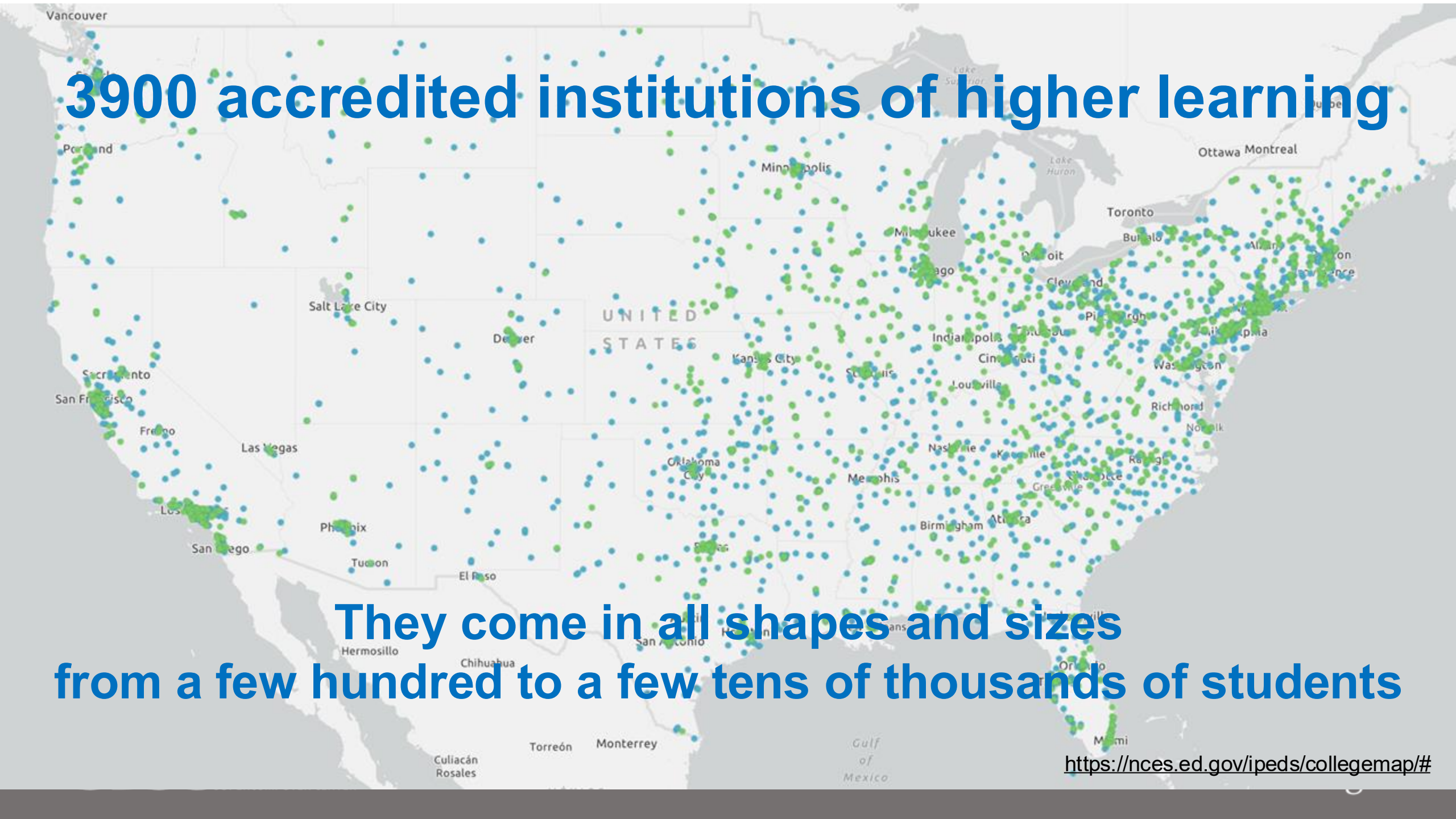
National Research Platform (NRP)

Incorporating PNRP NSF Cat II System

HPC and Data Science Summer Institute
August 8, 2025

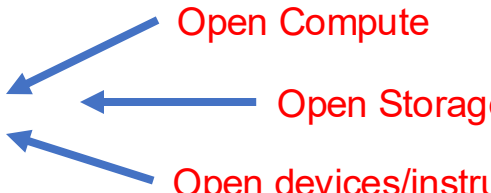
Mahidhar Tatineni
San Diego Supercomputer Center

Vision of NRP


3900 accredited institutions of higher learning
They come in all shapes and sizes
from a few hundred to a few tens of thousands of students
https://nces.ed.gov/ipeds/collegemap/#
Vancouver
Portland
Salt Lake City
Denver
UNITED STATES
Sacramento
San Francisco
Fresno
Las Vegas
Los Angeles
San Diego
Phoenix
Tucson
El Paso
Hermosillo
Chihuahua
Culiacán Rosales
Torreón
Monterrey
Minneapolis
Ottawa
Montreal
Toronto
Milwaukee
Detroit
Chicago
Cleveland
Buffalo
Pittsburgh
Kansas City
St. Louis
Indianapolis
Cincinnati
Louisville
Washington
Richmond
Norfolk
Nashville
Memphis
Oklahoma City
Birmingham
Atlanta
Miami
Gulf of Mexico

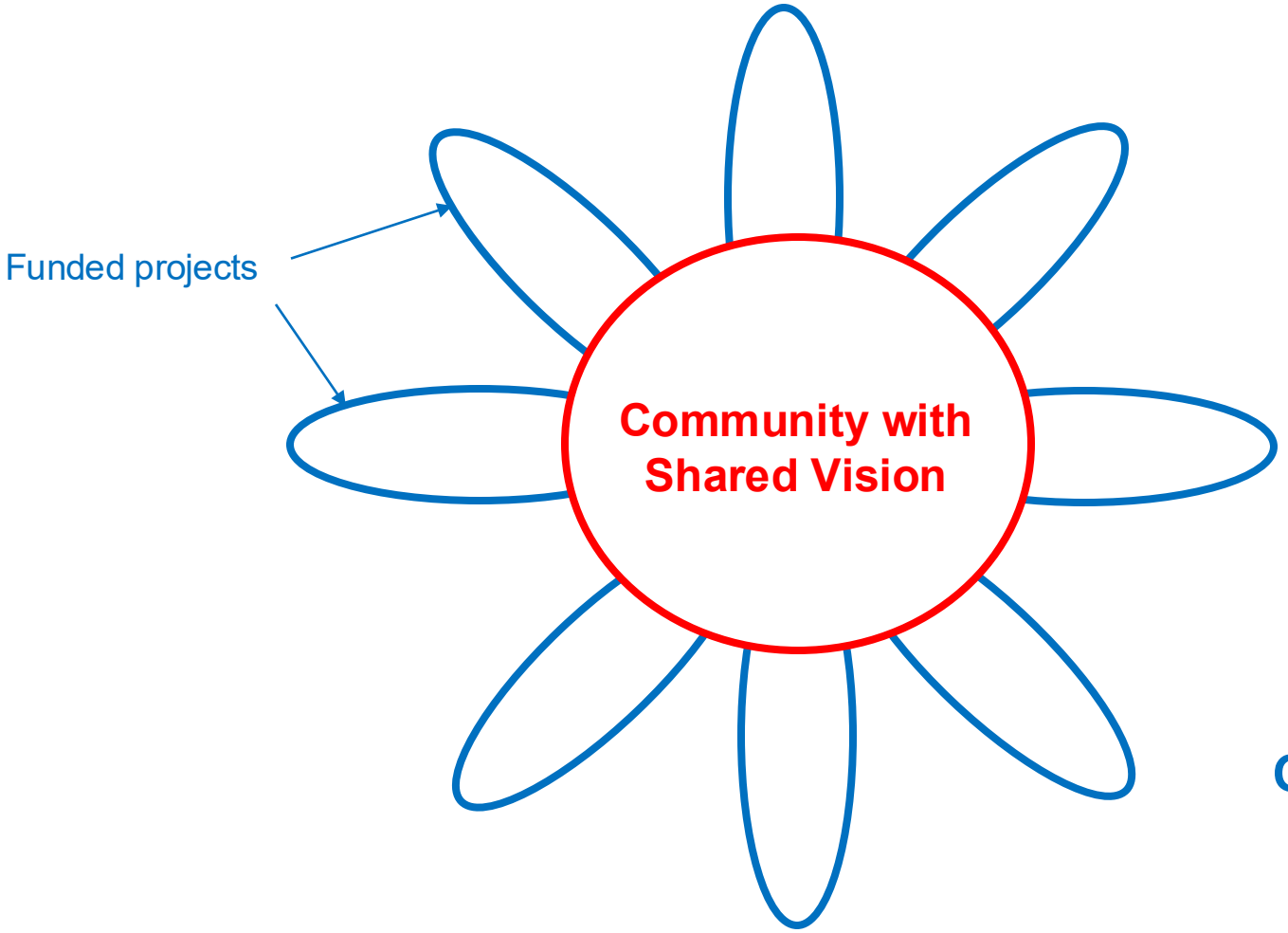
Long Term Vision

- Create an Open National Cyberinfrastructure that allows the federation of CI at all ~4,000 accredited, degree granting higher education institutions, non-profit research institutions, and national laboratories.
  - Open Science
  - Open Data
  - Open Source
  - Open Infrastructure



Openness for an Open Society

Community vs Funded Projects



Lot's of funded projects that contribute to this shared vision in different ways.

PNRP provides core services for the National Research Platform Community

Open Infrastructure is "owned" and "built" by the community for the community

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 only when necessary
  - Documentation and training for the community
- **Lots of software to reuse**
  - E.g. we operate JupyterHub for the community & show people how to deploy & customize their own
- **Maintain lots of topical Chat channels for the community to learn from each other and interact with each other**
  - Chat channels are monitored by professionals, but community is encouraged to interact and help each other
  - Starting to explore AI chatbots trained on the chats
- **LLM as a service**
  - We run about a dozen popular LLMs and provide popular APIs to upload any AI model from Hugging Face onto our resources.

PNRP Project

NSF Category II System

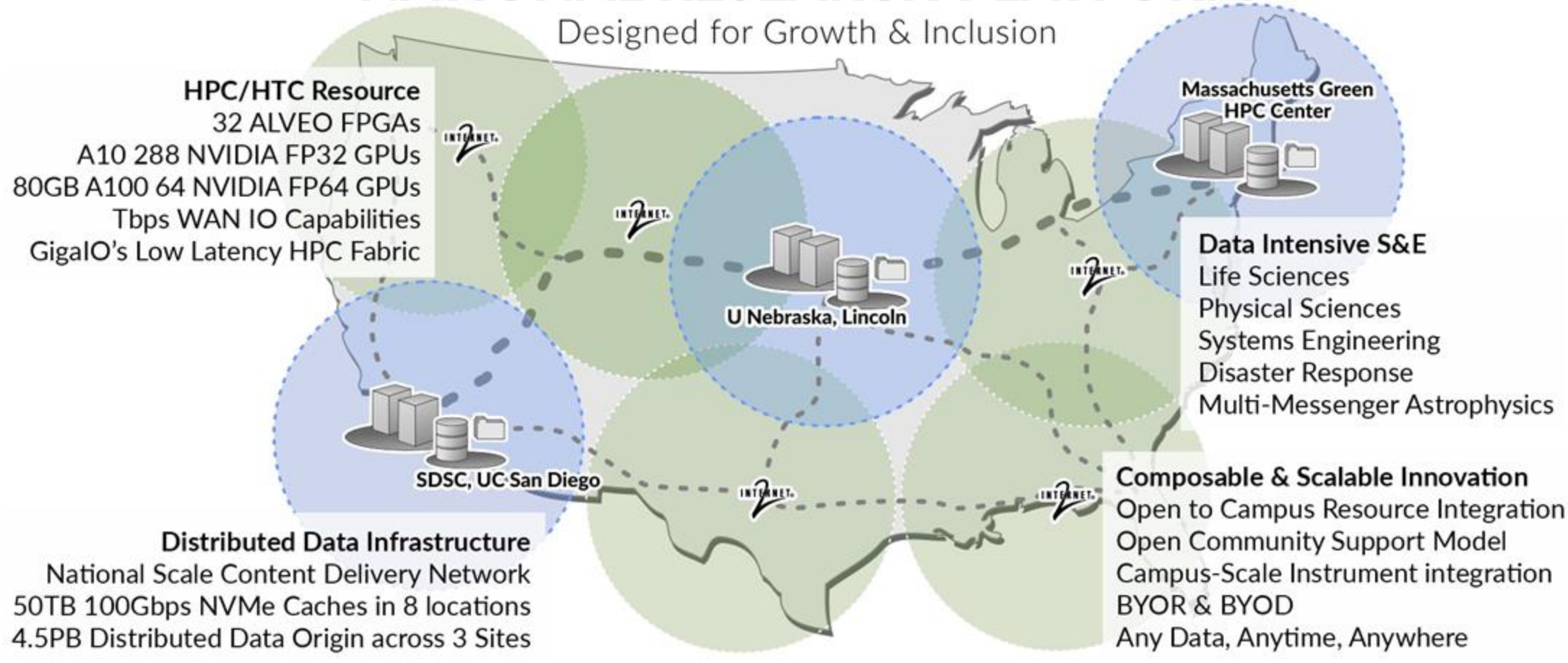
PNRP is designed & deployed as Category II innovative testbed

- NSF, Category II definition: “Resources proposed in this category will be initially deployed as a prototype/testbed system supporting S&E research through delivery of novel forward-looking capabilities and services”
  - 3-year Testbed Phase: Close collaboration with community to explore the PNRP system for transformational S&E
  - 2-year Allocation Phase: PNRP will be made available to the broader community via NSF supported allocation mechanism (e.g., ACCESS, PATh Facility, FABRIC, NAIRR, etc.)

- **Our proposal:**
  - A **nationally distributed system** with hardware deployed in 8 locations within the continental USA
  - An “**Open Cyberinfrastructure for all of Open Science**” …
    - Open horizontally via a Bring-Your-Own-Resources/Devices program
    - Open vertically for higher level service deployments on our Kubernetes layer
  - A composable hardware platform to support **R&D towards domain specific architectures**
    - PCIe hardware composability => allow for nodes that combine GPUs, FPGAs, and large volumes of NVMe

NATIONAL RESEARCH PLATFORM

Designed for Growth & Inclusion



5-year project: $5M for Acquisition and Deployment; $7.25M for Operations and Maintenance

PI = Wuerthwein; Co-PIs: DeFanti, Rosing, Tatineni, Weitzel

Funded as NSF 2112167

PNRP features innovations in processors, networking, system and data architecture, and operations

- I1: Innovative network fabric that allows disaggregated devices to behave like a single "node" connected via PCIe.
- I2: Innovative application libraries to expose FPGAs hardware to science apps at language constructs scientists understand (C, C++ rather than firmware)
- I3: A "**Bring Your Own Resource**" model that allows campuses nationwide to join their resources to the system.
- I4: Innovative scheduling to support urgent computing, including interactive via Jupyter.
- I5: Innovative Data Infrastructure, including national scale Content Delivery System like YouTube for science.

I3: Bring Your Own Resource (BYOR) innovation thrust of PNRP focuses on expanding the National Research Platform (NRP) Community

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

Select...

Sites	Nodes
84	**439**
Sites hosting NRP nodes	Nodes registered in Kubernetes
GPUs	**CPU Cores**
1,496	**29,254**
Total GPUs across all nodes	Total CPU cores across all nodes

NRP this month…

In January we were at 72 locations

UC San Diego

63 Campuses had active namespaces in 2024 on NRP

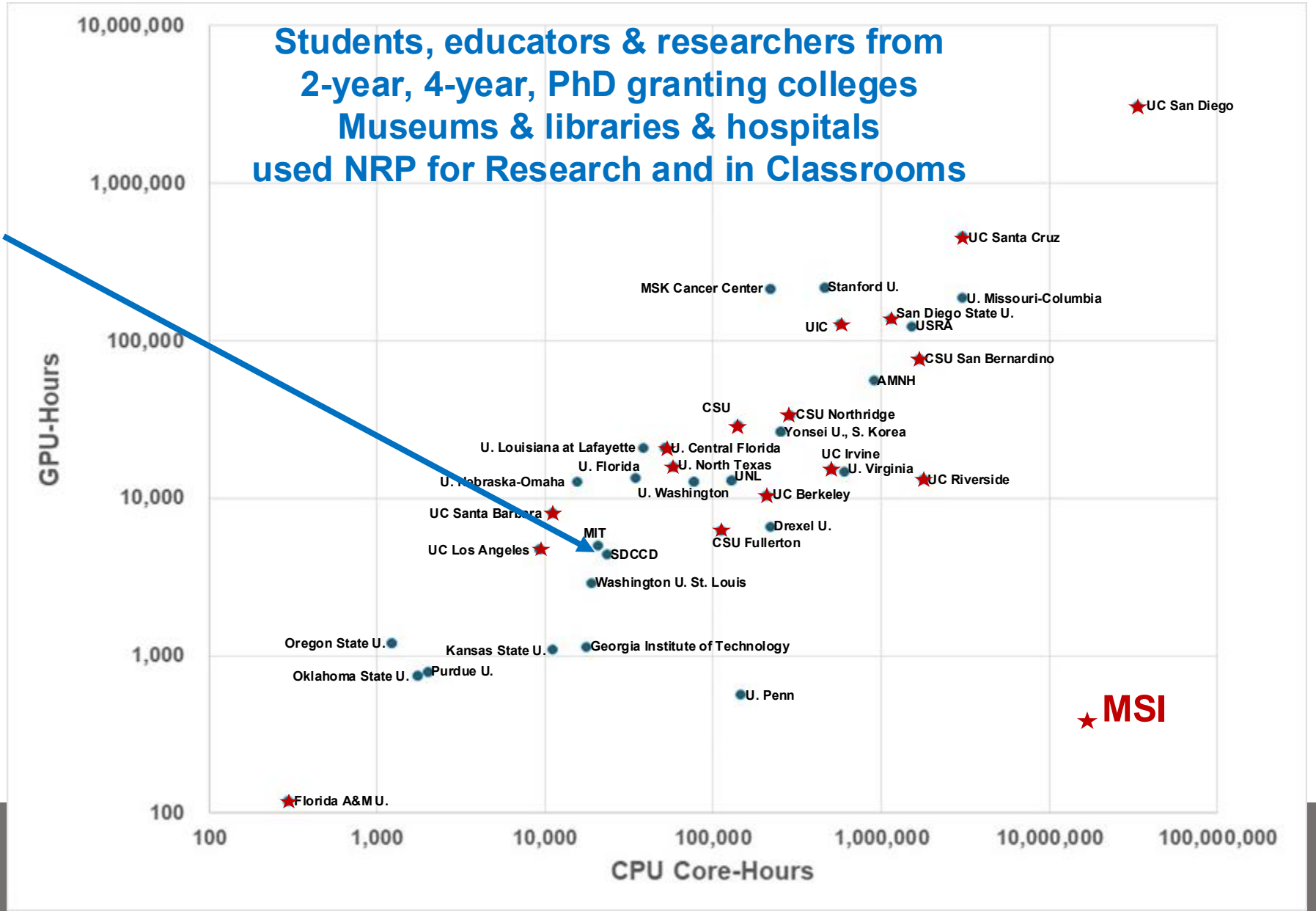
CENIC connected campuses:

San Diego Community College District

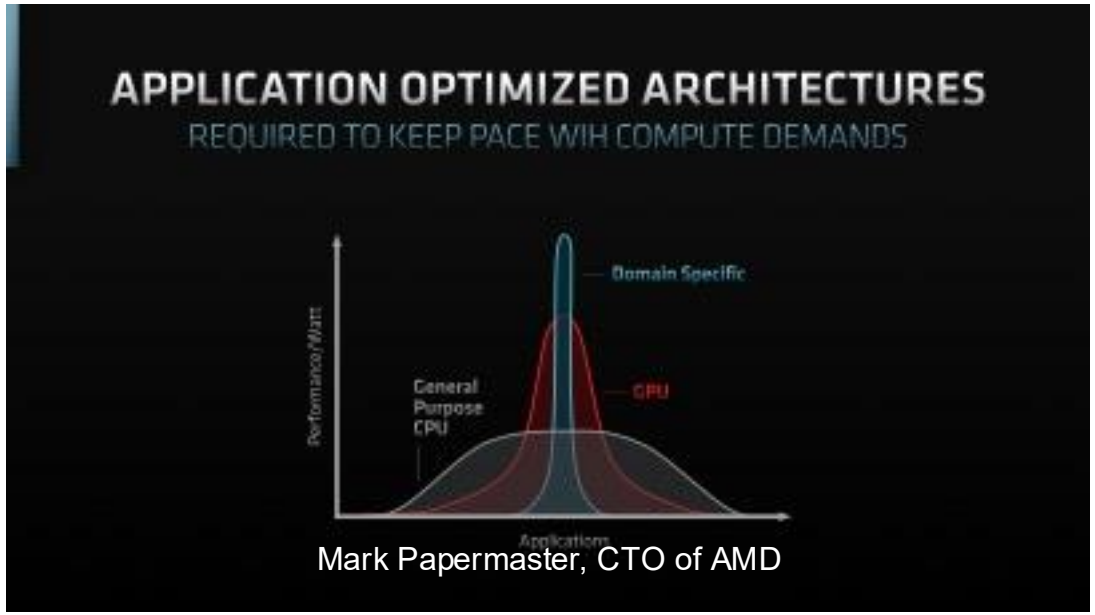
San Diego State University
CSU San Bernardino
CSU Northridge
CSU Fullerton
Cal Poly Humboldt

UC San Diego
UC Santa Cruz
UC Irvine
UC Riverside
UC Berkeley
UC Santa Barbara
UC Los Angeles

Stanford University



“end of Moore’s law” motivates new architectures



Performance improvements vs time slowed down by O(100)

I1 & I2 motivated by “end of Moore’s law”

PI, Tajana Rosing → **PRISM, a Jump 2.0 project funded by SRC is early user of FPGAs@NRP**

NRP supports FPGAs, P4 switches, NVIDIA DPUs & DGXs

Committed to be a “Playground” of technologies, easily deployed & operated via BYOR and BYOD.

Supporting Users with Multiple Development Environments

- We provide several development options for our users
- Gitlab – we have instructions for:
  - Building in Gitlab

  https://docs.nrp.ai/userdocs/development/gitlab/

  How to use your private repository

  https://docs.nrp.ai/userdocs/development/private-repos/
  - Integrate Gitlab and Kubernetes vi CI/CD jobs

  https://docs.nrp.ai/userdocs/development/k8s-integration/
- Coder open-source platform for creating and managing developer workspaces - used for FPGA code development; advanced code development on k8s cluster instead of local resources; develop with profiling level access; templates available for common dev setups
  - https://coder.com
- Jupyter-based development environment for FPGA codes with Xilinx tools installed

Classroom Support

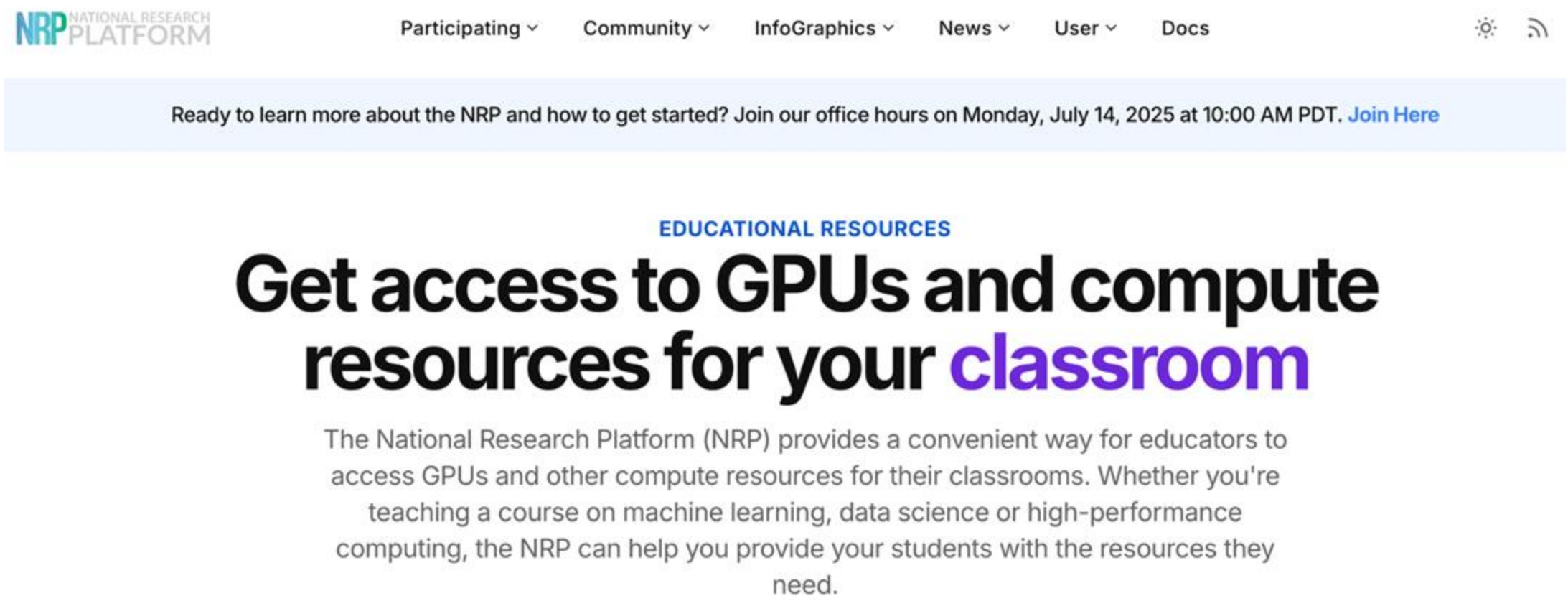
Slides Credit, from PEARC25 BOF:

Frank Wuerthwein: Director, San Diego Supercomputer Center
Derek Weitzel: Research Assistant Professor, School of Computing, University of Nebraska - Lincoln

The NRP Strategy for growth

- **Leverage existing social networks to engage colleges**
  - The Regional Research & Education Networks play a crucial role in that they are already engaged with the colleges in their regions, and with each other via the Quilt
- **Leverage the ongoing transformation of higher education** towards increasing use of data & compute as essential part of classroom education.
  - AI is an accelerant of a transformation that preceded ChatGPT
- **Reduce TCO of compute & data infrastructure for colleges**
  - We provide scalable remote system administration, user support, and training
  - Lean into the CC* program by the NSF to bring AI hardware to places that never had such capabilities before.
- **As we solve the AI education TCO pain point for colleges, we advance their research capabilities as a desired side-effect**

Education Website: nrp.ai/education



Education Resources

Educational features of the NRP

The NRP offers many features that are useful for educators. Here are some of the key features of the platform:

Hosted JupyterHub

No need to create your own JupyterHub, you can use the NRP's hosted JupyterHub to run your notebooks and share them with your students.

Learn more

Private JupyterHub

If you prefer to have your own private JupyterHub, the NRP can help you set one up for your classroom. Private JupyterHub allows you to customize the environment to meet your specific needs.

Learn more

Hosted LLM Service

The NRP provides a hosted LLM service with many open models to choose from.

Learn more

Access to GPUs

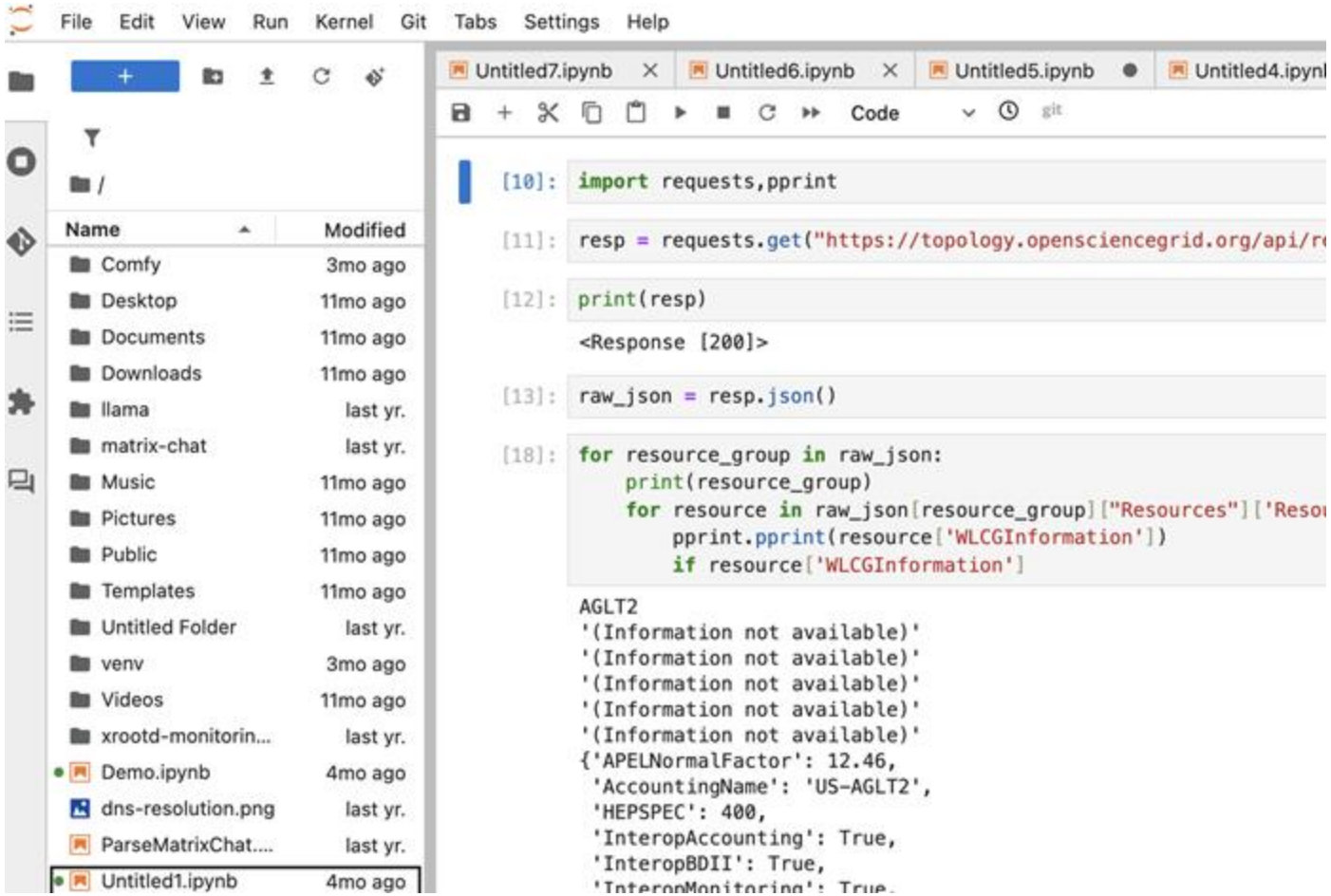
The NRP provides access to modern GPUs for use in your classroom. Whether you're teaching

Coder Hosted IDE

Coder is a powerful web-based integrated development environment, based on VS Code,

Hosted Jupyterhub

- NRP hosts a general purpose Jupyterhub
- Comes pre-built with many AI and ML images
- Professionally maintained and monitored for uptime and issues
- Configurable number of GPUs, and types of GPUs



Private Jupyter - 113 In CY 2024

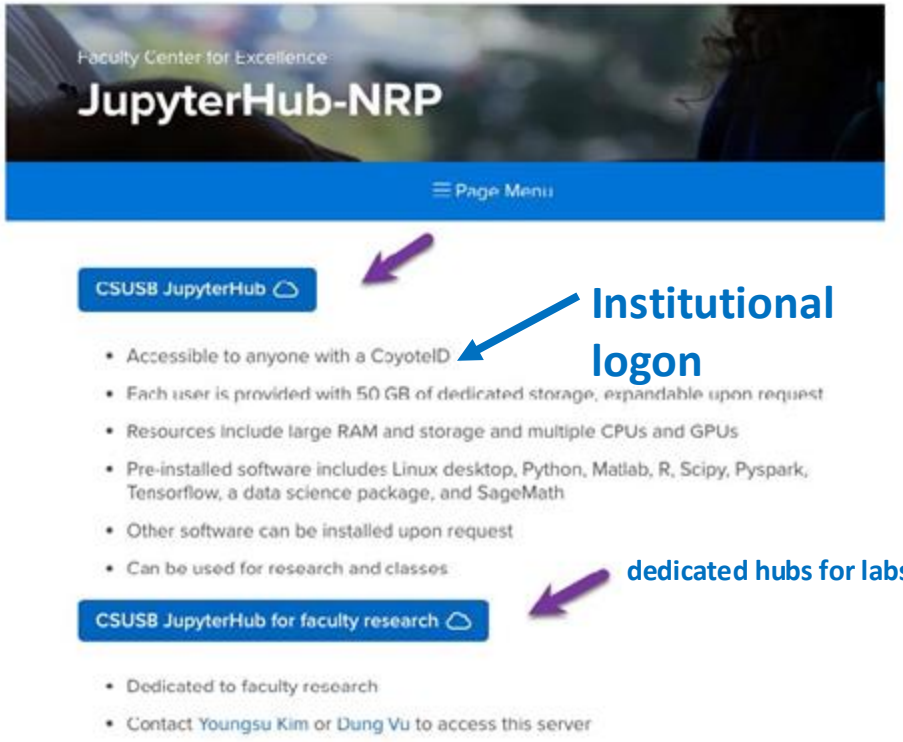
Same Jupyter software as the hosted version, but:

- The instructor controls the images and can “bake” in software / lessons
- The instructor controls access
- Can configure larger (or smaller) default storage as required for the course
- Instructor can control the requested resources

But

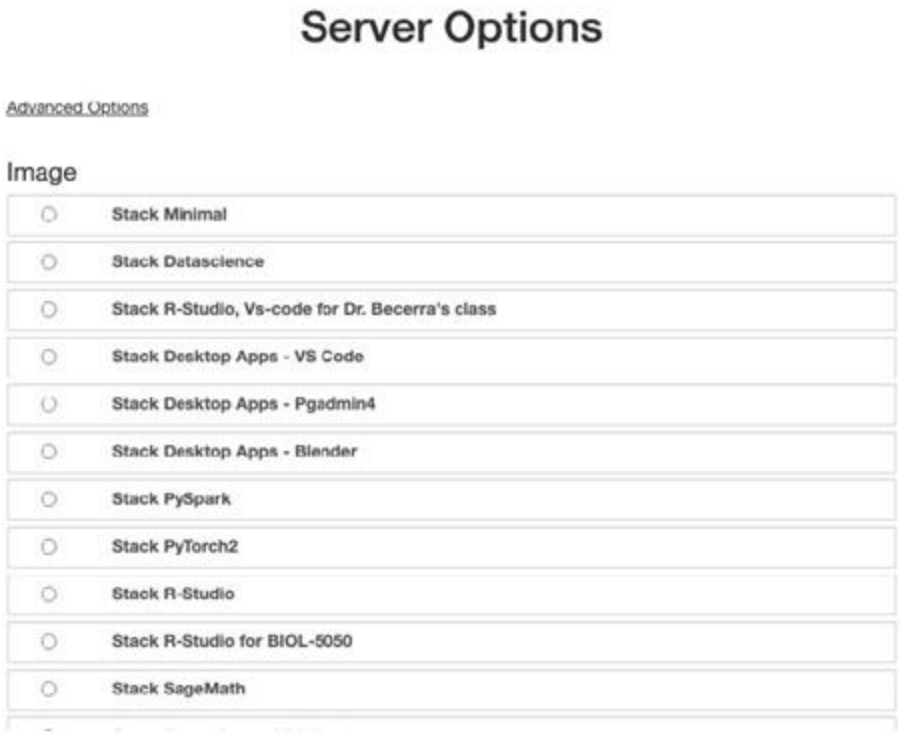
- Not professionally monitored or maintained
- It’s on the instructor to configure and maintain the Jupyterhub (though, we can help)

JupyterHub User Interface at CSU, San Bernardino



Institutional logon

dedicated hubs for labs



Wide range of dedicated hubs

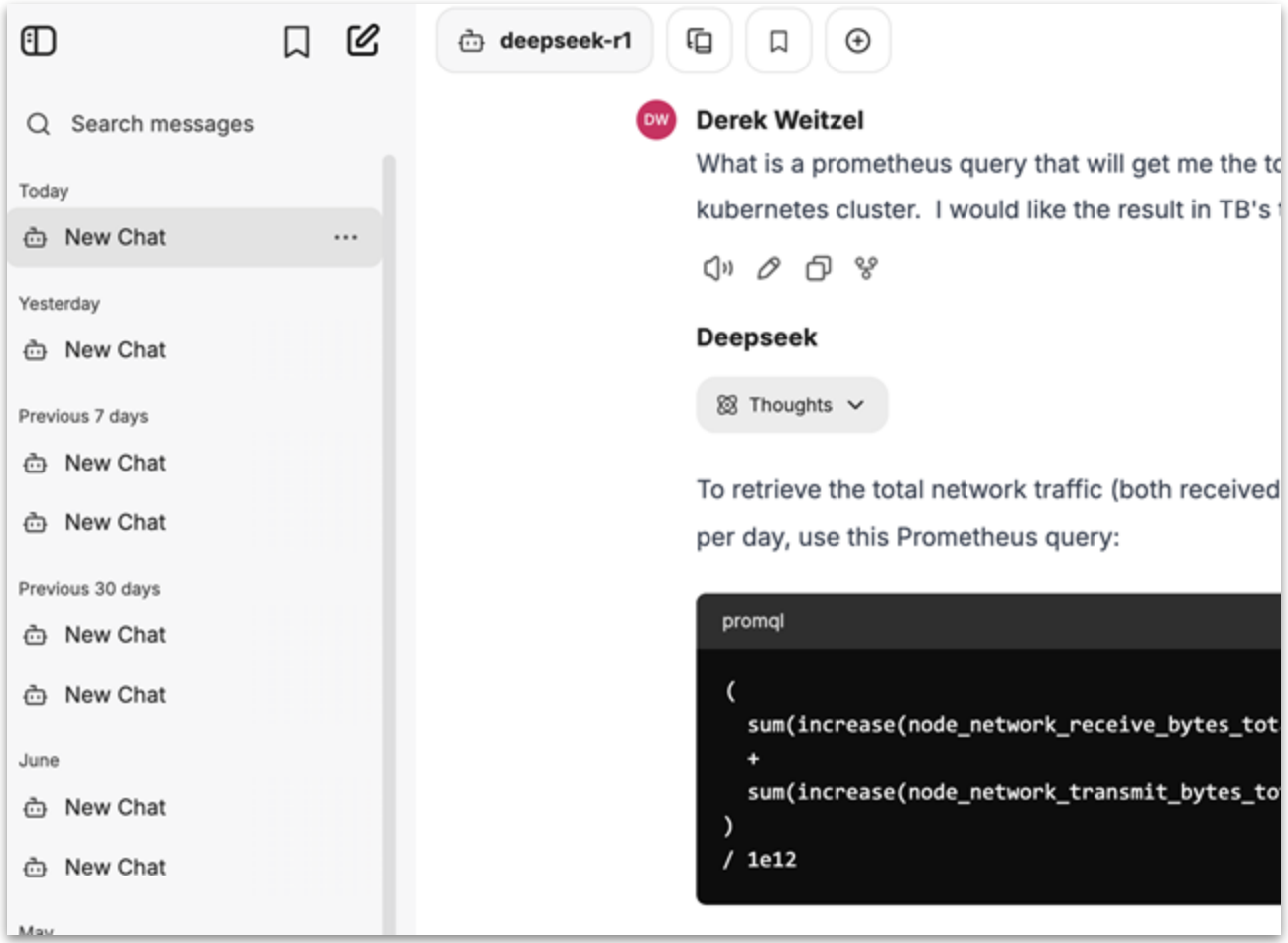
https://csusb-metashape.nrp-nautilus.io: 3D modeling
https://csusb-vasp1.nrp-nautilus.io Viena Ab initio Simulation package (VASP)
https://csusb-cousins-lab.nrp-nautilus.io: VASP simulation
https://csusb-becerra.nrp-nautilus.io AI/ML project
https://csusb-biol-5050.nrp-nautilus.io: Biology course
https://csusb-cse-salloum.nrp-nautilus.io Summer Research
https://csusb-drhamoudahub.nrp-nautilus.io Data Analytics
https://csusb-ratnasingam.nrp-nautilus.io Data Analytics
https://csusb-zhang.nrp-nautilus.io AI/ML project

Hosted LLM Service

The NRP hosts some open-model LLMs available

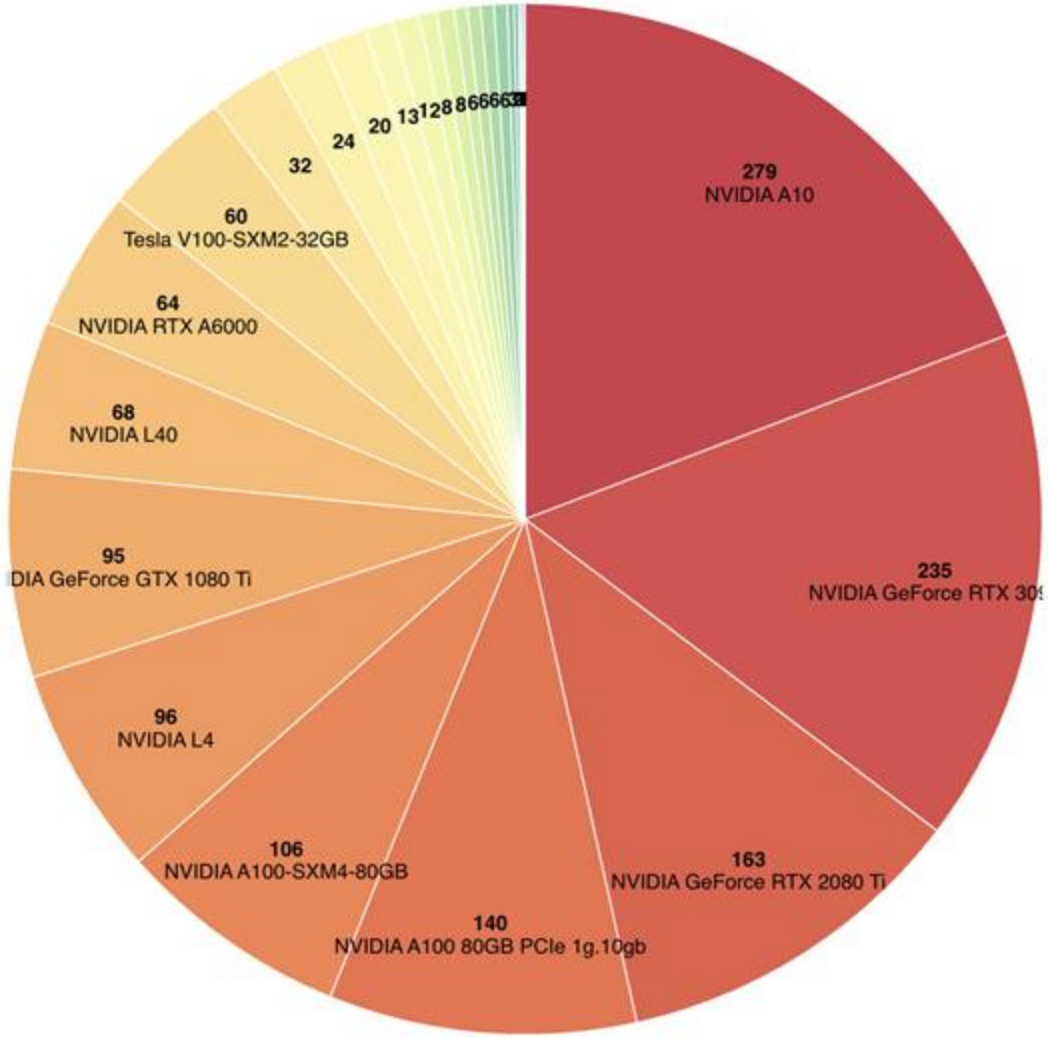
Query through our hosted chat interfaces.

Or… use an API to perform many queries.



GPUs - Kubernetes

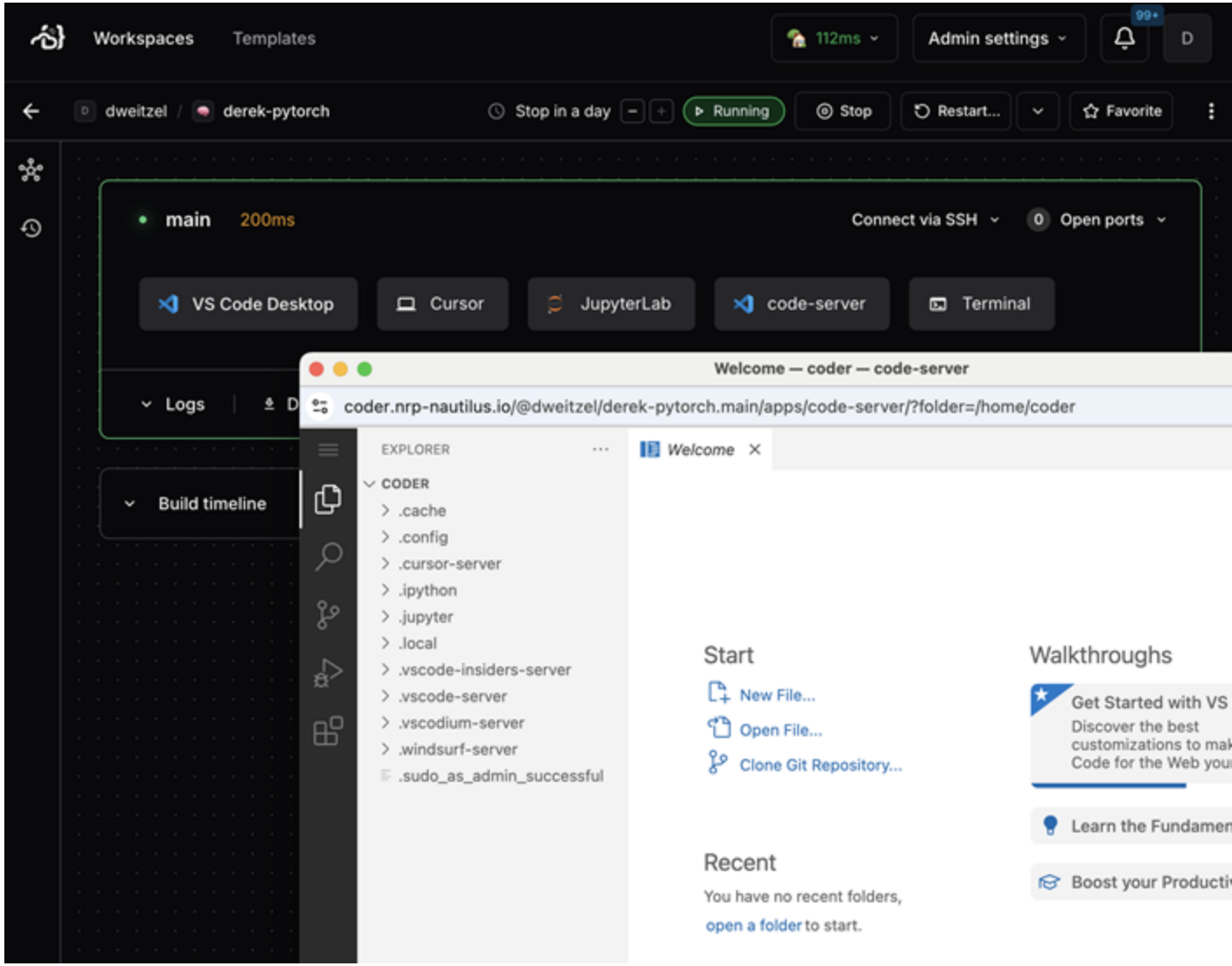
GPU Type	Quantity
NVIDIA A10	279
NVIDIA A100 80GB PCIe	266
NVIDIA GeForce RTX 3090	235
NVIDIA GeForce RTX 2080 Ti	163
NVIDIA L4	96
NVIDIA GeForce GTX 1080 Ti	95
NVIDIA L40	68
NVIDIA RTX A6000	64
Tesla V100-SXM2-32GB	60
NVIDIA RTX A4000	32
NVIDIA GeForce RTX 4090	24
NVIDIA TITAN Xp	13
...	
Total	**1455**



Coder

Provides an on-demand jupyter/VS Code environment.

Useful for GPU/FPGA development where you need the GPU/FPGA locally to run tests.



PNRP offering in National AI Research Resource (NAIRR) Classroom

- Jupyter resources for classes that need AI/ML – can be teaching about AI/ML or using AI/ML in domain science
- Can use institutional authentication methods leveraging InCommon
- Professor and TA will have admin status to create namespaces for classes
- Support/Community interaction through Matrix channels – one for the class that includes students so Professor/TA can help them and second for Jupyter admins that includes admins from other classes
- Direct ticketing system available for any issue needing longer follow ups
- GPU nodes each with 8 A10 GPUs, 512GB of RAM, 2 AMD EPYC 7502 CPUs, and 8TB of NVMe.
- Depending needs, other GPU resources can be provisioned via the NAIRR pilot and made available through this platform

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.