

National Resources for Al and Large-scale Computing



Who are these resources for?

You! (and your campuses)

No experience necessary! *No cost to you!*

- Researchers and instructors from U.S.
 2- or 4-year academic institutions or non-profit organizations
 - (sometimes even broader eligibility in some cases)
- Any or no source of funding for the research
 - (some limitations in specific cases)
- Any application domain of AI
 - And non-Al work, in many cases
 - (some limitations in specific cases)



What do you need to get started?

- You need to bring the idea for the work and the ability to carry it out
 - While resource access is no cost, you need to cover costs for staff time
- Information about your project
 - Title & Abstract
 - Personnel
 - Project proposal or description
 - Instructions vary depending on the scale of work and the program providing resource access
 - Supporting grant information
 - If applicable to your project

) C 🗱 submit-nairr.xras.org/opportunities/534068/requests/new 🖄 🖞 🛃
🕲 NSE Next-Gen ES 🙆 ACCESS - RAMPS 🚺 My Plaid Student - 🛛 🤻 RAMPS JIRA 🌒 SciENcy: Science 🕲 XACCT Admin 💩 T1-RAMPS Shured 👋 🗅 Al E
Request Information
Request Title: *
Enter request title here
Public Abstract:
Type your public abstract here
Keywords (separated by commas):
Enter keywords separated by commas like astronomy, stars
Fields of Science
To add a field of science select one from the list and check whether or not it is a primary one. To add more click "Add Another Field of Science". To remove
one click "Remove".
Primary Please select one V Remove
Add Annahan Didd of Calmer
Add Another Held of Science
Project Personnel
To add personnel to this request, first select one or more roles from the selection list, then search for a user by using her ORCID id or name. Search will only
work if the individual has already logged into this site with their ORCID id. If no matches are found, use the "Create New User" button to enter this user's
information and associate them with the role(s). Any added user will be shown in a table below the search field. To remove a particular user click on Remove when the personnel table is visible. A project can have only one Project Lead.
Beniest Land
Project Lead Project Admin
Search for people
Search for people Find people by ORCD ID, last name, or first name, e.g., "j smith"
Search for people Find people by ORCD ID, last name, or first name, e.g., "j smith" Create New Uter
Search for people Find people by ORCD ID, last name, or first name, e.g., "j smith" Create New User
Search for people Find people by ORCID ID, last name, or first name, e.g., "j smith" Create New User Other Collaborators
Search for people Find people by ORCD ID, last name, or first name, e.g.,") smith* Create New User Other Collaborators Include name, institution, and email for all collaborators listed.





NSF ACCESS



What is NSF ACCESS?



ACCESS has been established by the U.S. National Science Foundation (NSF) to connect researchers and educators to the resources and services they need to accomplish their objectives.

access-ci.org allocations.access-ci.org



Allocations on NSF ACCESS

- Visit https://allocations.access-ci.org/
- Select Get Your First Project Here in the first box
- You'll typically have an Explore ACCESS project within a day or so
- Many projects go from submitting their project request to completing their first resource job in about 10 days.
- Select *Resources* from top navigation menu to explore resources

If you're new to nationalscale resources, you need to tell the provider about your project to get an "**allocation**" to use a resource. It's an amount of resource units (core-hours, GPU-hours, etc.) that you have permission to spend to pursue the goals of your project.



NSF ACCESS Multi-core Compute

• Anvil (Purdue)

Delta (NCSA)

- 1,000 AMD Milan nodes, 128 cores per node, some large memory nodes
- Bridges-2 (PSC)
- 504 AMD Rome nodes, 128 cores per node, large memory nodes available; extreme memory (4 TB) nodes allocated separately
- 124 AMD Milan nodes, 128 cores per node
- Expanse (SDSC) 728 AMD Rome nodes, 128 cores & 1 TB NVMe per node
- KyRIC (U Kentucky) Five large-memory (3 TB, 6 TB) nodes, 300 TB storage
- Launch (Texas A&M) 35 AMD EPYC 9124 nodes
- Stampede 3 (TACC) 1,848 nodes with Intel Sapphire Rapids, Ice Lake, and Skylake



NSF ACCESS GPU Computing

- Anvil GPU (Purdue) 16 nodes, 4 NVIDIA A100 GPUs each
- Bridges-2 GPU (PSC) 33 nodes, 8 NVIDIA V100 GPUs & 7.68 TB NVMe per node
- Delta GPU (NCSA) 4 node configs: 100 nodes w/ 4x A100s; 100 w/ 4x A40 GPUs; five w/ 8x A100s; one w/ 8x AMD MI100 GPUs
- **Delta AI (NCSA)** 114 nodes, each with four Grace Hopper chips
- Expanse GPU (SDSC) 52 nodes, 4 NVIDIA V100 GPUs each



NSF ACCESS Novel / Innovative Computing

- ACES (Texas A&M U) Composable PCIe fabric with Intel Sapphire Rapids cores, Graphcore IPUs, NEC Vector Engines, Intel Max GPUs, Intel FPGAs, Next Silicon co-processors, NVIDIA H100 GPUs, Intel Optane memory
- FASTER (Texas A&M U) 180 nodes on a composable fabric, 2x Intel Ice Lake processors each, 260 NVIDIA GPUs (five different architectures)
- Jetstream2 (Indiana U) Cloud environment with AMD Milan nodes, and 90 nodes with 4x A100 GPUs

- Neocortex (PSC) 2 Cerebras CS-2 Wafer Scale Engine systems
- Ookami (Stony Brook U) 176 nodes with Riken/Fujitsu A64FX processors; additional nodes with AMD Milan, Thunder X2, and Skylake/V100 architectures
- **Voyager (SDSC)** 42 Intel Habana Gaudi training nodes, each with 8 training processors

allocations.access-ci.org/resources

NSF ACCESS Storage

Storage options also available for most compute systems, and capacity is awarded alongside compute allocations for those resources. The following storage resources can be requested separately.

- **Granite (NCSA)** 19-frame Spectra Tfinity tape library; 3.6 PB available for ACCESS allocations
- **Open Storage Network (OSN)** Cloud object storage resource, comprised of geographically distributed pods, accessed via S3 interfaces
- Ranch (TACC) Large-scale, tape-based archival storage system



allocations.access-ci.org/resources

NSF ACCESS Affiliated Resources

Beyond the preceding resources, which are all allocated via ACCESS-managed processes, you can also find out about additional resources and services via the ACCESS Resource Catalog

- **Open Science Pool —** High-Throughput Computing environment, no allocations necessary
- **FABRIC** Distributed infrastructure with configurable in-network computing capabilities
- Sage Al@Edge cyberinfrastructure with deployed and configurable sensor networks
- **SGX3** Science Gateway consulting and services

ACCESS



Your On-Ramp to NSF ACCESS

- On-Ramps are a new feature developed by the ACCESS Allocations team that lets you bring ACCESS resource information to your campus community.
- Our goal was to help researchers find out about ACCESS in the first place they go for IT information their campus websites.





How It Works

- Visit our <u>On-Ramps page</u>
 - allocations.access-ci.org/on-ramps
- Copy the 17 lines of JavaScript
- Paste the JavaScript into your page about ACCESS, and save

Content is dynamically generated from ACCESS sources, so maintenance is not an issue.

You'll always have the latest info.



"Baylor ITS Research Technology announces Baylor's participation in the National Science Foundation's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program. ..."



On-Ramps in the Wild

- Rocky Mountain Advanced **Computing Consortium**
- **Baylor University ITS** .
- Illinois Computes
- Jackson State University .
- University of Hawai'i .
- Texas A&M University
- University of Central Florida
- Columbia University
- University of Kentucky
- Keystone REN
- University of Houston .
- **Boston University**
- Texas Tech University
- MS-CC .
- **NSF NCAR CISL** .
- East Tennessee State University
- Stanford University .
- **UC Riverside**





The NSF ACCESS (Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support) program connects researchers and educators with the nation's most powerful supercomputers and data systems at no cost. Whether you're conducting large-scale simulations, analyzing massive datasets, or exploring new frontiers in computation research. NSF ACCESS can elevate your work.

other cyberinfrastructure services to tackle complex computational challenges

Explore NSF ACCESS Support and Resources

Data Management

Amazon Web

NSF ACCESS

Resources

Contact Us

Services



🔕 NSF Next-Gen TS. 🖉 ACCTSS - RAMPS. 🚺 My Flad Student - 🥒 RAMPS JRA 関 SciEncy. Science. 🕲 XACCT Admin 💩 TI-RAMPS Shared

* 5 0

IF C AT BOOK

Q

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

CYBERINFRASTRUC TURE RESOURCES

NSF ACCESS Resources/NCAR

Discover local entryways to the nation's cyberinfrastructure. On-Ramps is an online tool to browse available ACCESS resources to facilitate an increasing educational and research community for utilization of its full range of HPC options that provides educators and researchers with powerful HPC resources when and where they need them. It extends the ACCESS program's reach, and empowers community members by giving them a larger toolbox to choose from.

RESEARCH AND CYBER COMPUTING

Information

Alumni



NAIRR Pilot



What is the National Al Research Resource (NAIRR) Pilot?

- NAIRR Task Force established by National AI Initiative Act of 2020, launched in June 2021, co-chaired by OSTP and NSF
- NAIRR Task Force's final report issued in Jan. 2023, providing a roadmap for standing up a national AI research infrastructure
- White House issued Executive Order on Oct.
 30, 2023, with 90-day window to launch NAIRR
 Pilot
 - Among many AI-related directives to federal agencies





nairrpilot.org/opportunities/allocations

Allocations in the NAIRR Pilot

- Visit <u>https://nairrpilot.org/</u>
- Under "Current Opportunities," select
 - Researcher Resources Call, for research projects
 - Classroom/Educators Resources Call, for classroom activities
- Requests require a 3-page proposal
 - See website for proposal instructions
- Feel free to submit a help ticket to NAIRR
 - If you have proposal or resource questions

🗧 😑 🔹 🔯 NAJRR Pilot - Home	× +			`
← → O ts naimpilot.org				* 0 🧶 1
88 Ø NSF Next-Gen ES_ 🛆 ACCES	S - RAMPS 🚺 My Plaid Student 🛹 RAMPS JIRA 🚯 SciEM	icv: Science 🚷 XACCT Admin 🛕 T1-RAMPS 5h	ared 🤷 NCAR KPIs 2023 📴 Daily B Items (Dav	>> 🗀 All Bookmarks
NAIRR Pilot	National Antificial Intelligence Research Researce Pilot The NAIRR Pilot aims to conn training resources needed t agencies are collaborating v implement the Pilot as a pro	nd Anfhiod Intelligence ch Resources Plot Researcher Resources Call Classroom Educator The NAIRR Pilot aims to conn training resources needed to open Data, Models, agencies are collaborating v and More implement the Pilot as a pro- n eventual full NAIRR implementation.		t myNainAllocation
	250+ Research projects supported	40 States + DC repr	resented	
	22 NAIR Classroom awards	20 Infrestructure and data demonstration projects	Community workshops (more scheduled)	
	Learn more about NAIRR Pilot Subscribe for			



NAIRR Pilot Private Sector Resources

These resources represent leading-edge offerings from corporate and non-profit organizations. So new we're still trying to decide how to classify them.

Cloud Providers Amazon Web Services Google Cloud Platform Microsoft Azure

GPU Systems NVIDIA DGX Cloud Training Systems

Cerebras CS-2 SambaNova Suite

Inference Services

Anthropic Groq LPU Inference Engine OpenAI SambaNova Cloud **Tools and Software**

DataBricks

Eleuther Al

Hugging Face

OpenMined

Weights & Biases





nairrpilot.org/opportunities/allocations

NAIRR Pilot GPU and CPU Resources (federal)

These resources comprise mostly homogeneous partitions or are dominated by a single processor type. The number and types of hardware vary from resource to resource.

<i>AMD Milan</i> Purdue Anvil CPU TACC Lonestar-6	AMD Rome PSC Bridges-2 CPU SDSC Expanse CPU	<i>Intel Cascade Lake</i> TACC Frontera	Large-Memory Nodes PSC Bridges-2 EM (Extreme Memory)	
TACC Frontera GPU 360 RTX-5000s	48 A100s PATh Facility A100s & AME) Milan	Purdue Anvil Al 80 GH H100s	
SDSC Expanse GPU 208 V100s	Purdue Anvil GPU 64 A100s TACC Lonestar-6 GPU		NCSA DeltaAI 456 GH H100s	
NVIDIA V100 PSC Bridges-2 GPU 280 V100s	NVIDIA A100 / A40 NCSA Delta GPU 400 A100s & 400 A40s		NVIDIA H100 / Grace Hopper TACC Vista 608 GH H100s	

ACCESS

nairrpilot.org/opportunities/allocations

NAIRR Pilot AI, Cloud, and Composable Resources

The NAIRR Pilot includes resources that offer alternative architectures designed to support a wide range of needs beyond the more conventional HPC architectures.

AI Accelerators

PSC Neocortex

Two Cerebras CS-2 systems

SDSC Voyager

42 8x Intel Habana Gaudi training nodes

DOE ANL AI Testbed

four different architectures

- Cerebras CS-2
- SambaNova DataScale SN30
- Graphcore Bow Pod 64
- Groq LPU Inference Engine

Cloud

Indiana Jetstream-2 GPU

360 A100 GPUs

Composable & Distributed

Texas A&M U ACES

composable system with a wide range of processor and accelerator options

FABRIC

distributed compute / network platform



NAIRR Pilot Classroom Resources

The NAIRR Classroom opportunity provides resources aimed at undergraduate or graduate courses or shorter training sessions that include AI subject matter.

NIH Cloud Lab

- Targeted to instructors in biomedical courses
- Up to 90 days of access to AWS, GCP, or MS Azure, plus up to \$500 cloud credits / student

Prototype National Research Platform (PNRP)

• Offers GPU resources for your students and a Jupyter platform

Vocareum Al Notebook

• Licenses for advanced, training-oriented, Jupyter-based notebooks

Indiana Jetstream-2 GPU

• Provides flexible, on-demand, programmable tools and services







NSF National Center for Atmospheric Research (NSF NCAR)

Allocations at NSF NCAR

• SMALL

- Requires NSF research award
- 2,000,000 core-hours **OR** 5,000 GPU-hours

• EXPLORATORY

- Support for grad students, post-docs, certain unfunded activities
- 1,000,000 core-hours **OR** 3,000 GPU-hours

CLASSROOM

- Support for classroom, training activities
- Same limits as Exploratory projects
- No funding constraints

• DATA ANALYSIS

- Requires need to analyze NCAR-hosted data set
- Casper only (no Derecho access)
- No funding constraints

LARGE UNIVERSITY PROJECTS

- Require NSF research award
- More than 2 million core-hours OR more than 5,000 GPU-hours
- Reviewed twice annually by panel
- Next deadline: September 2025

All university allocations require

- US-based project lead
- Academic or non-profit institution affiliation for the project lead
- Must be work in the Earth system sciences or related activities



arc.ucar.edu/allocations

NSF NCAR's Compute and Data Services

Derecho

- 2,488 AMD Milan nodes, each w/128 cores and 256 GB
- 328 40GB A100 GPUs (82 4-GPU nodes)
- 60 PB scratch file system

Casper

- For high-throughput workflows, data analysis, and AI/ML
- ~100 V100, A100, and H100 GPUs

NCAR Data (data.ucar.edu)

- Find datasets in the more than 10 PB of curated data collections housed at NSF NCAR
- Download data or analyze on Derecho & Casper





arc.ucar.edu/allocations



And That's Not All



Other Resource Options

Beyond the programs and sites already mentioned, there are yet more options for those in need of computing support. Many of these resources are aimed at large-scale problems.

- Frontera & LCCF Texas Advanced Computing Center (TACC), <u>https://tacc.utexas.edu/</u>
- Department of Energy National Labs
 - National Energy Research Scientific Computing Center (NERSC), <u>https://nersc.gov/</u>
 - Argonne Leadership Computing Facility (ALCF), https://www.alcf.anl.gov/
 - Oak Ridge Leadership Computing Facility (OLCF), https://www.olcf.ornl.gov/
- Closer to home
 - A number of regional organizations provide access to research computing and data services.



Research-Ready Software Environments

- Most sites have standard HPC and AI software tools ready to go
- Many sites now hosting open-source LLMs
 - Llama, DeepSeek, and more
- Most sites willing to install open-source software upon request
 - Some commercial software available, esp. if you have a paid license





More Than Just Hardware

- All the resources mentioned are "full service" — not only hardware but also support teams, training courses, and related activities.
- Many separate training and support efforts being offered through the NSFfunded CyberTraining and SCIPE programs.
- Start with the program or provider for resource-specific training.
- Try <u>HPC-ED</u> for general training resources.







QUESTIONS?

David L. Hart dhart@ucar.edu

