



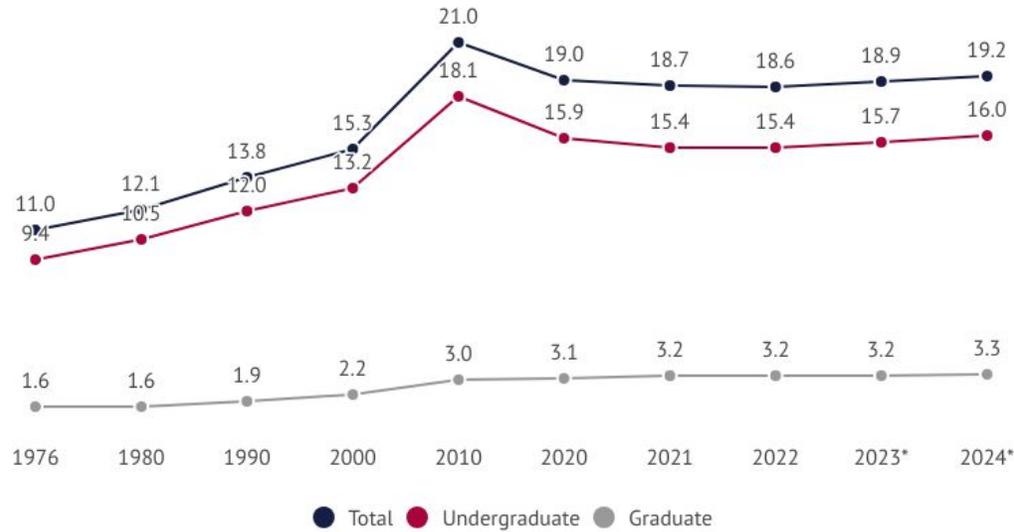
# AI Infrastructure for All

**Frank Wuerthwein**  
**Director**  
**San Diego Supercomputer Center**

In the 2023-24 academic year, there were  
**1,632 public (blue) & 1,938 private non-profit (green)**  
 accredited degree granting colleges in the USA  
**3,580 non-profit colleges total**

Historical College Enrollment, Selected Years

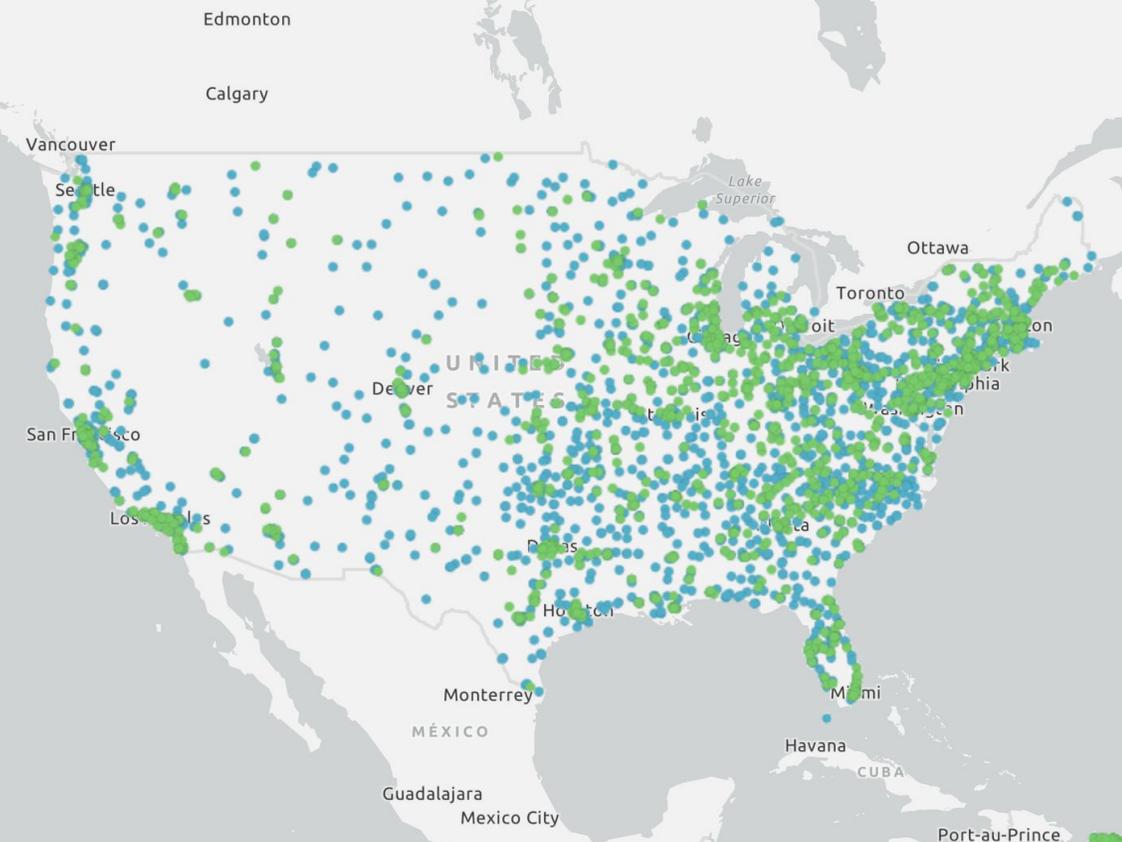
(in millions)



Education Data Initiative source: U.S. Department of Education

\*Estimate

**~20 Million students**  
**40% in community colleges**  
**43% in 4-year colleges**  
**17% in graduate/professional school**



# UC San Diego Digital Assets in the Classroom

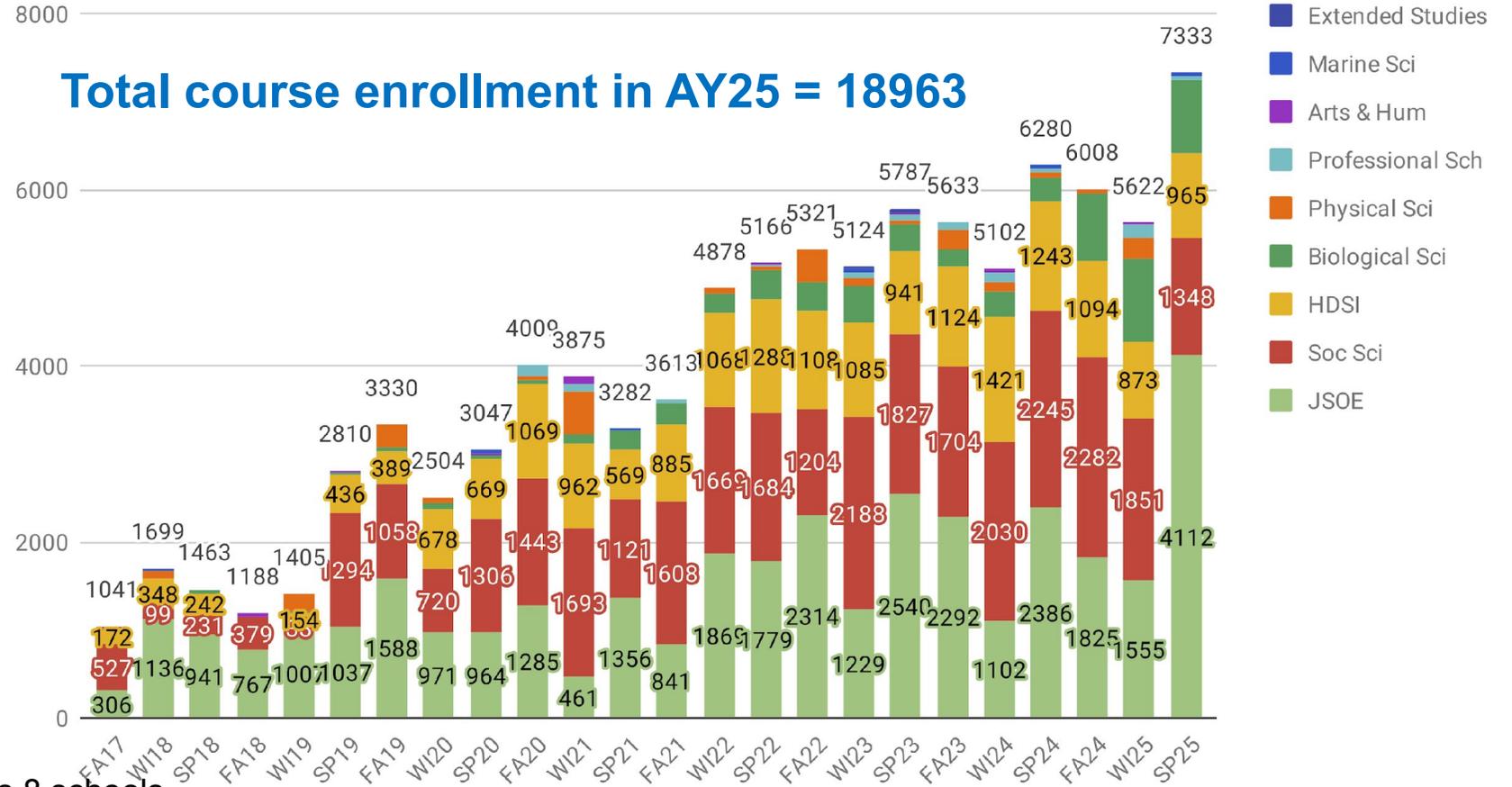
24% of all UCSD undergraduates used compute & data infrastructure in AY25 as part of their classroom education

Total in AY24	11092
UG Students	8538
Grad/Prof Students	2554

Some students take multiple courses/AY

34,955 undergraduates enrolled in Fall 2024

#Students by School/Division, Quarter



In AY25: 132 courses, 81 instructors, across 8 schools

# The beginning of a revolution in education

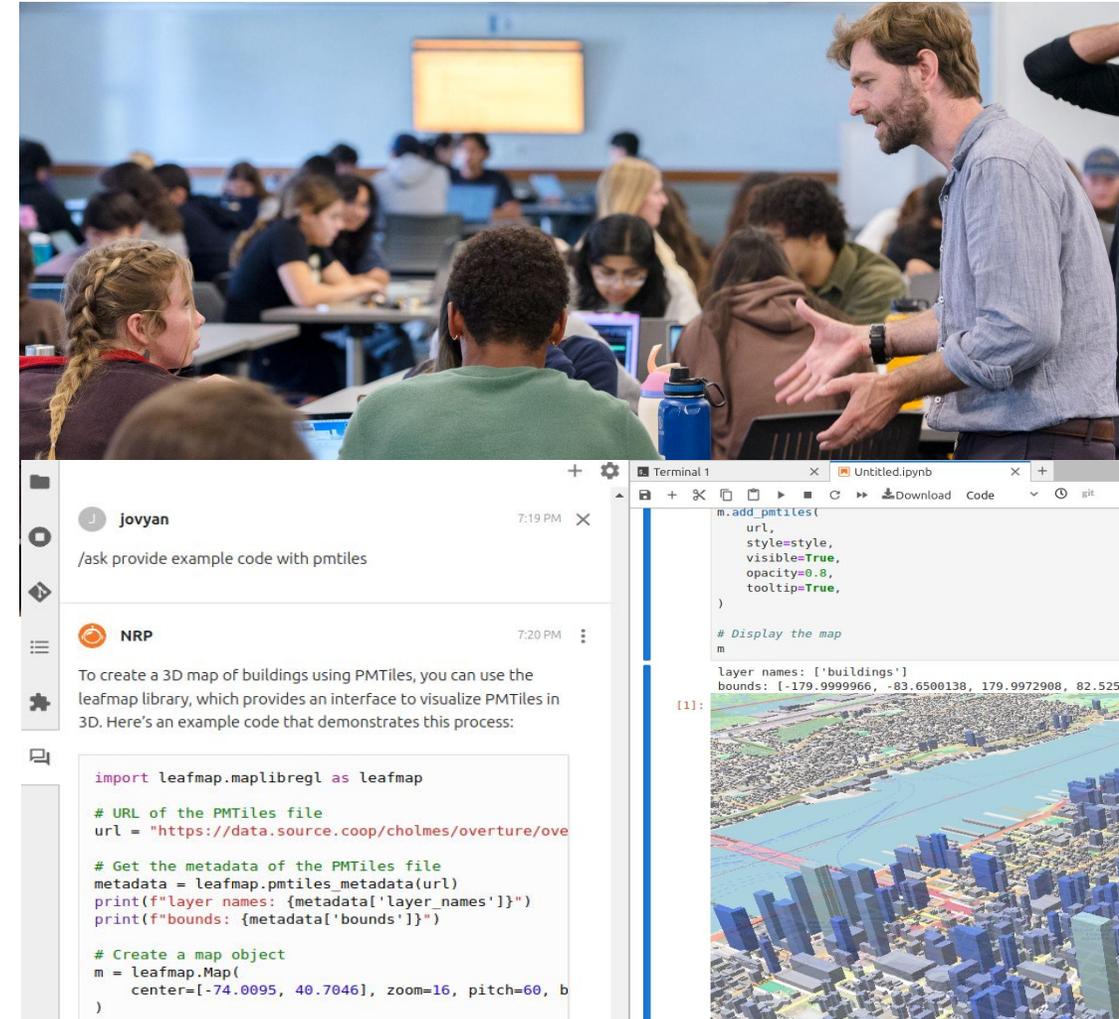
- AI redefines what's possible to teach in a classroom
  - Programming at the fingertips of every student
  - Simulations, data analytics, and visualizations on any topic
- ⇒ **Students transition from consuming to experiencing education**
- AI makes personalized tutoring before, during, and after the classroom session possible.
- ⇒ **More students understand more as a result of personalized attention, even in classes with hundreds or thousands of students**

# Redefining what's possible

Carl Boettiger, UC Berkeley  
NAIRR Classroom Allocation on NRP

## Coding with large language model

- 122 Students
- Active learning classroom
- Students visualize the Manhattan Skyline in 3D with the help of AI during class



# Streamlining access and use...



NATIONAL DATA PLATFORM

## National Data Platform Education Hub

- Intuitive wizards to create compute-ready classrooms for your students to use
- Full range of digital assets available via reusable and reproducible materials that use AI, data, and compute services
- Formal and informal learning materials
- Group data challenges with large data

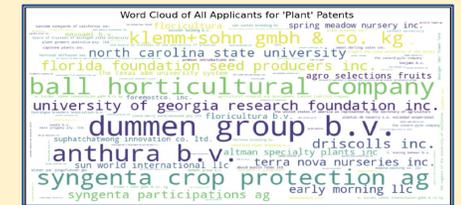
<https://www.nationaldatapatform.org/>

### UC San Diego DSE 203 (Amarnath Gupta)

- **Exercise:** Constructing knowledge graphs with large datasets



Example Student Project with NAIRR USPTO Dataset



### UC San Diego BENG 211 (Benjamin Smarr)

- **Exercise:** Time series analysis of physiology data

The screenshot shows the National Data Platform interface. On the left, there's a sidebar with 'NATIONAL DATA PLATFORM' and a file directory. The main workspace shows a Jupyter notebook with the following content:

```
[4]: df_wideBiomarkers = pd.read_csv(fp_biomarkersWide)
df_wideBiomarkers
```

	Date	% SCFA Acetate	% SCFA Butyrate	% SCFA Propionate	% SCFA Valerate	237_mg/L	AID ratio_None	ALT_U/L	APO A1_mg/dL
0	1993-05-01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1996-11-27	NaN	NaN	NaN	NaN	NaN	1.3	NaN	NaN
2	1997-05-08	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	1998-04-15	NaN	NaN	NaN	NaN	NaN	1.4	NaN	NaN
4	1999-04-01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...
327	2025-02-05	47.0	24.0	22.0	7.2	NaN	NaN	22.0	NaN
328	2025-02-12	NaN	NaN	NaN	NaN	NaN	NaN	14.0	NaN
329	2025-02-28	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
330	2025-05-20	NaN	NaN	NaN	NaN	NaN	NaN	17.0	NaN
331	2025-06-13	NaN	NaN	NaN	NaN	57.0	NaN	15.0	NaN

332 rows x 10 columns

**Biomarker Time Series Summary**

```
[5]: num = df_wideBiomarkers.copy() # Make a working copy so we don't mutate the original DataF
# Convert any non-numeric, non-datetime columns to numeric.
# - Strings (or mixed types) become numbers where possible and NaN otherwise.
# - Datetime columns are left as-is so their min/max still work.
for c in num.columns:
    if not np.issubdtype(num[c].dtype, np.number) and not np.issubdtype(num[c].dtype, np.datetime64):
        num[c] = pd.to_numeric(num[c], errors="coerce")
```

# Let's play with some numbers ...

- 20M students attend college in the USA
- 24% of UC San Diego undergraduates have used data & compute in their classroom in AY25

**=> There are ~5M college students in the USA that need data & compute resources for their classroom education**

# Operations is expensive ... ... unless it is amortized across large scales

- Of the 3,580 accredited, degree granting higher education non-profit institutions less than 200 are research intensive (R1), and have the scale to afford a group of sysadmins, cybersecurity, user support, ... professionals.
- ⇒ **Less than 6% of colleges nationwide can afford AI infrastructure ops**
- **What if we aggregated the system administration, cybersecurity, and user support across 1,000 colleges ?**
- Colleges own their AI hardware, but we centralize the operations to benefit from economies of scale, just like the cloud providers.

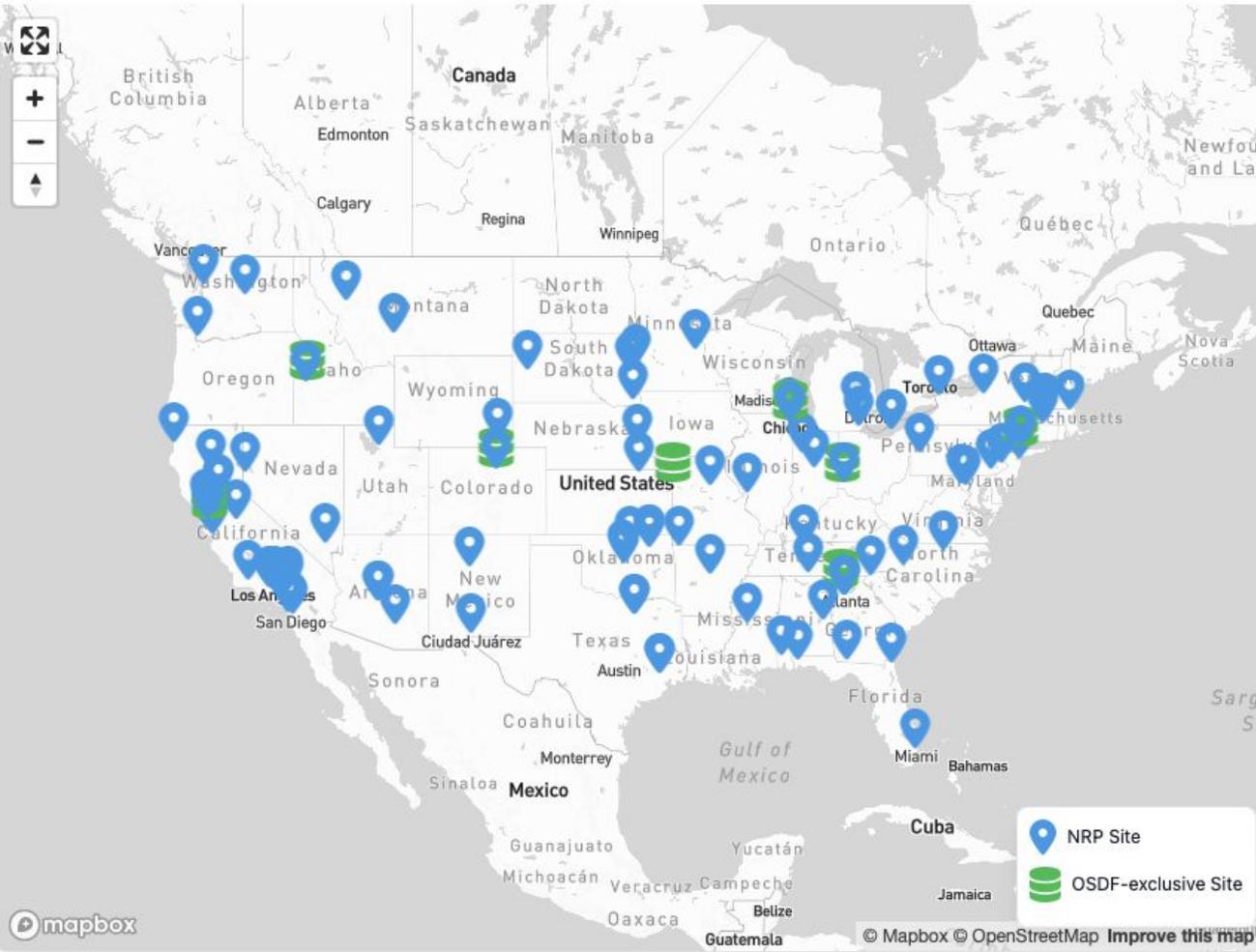
**National Research Platform (NRP), a prototype to achieve this vision**

# A possible path for a new college

- **Some faculty start exploring NRP**
  - <https://nrp.ai/training/>
  - <https://nrp.ai/documentation/>
    - The LibreChat LLM interface is accessible without an account via your InCommon login
  - Zoom based office hour every other week on Tuesday 10am PST
    - Next office hour on January 27<sup>th</sup>
- **When there is sufficient interest among the faculty, college buys hardware for NRP to operate.**
  - You rack up hardware in your data center
  - NRP team operates hardware from IPMI up

# NRP offers the community

- **To run your hardware from IPMI up**
  - OS maintenance, security monitoring, ...
- **Researchers, Educators, and students see a global scale Kubernetes cluster**
  - While the cluster is shared, we restrict use of individual hardware to owners when necessary
  - Documentation and training for the community
- **Lot's of software to reuse**
  - E.g. we operate JupyterHub for the community & show people how to deploy & customize their own
    - 46 community-run JupyterHubs now on NRP
- **Maintain lots of topical Chat channels for the community to learn from each other and interact with each other**
  - Chat channels monitored by professionals, community is encouraged to interact & help each other
  - Starting to explore AI chatbots trained on the chats
- **LLM as a service**
  - We run most popular LLMs as services for all
  - ... and provide popular APIs to upload any AI model from Hugging Face onto our resources.
- **AI workflows via NDP**
  - Reusable workflows that incorporate curated data & AI into classroom exercises



# NRP NATIONAL RESEARCH PLATFORM

The National Research Platform is a partnership of more than 50 institutions, led by researchers at UC San Diego, University of Nebraska-Lincoln, and Massachusetts Green High Performance Computing Center and includes contributions by the National Science Foundation, the Department of Energy, the Department of Defense, and many research universities and R&E networking organizations in the US and around the world.

Select a site or click on a site in the map

Map Customization - Select Multiple Sites

Or select by Regex pattern

<b>Sites</b> 127 Sites hosting NRP nodes	<b>Nodes</b> 497 Nodes registered in Kubernetes
<b>GPUs</b> 1,448 Total GPUs across all nodes	<b>CPU Cores</b> 31,826 Total CPU cores across all nodes

**75% growth since January 2025**

**In January we were at 72 locations and 22 PB of disk space**

## NRP as of today ...

<https://dash.nrp-nautilus.io> 11

## Example: CSU San Bernardino

- < 1,100 faculty members
- < 19,000 students
- < Serves 2 of CA's largest counties
- < Hispanic Serving Institution
- < 57% Pell Grant recipients
- < Many student oriented projects

**Non-R1 Institution**



# JupyterHub User Interface at CSU, San Bernardino



CSUSB JupyterHub ☁

- Accessible to anyone with a CoyoteID
- Each user is provided with 50 GB of dedicated storage, expandable upon request
- Resources include large RAM and storage and multiple CPUs and GPUs
- Pre-installed software includes Linux desktop, Python, Matlab, R, Scipy, Pyspark, Tensorflow, a data science package, and SageMath
- Other software can be installed upon request
- Can be used for research and classes

**Institutional logon**

CSUSB JupyterHub for faculty research ☁

- Dedicated to faculty research
- Contact Youngsu Kim or Dung Vu to access this server

**dedicated hubs for labs**

**Wide range of dedicated hubs:**

## Server Options

Advanced Options

Image

<input type="radio"/>	Stack Minimal
<input type="radio"/>	Stack Datascience
<input type="radio"/>	Stack R-Studio, Vs-code for Dr. Becerra's class
<input type="radio"/>	Stack Desktop Apps - VS Code
<input type="radio"/>	Stack Desktop Apps - Pgadmin4
<input type="radio"/>	Stack Desktop Apps - Blender
<input type="radio"/>	Stack PySpark
<input type="radio"/>	Stack PyTorch2
<input type="radio"/>	Stack R-Studio
<input type="radio"/>	Stack R-Studio for BIOL-5050
<input type="radio"/>	Stack SageMath

<https://csusb-metashape.nrp-nautilus.io>: 3D modeling

<https://csusb-vasp1.nrp-nautilus.io>: Viena Ab initio Simulation package ( VASP)

<https://csusb-cousins-lab.nrp-nautilus.io>: VASP simulation

<https://csusb-becerra.nrp-nautilus.io>: AI/ML project

<https://csusb-biol-5050.nrp-nautilus.io>: Biology course

<https://csusb-cse-salloum.nrp-nautilus.io>: Summer Research

<https://csusb-drhamoudahub.nrp-nautilus.io>: Data Analytics

<https://csusb-ratnasingam.nrp-nautilus.io>: Data Analytics

<https://csusb-zhanq.nrp-nautilus.io>: AI/ML project

# Community Colleges

## ... a crucial audience for NRP

- **CCC system dominates the total college student population in CA**
  - 70% of all of California's college students
- **CCC transfer students to CSU & UC drive social mobility in CA**
  - 30% of UC San Diego's incoming class are graduates from CCC
  - 50% of San Diego State Universities incoming class are graduates from CCC
- **California has strong path from K-12 to CCC to UC/CSU**
  - Public high schools in CA often have course offerings with their local CCC
  - High School students receive computing accounts from community colleges

**NRP presently working with  
San Bernardino, San Diego, and Los Angeles Community College Districts**



# Strategy for Scaling out

Engage with colleges from the network via the RENs  
and  
from the curriculum via ADSA (Academic Data Science Alliance)



CENIC and NRP Team visit LA Community College District to deploy hardware



**CENIC is a 501(c)(3) with the mission to advance education and research statewide by providing the world-class network essential for innovation, collaboration, and economic growth.**

**Charter Associates:**

- California K-12 System
- California Community Colleges
- California State University System
- Stanford, Caltech, USC
- University of California System
- California Public Libraries
- Naval Postgraduate School

2M students enrolled  
~10% of US college students

**Other Members Include:**

- Scientific and Cultural Institutions
- Private Colleges and Universities
- Hospitals and Specialized Medical Institutions
- Biomedical, Space, Environmental Research Organizations
- Tribal Nations

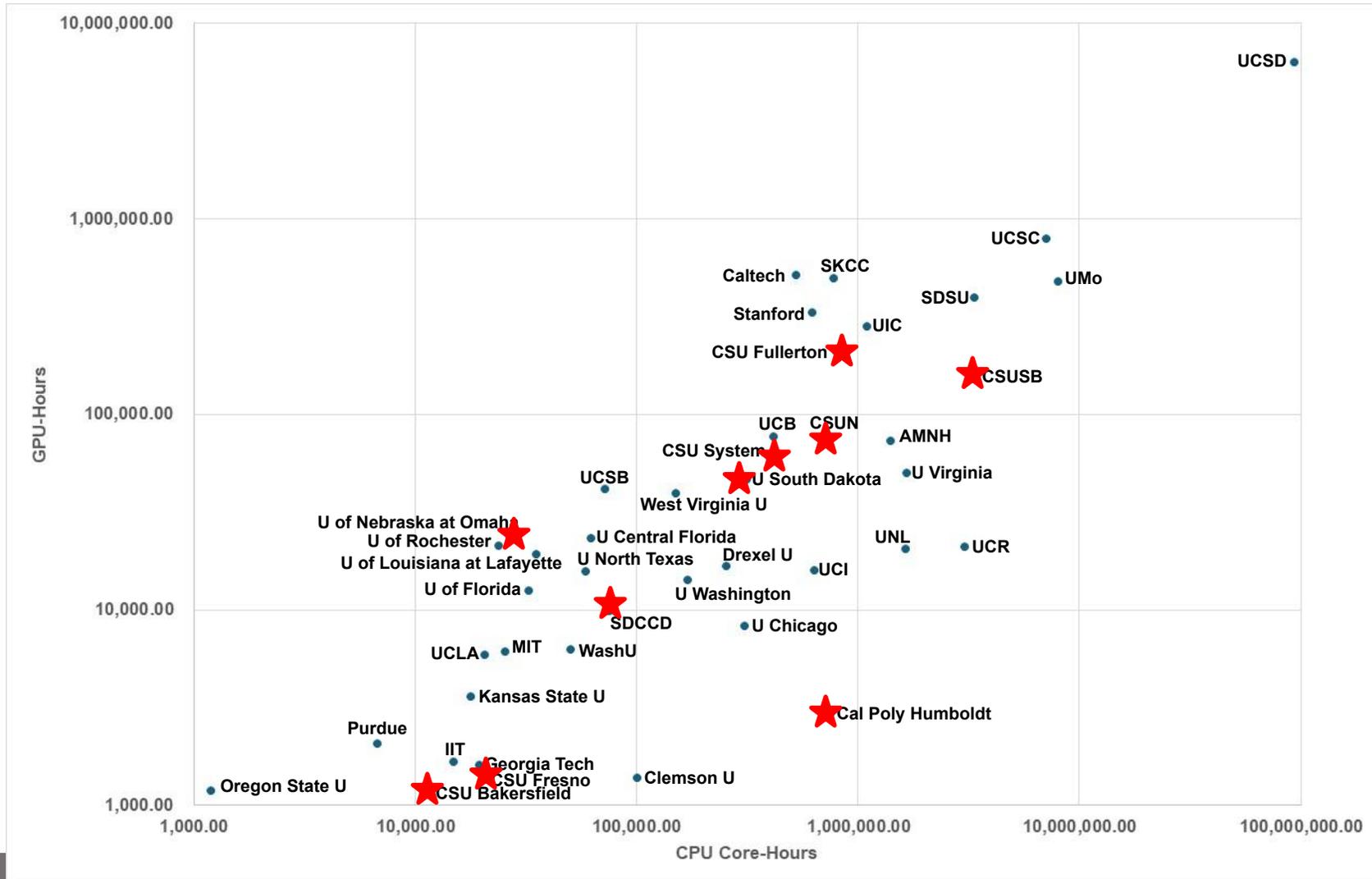


California  
STATE LIBRARY  
EDUCATION  
Preserving Culture, Inspiring Learning



# National NRP Faculty Usage by Campus Last Two Years

★ Non-R1 Institutions



# Research & Engagement w Industry

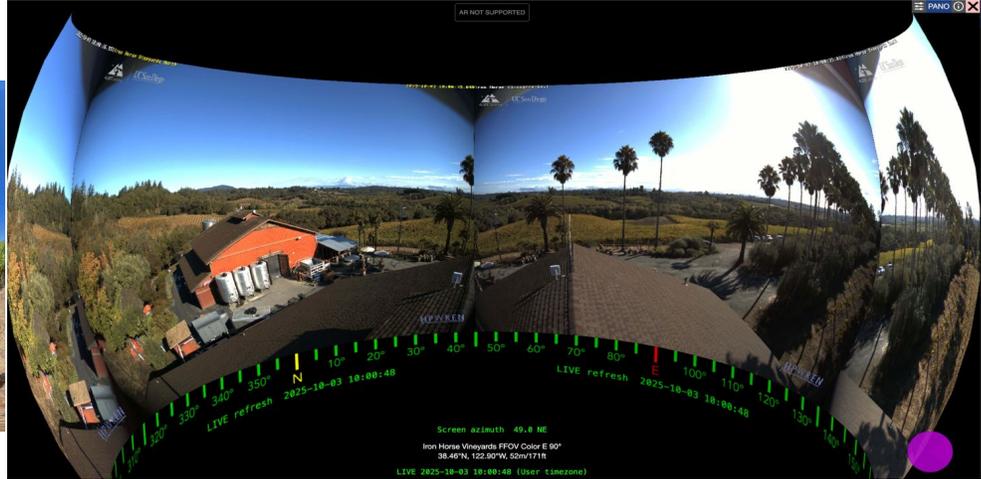
- Community Colleges and CSUs have a strong local focus
  - Serve local industries
- **We aspire to create research capacity & engagement with local industries via their ownership of AI Infrastructure**
- **Example:**
  - Agriculture Technology => Instrumenting Iron Horse Vineyards in Sonoma

<https://www.sdsc.edu/news/2025/PR20250113-data-driven-ag.html>

# Instrumenting Iron Horse Vineyards



Tom taking 360 video on e-bike



David, Emily & John Chose Sites for Moisture, Temperature and Conductivity Sensor w/LoRaWAN Transceiver & Solar Charger & CO2 Sensors

adding cameras for wildfire detection



# Path towards Sustainability

# We are in phase 1 of 3 phases

- **Phase 1:**
  - **Establish Concept and scale out to 100 colleges and 10,000 students/year**
  - Today we operate hardware at close to 60 colleges
  - Expecting to complete phase 1 in 2-3 years.
  - This “pilot phase” is funded by the National Science Foundation
- **Phase 2:**
  - **Scale out to 1,000 colleges and 1 Million students/year**
  - Scale out to 24x7 support for system & cybersecurity, but only 9-5 for user support & training
  - Expecting this to be a 5-10 year program
  - Estimating 10,000 GPUs needed to serve this population
- Phase 3: Sustainability through 501(c)(3) and membership fees
  - **At 1 Million students per year, even a modest fee of \$10/(student\*year) sustains the program**

Longer term: include high schools and public libraries

# Summary & Conclusions

- **Opportunity is a requirement for Social Mobility**
  - Education is the most effective guarantor of social mobility
  - Data & Compute is increasingly required for Education

**We provide a Data & Compute Platform for Higher Ed that is cost-effective and aspire to make it available to all colleges in the USA**

- **We aspire to engineer transformative change by creating a long term sustainable AI Infrastructure that colleges nationwide can own together**

# Acknowledgements

The NRP is supported by NSF grants OAC-1541349, OAC-1826967, OAC-2030508, OAC-1841530, OAC-2005369, OAC-21121167, CISE-1713149, CISE-2100237, CISE-2120019, & OAC-2112167

And by CENIC, Pacific Wave, MREN, GPN, NYSERNet, FLR, NEREN, SunCorridor, OARnet, SCLR, the Albuquerque GigaPoP, and Internet2

As well as a long list of Universities and colleges that host hardware, and share their hardware with the community.