15 GPU-Consuming ML/AI Academic Research Nautilus Namespaces: Last Six Months-Peaking at Over 700 GPUs




NRP NATIONAL RESEARCH PLATFORM

For More Details on Nautilus Applications, Including ML/AI Namespaces Like the Ones Above
See my 4NRP Talk: www.youtube.com/watch?v=1yUz0BwObGs&list=PLbbCsk7MUIGdHZzgZqNbZkV7KGVZ7gn1g&index=19

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



SDSC Racked FIONAs:

- 139 32-bit GPUs (10% NRP)
- 2048 CPU-cores
- 10/25/40G Networking
- *Not* Federated With NRP

Software Used By Students

TensorFlow, PyTorch, Caffe, jupyter, NumPy, ANACONDA, NVIDIA CUDA, docker, kubernetes

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

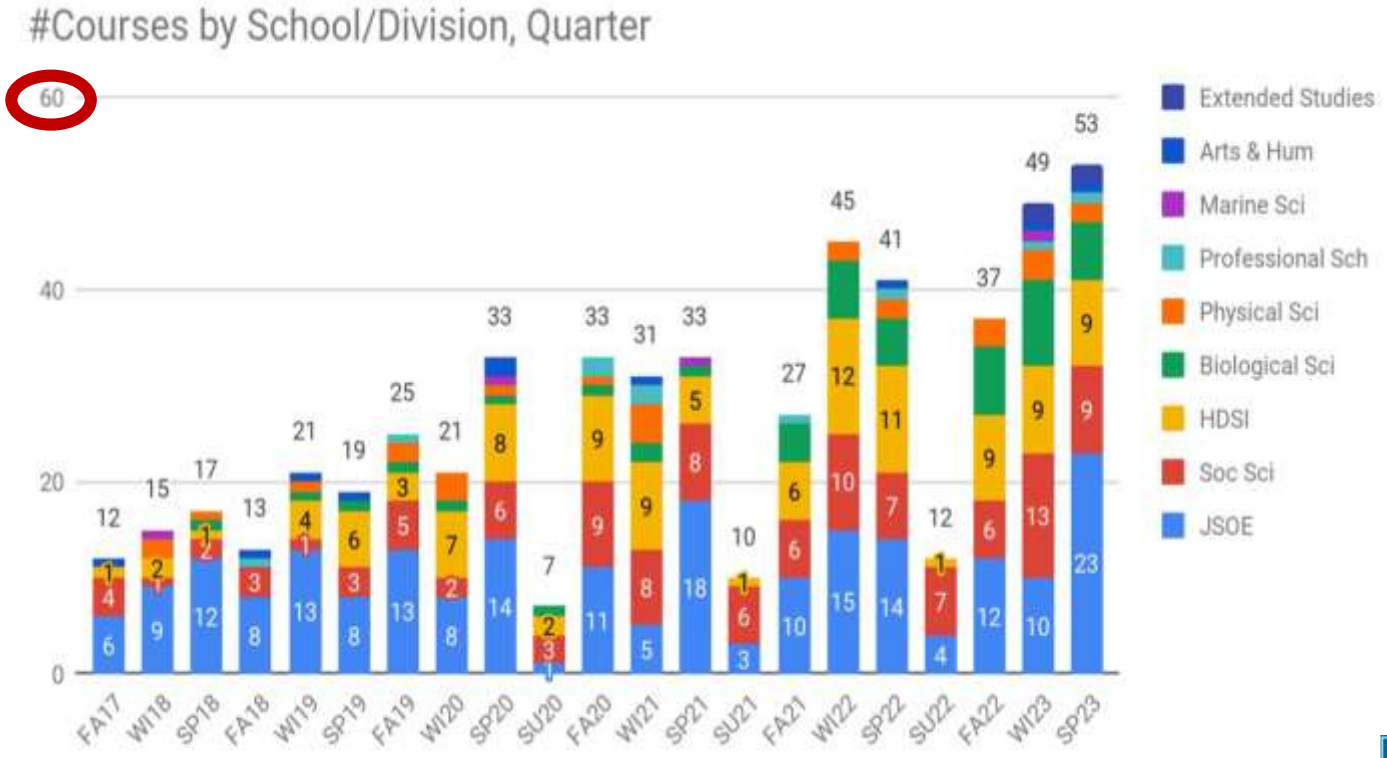
**Source: Adam Tilghman, UCSD ITS**

Data Science/ML Platform - 5/2023 Update

NRP NATIONAL RESEARCH PLATFORM

IT SERVICES UC San Diego

# UCSD's DS/ML Platform Supported Nearly 60 Courses Spring Quarter 2023




**Source: Adam Tilghman, UCSD ITS**

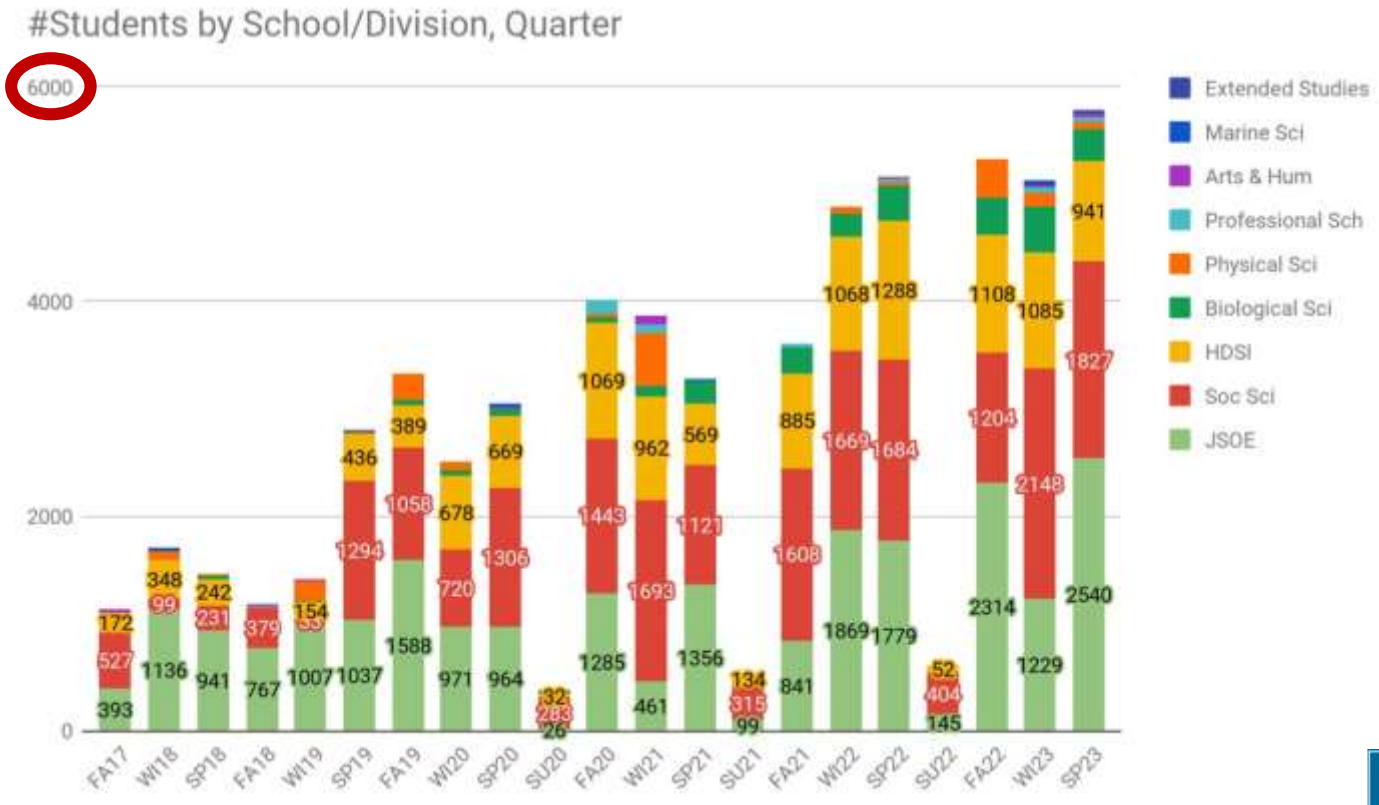
Data Science/ML Platform - 5/2023 Update

NRP NATIONAL RESEARCH PLATFORM

IT SERVICES

UC San Diego

# UCSD's DS/ML Platform Supported Over 6,000 Students Spring Quarter 2023




**Source: Adam Tilghman, UCSD ITS**

Data Science/ML Platform - 5/2023 Update

NRP NATIONAL RESEARCH PLATFORM

IT SERVICES

UC San Diego

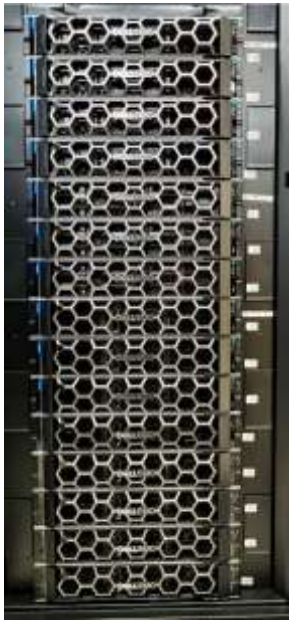
# Selected Courses, Spring 2023

- Advanced Computer Vision
- Bioinformatics for Immunologists
- Computational Physics: Probabilistic Models/Sim.
- Data Analysis/Design for Biologists
- Data Science/Spatial Analysis
- Deep Learning and Applications
- Intro to Causal Inference
- Neural Networks/Pattern Recognition
- Numerical Analysis for Multiscale Biology
- Robot Manipulation and Control

NRP NATIONAL RESEARCH PLATFORM

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

SDSU Information Technology Division



CENIC

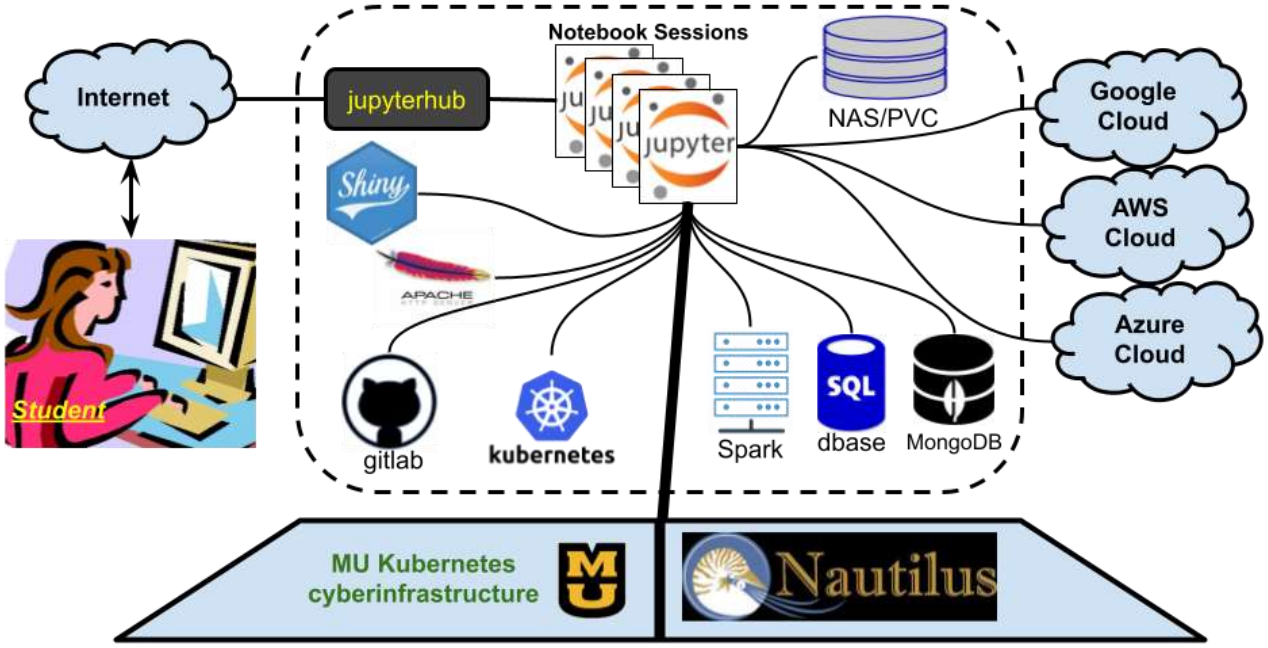
jupyter TensorFlow PyTorch kubernetes NVIDIA docker

- Dell PowerEdge Cluster for Instructional Use, with 15 Nodes with 32 A100 GPUs (managed as 232 independent GPUs), 768 CPU cores, and 960 TB Storage.
- A Learning Resource that *Surges* for Research and Instruction
- System Administration as a Service from the NRP
- Friendly User mode this Spring with
  - ~90 Users in Computational Engineering 361 & Astronomical Techniques ASTR 350
  - Research in Developing Synthetic Medical Data for Machine Learning Training

NRP NATIONAL RESEARCH PLATFORM

Slide courtesy of Jerry Sheehan, SDSU CIO

# University of Missouri is Leveraging Jupyter Hubs on Nautilus for STEM & AI/ML Experiential Learning



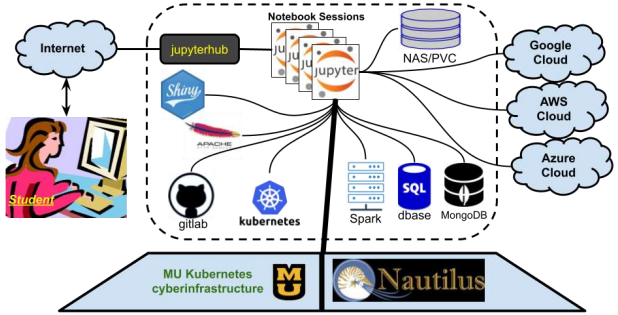

Has 16 GPUs On-Site Now -
Adding 44 with a New CC* Award


Grant Scott
https://scottgs.mufaculty.umsystem.edu/


NRP NATIONAL RESEARCH PLATFORM

University of Missouri

# Customized Jupyter Hubs on Nautilus Support Multiple Courses and Certificates




- Computer Science Undergraduate Classes
- Computer Science HPC Classes
- Undergraduate Data Science
- HPC Emphasis Graduate Data Science
- State & Federal Government Training Programs

DEPARTMENT OF HIGHER EDUCATION & WORKFORCE DEVELOPMENT

NRP NATIONAL RESEARCH PLATFORM

**Grant Scott**
https://scottgs.mufaculty.umsystem.edu/

University of Missouri

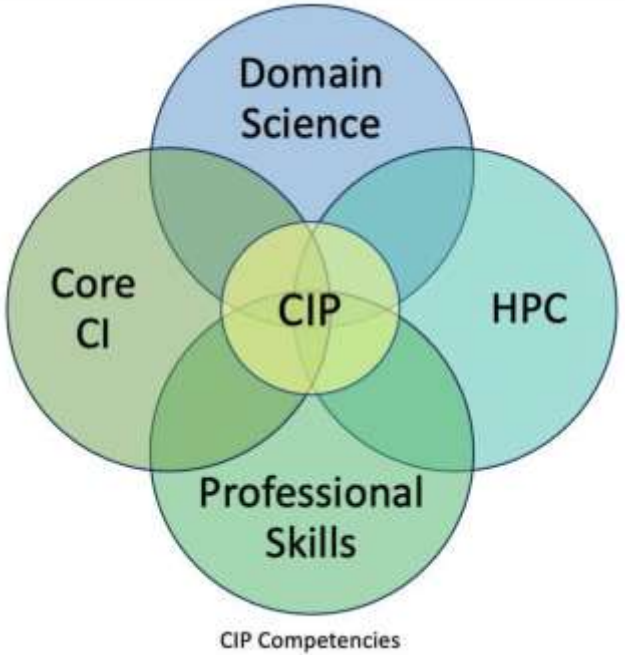
# But Who Will Train the Trainers?
# NSF is Providing CyberTraining Funding

October 20, 2022

## Three California Universities Commit to Growing a Community of Cyberinfrastructure Professionals

SDSC at UCSD, SDSU & CSUSB have been awarded $6.7 million for training and mentoring a cohort of CI professionals.

Our training program cohort will be drawn from populations including domain scientists, interdisciplinary research professionals, faculty, students and other NSF workforce individuals and underrepresented STEM communities."
--PI Mary Thomas, SDSC



CIP Competencies

NSF award #2230127

NRP NATIONAL RESEARCH PLATFORM

# GPT Can Be an AI Force Multiplier

## GPT-4 is Great at Coding.

## Integrating Chat GPT with Data Analysis Tools

## Run Interactive Sessions With ChatGPT In Jupyter Notebook

## A.I. Course Creation: How to Use ChatGPT to Create eLearning Content

# Help Us Find The Minds We Need From the CENIC Family: Join the Growing Users of the CENIC-Connected CI Commons

- UC Berkeley
- UC Davis
- UC Irvine
- UC Los Angeles
- UC Merced
- UC Riverside
- UC San Diego
- UC San Francisco
- UC Santa Barbara
- UC Santa Cruz

Minority Serving Institutions

- Cal Poly Humboldt
- Cal Poly Pomona
- Cal State Fullerton
- Cal State Monterey Bay
- Cal State Northridge
- Cal State San Bernardino
- San Diego State
- San Jose State

- Caltech
- Harvey Mudd College
- Stanford U.
- USC

- Naval Postgraduate School

Future Users From:
- 15 More CSU
- Community Colleges
- High School
- Libraries
- Planetariums
- Museums

NRP NATIONAL RESEARCH PLATFORM

# The CENIC-Connected CI Commons Can Help Support The Engagement of California's Diverse Research & Education Communities

POLITICS

## As Congress grapples with AI regulation, will California step up?



Sophia Bollag

July 17, 2023

San Francisco Chronicle




AI summit at the Fairmont Hotel in San Francisco in June,

NRP NATIONAL RESEARCH PLATFORM

# Exploring the Opportunities to Utilize the CENIC-Connected CI Commons For Rapid Growth in California-Wide AI Research, Education, and Training

- **Panelists**
  - **Sam Sudhakar, CSU**
  - **Terry Loftus, K-12**
  - **John Hetts, CCC**
  - **Josh Chislom, Libraries**

NRP NATIONAL RESEARCH PLATFORM

# NRP Support and Community:

- US National Science Foundation (NSF) awards and subawards to UCSD
  --CNS-1456638, CNS-1730158, CNS-2100237, CNS-2120019
  --ACI-1540112, ACI-1541349, OAC-1826967, OAC-2029306, OAC-2112167
- DOD DURIP awards to UCSD
- UC Office of the President, Calit2 and Calit2's UCSD Qualcomm Institute
- San Diego Supercomputer Center and UCSD's Research IT and Instructional IT
- CENIC, Pacific Wave/PNWGP, StarLight/MREN, The Quilt, Great Plains Network, NYSERNet, Open Science Grid, Internet2, DOE ESnet, NCAR/UCAR & Wyoming Supercomputing Center, AWS, Google, Microsoft, Cisco, Juniper, Arista

