CENIC AIR

CENIC ARTIFICIAL INTELLIGENCE RESOURCES

# Frequently Asked Questions

This FAQ answers questions about the CENIC AI Resource both from faculty members who would like to use CENIC AIR for their research and instruction and from senior campus leaders and IT directors responsible for architecting and deploying network infrastructure.

> **An Introduction to CENIC AIR:** A brief definition of the CENIC AI Resource (CENIC AIR), why it was created, what it consists of, and how it relates to the National Research Platform (NRP).
>
> **CENIC AIR for Research and Instruction:** This section addresses questions from researchers and faculty at universities and colleges on how to access the computers and storage provided by CENIC AIR institutions.
>
> **Connecting Your Institution to CENIC AIR:** This section addresses infrastructure-related questions about connection bandwidth, implementing a Science DMZ, adding compute and storage resources, obtaining funding, and the roles and responsibilities of your institution, CENIC, and the National Research Platform regarding design, implementation, maintenance, and support.

CENIC AIR is a research endeavor, not a product. While the intention is to develop a sustainable, long-term production infrastructure, equipment can be recalled and repurposed for use by owners outside Nautilus at will.

Any questions can be directed to **info@cenic.ai**

---

---

# An Introduction to CENIC AIR

Artificial Intelligence, Machine Learning, and Data Science (AI/ML/DS) are opening new career paths for students at all levels who will become the next facilitators of academic research and private sector innovation and have the potential to transform entire industries, statewide labor forces – even society as a whole – in positive and beneficial ways.

To facilitate this transformation, there is a need for a national-level research and education AI resource that is open-access, scalable, and grown through the contributions of its user community.

Such a resource, including the infrastructure, and the communities that use it, already exists in the United States as the **National Research Platform (NRP)**. The NRP is an AI Research and Education infrastructure built, owned, and supported by the community for the community, with support from the National Science Foundation.

Given the importance of CENIC in connecting the research & education community in California, we refer to the California portion of the NRP as the **CENIC AI Resource (CENIC AIR)**. CENIC AIR provides California's research *and education* communities a platform to enable their faculty and students to contribute constructively to this transformation and collaborate with colleagues nationwide.


*FIGURE 1: A map of CENICs network and the CENIC AIR sites as of mid-2024. New sites at CSU Fullerton, Cal Poly Humboldt, CSU Monterey Bay, and others will be shown once operational.*

Part of the NRP, called the Prototype National Research Platform (PNRP), is providing educational resources to the National Artificial Intelligence Research Resource (NAIRR) Pilot through NAIRR Classroom (Press Release). While the NAIRR pilot is focused on providing a national allocation mechanism for AI research and education resources, NRP and CENIC AIR are focused on increasing ownership of AI resources for education and research across educational institutions, including California's many PhD-granting universities, CSUs, community colleges, and eventually K12 school districts and public libraries.

CENIC AIR is the CENIC-connected part of the National Research Platform providing California's research and education communities a way to enable their faculty and students to develop and refine topical AI curricula, constructively contribute to the transformations promised by these new technologies, and collaborate extensively with colleagues nationwide over the NRP infrastructure. Supplementing the NRP's leadership and staff, CENIC AIR adds CENIC's network planning, engineering, operations, and outreach activities to these efforts, all of which are key to expanding AI resources beyond the partner research-focused campuses.

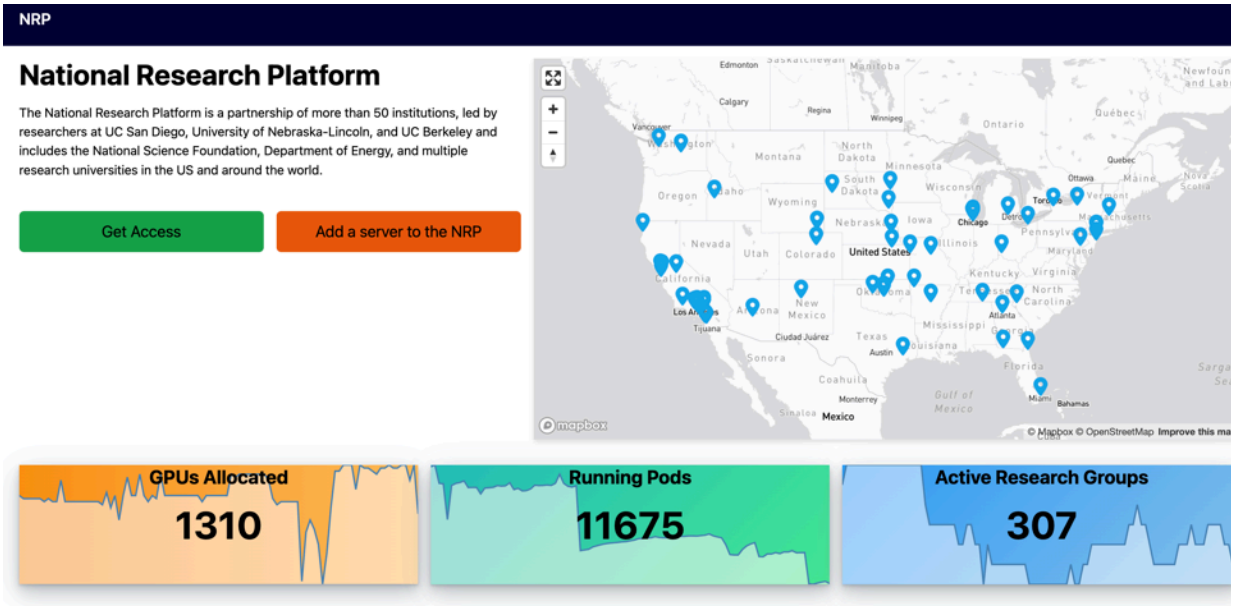


*FIGURE 2: NRP's Dashboard*

# CENIC AIR for Research and Instruction

## *CENIC AIR and Its Purpose*

### *1.1 What is CENIC AIR?*

CENIC AIR (cenic.org/initiatives/cenic-air ) is the CENIC-connected subset of the National Research Platform (NRP), a partnership of more than 50 institutions nationwide and growing. It is led by researchers and cyberinfrastructure professionals at UC San Diego and partner campuses, and it is supported in part by grants from the National Science Foundation. At the core of NRP is *Nautilus*: a large, distributed compute and storage cluster optimized to run data science/AI applications. Institutions voluntarily contribute to Nautilus' hardware and software, becoming co-owners. These resources are interconnected and accessed via regional and national research and education networks such as CENIC's CalREN and Internet2. System administration and remote operations are handled by IT professionals funded in major part by awards to the San Diego Supercomputer Center and the Qualcomm Institute at UC San Diego.

CENIC AIR aims to help CENIC-connected university researchers, faculty, and their students gain access to Nautilus, whether or not the campus/researcher/faculty provides computational or storage resources for Nautilus. For institutions that choose to contribute resources, CENIC Engineering can assist campuses with building out their networks to host such resources and help ensure access. Other regional networks part of the NRP are considering offering similar assistance.

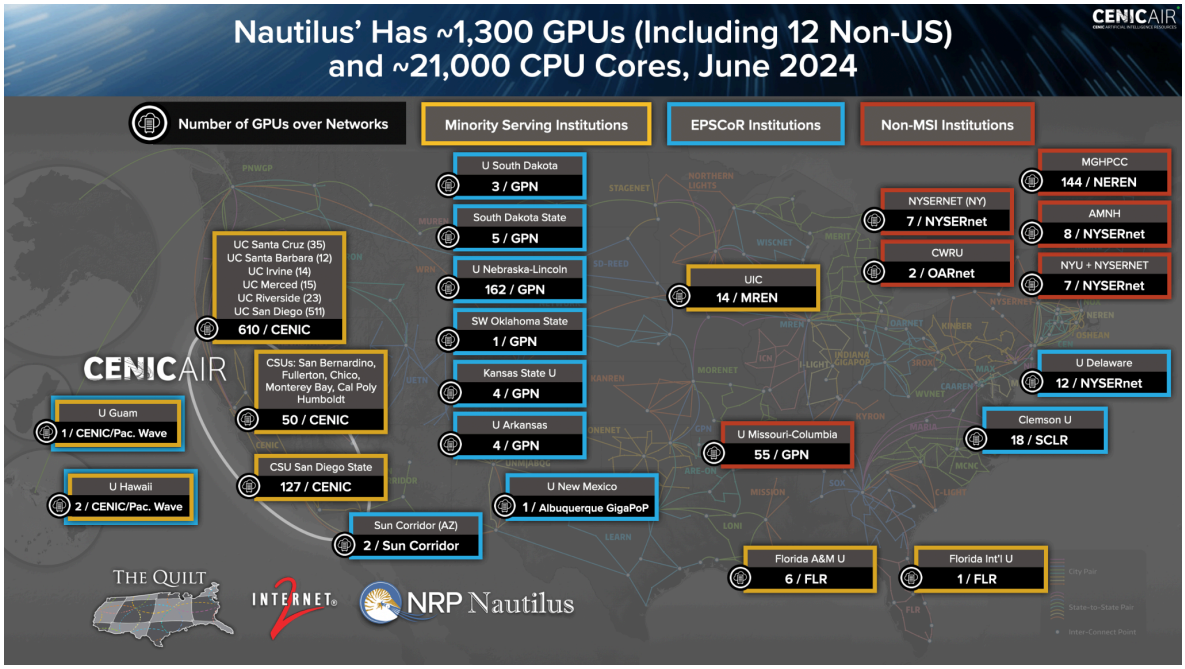


*FIGURE 3: Nautilus Compute Resources Nationwide*

Nautilus is currently hosted on >50 campuses and networking organizations' sites. It currently has online >21,000 CPU cores and >1,200 GPUs plus 13 petabytes of shared high-speed storage. There is access documentation on the NRP website, Nautilus users can also get live technical support via a set of online channels. There are channels dedicated to instructors, system administrators, and so on, in which users and the NRP administrators engage, learn from, and help each other.

CENIC AIR encourages researchers and faculty at non-profit institutions of higher education served by CENIC, and their students, to join this initiative and use their CILogon credentials to explore CENIC AIR resources (CILogon is an integrated identity and access management platform--see cilogon.org and NRP for more).

There is no cost to utilize the NRP's hundreds of existing data science and AI-focused computers for research and education use. When institutions add resources to NRP, they can provide their users with a guaranteed level of access. Users can still access NRP resources without their institutions contributing, but access is on an "as available" basis. The NRP team is building documentation for using Nautilus as well as for JupyterLab notebooks – which enable researchers, faculty, and students to have access to CENIC AIR infrastructure even if their institution doesn't presently house a CENIC AIR Science DMZ. (See https://jupyter.org for more general information on JupyterLab.) Nautilus has its own JupyterHub available to users that can allocate up to 8 GPUs.

***Like NRP, CENIC AIR is a program; the computers in Nautilus and the network infrastructure that connects them comprise the resource.***

## *1.2 Does it cost my institution anything for me to use Nautilus?*

No. Nautilus is provided at no cost to faculty, staff, and students at non-profit educational institutions. It is a long-term experiment in community cyberinfrastructure building and access supported in large part by several National Science Foundation grants. Faculty/staff gain access to Nautilus by first joining Nautilus at the NRP website (see join/contact to get on the mailing list and further down the page click on Matrix Chat under Quick Access). You use your CILogon credentials to access Nautilus. There are JupyterLab examples and other documentation (mostly oriented toward experienced data scientists at the moment). Students gain access through their instructors.

Many campuses are adding equipment to Nautilus (see funding information in Section 2.15). If you want to add resources to Nautilus, your institution must have implemented a Science DMZ, install the equipment in that Science DMZ, and allow Nautilus's operations staff access for monitoring, software installation, and maintenance. CENIC is developing several implementation and support models so CENIC engineers can help member institutions assess their existing CalREN connectivity for purposes of implementing and supporting a new, CENIC AIR Science DMZ; some of these models can be CENIC-managed through the network to

varying degrees, as contracted. There will be charges associated with these fuller-service CENIC-managed options.

## *Accessing Compute and Storage Resources*

### *1.3 How do I get access to Nautilus?*

The CENIC AIR website will provide a growing number of user-friendly tutorials as they are developed. For the adventurous now, go to the NRP website and pull down the Join/Contact tab. That will get you on the NRP mailing list. You can also start using Nautilus with your CILogon sign-on (see docs.nrp.ai). Experienced users with containerized applications ready to go should click on the Get Quick Access options on the NRP website. There are also JupyterLab examples and brief instructions (also at docs.nrp.ai).

### *1.4 Will CENIC AIR participants and users have access to other NRP resources, and if so, how will that work and be technically supported and maintained?*

Yes. Compute jobs can "burst" to computers and storage beyond California and into the national-level infrastructure and community resources of the NRP. Nautilus uses the open-source Kubernetes software to orchestrate user-containerized applications based on users' requests. In general, users need not be concerned about where the computation physically takes place. (There are, however, options to customize locality and resource types available to advanced users.) In addition, some Nautilus users have successfully ported their software containers to run in commercial clouds when they obtain the necessary credits.

Thanks to National Science Foundation grants and existing campus support, Nautilus' compute capacity is comfortably provisioned except for the very popular, most expensive A100/H100 GPUs which are in great demand by some of the researchers. Nautilus can accommodate thousands of researchers, faculty, and students with modest data science aspirations. The resources in Nautilus are very dynamic by design, old resources disappear, new resources get added, owners of resources may temporarily restrict access, and federal agency-provided resources may be subject to allocation mechanisms. Nautilus capacity continues to grow, year after year as a result of being community-owned and driven.

If overall utilization ever becomes overly high, some researchers from institutions not contributing resources may be required to run in "background scavenger" mode or request allocations from federal agencies.

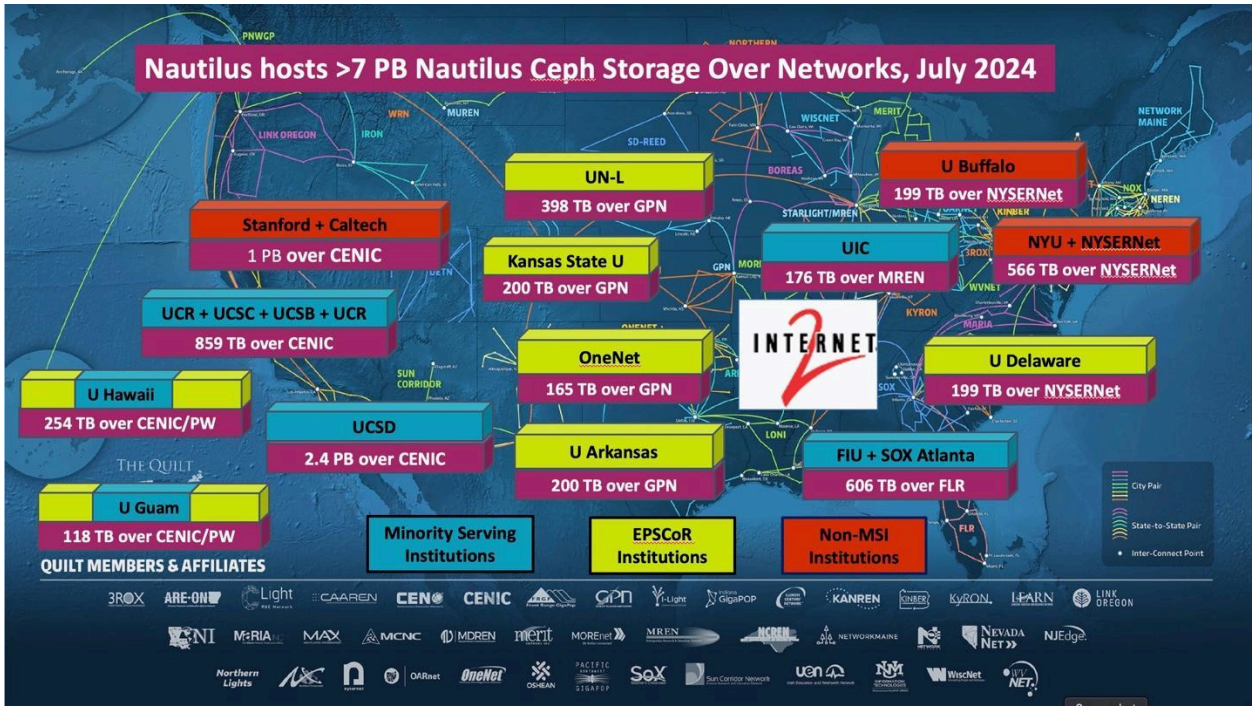


*FIGURE 4: Nautilus Storage Resources Nationwide*

## *1.5 Where is my data stored when I use CENIC AIR?*

Nautilus maintains its own storage intended to hold data while you are computing or sharing (also referred to as "hot storage"). Some storage is held in regional Ceph pools (see ceph.io).

There are various means to import your data into Nautilus, such as the Open Science Data Federation or Globus. Naturally, there are rules about what is deemed appropriate use of Nautilus' storage (download from the NRP website). For example, some of the storage offerings for the NRP are: S3 compatible storage, Ceph block devices, CVMFS, Linstor, and Seaweed FS. The discussion of storage technologies and their trade-offs are beyond the scope of this FAQ. Local scratch storage is also provided. Note: archival storage is not provided nor is storage backed up by Nautilus; it is assumed you can reload your data if needed, and that you will delete unused storage on Nautilus.

## *1.6 How does public cloud access relate to CENIC AIR?*

Researchers, faculty, and students at your institution can make use of commercial cloud services (e.g., AWS or Azure) via CloudBank, which provides information, education, training, and allocations for public clouds -- currently, Amazon Web Services, Google Cloud, IBM Cloud, and Microsoft Azure. Information, education, and training resources are available to anyone in the research community. Cloud credits obtained via programs like CloudBank or the NAIRR Pilot may be used to integrate cloud resources into Nautilus dynamically on demand.

Software developed using Nautilus can be transferred to run in the commercial cloud, in general. Detailed documentation on how to integrate cloud resources into Nautilus or develop software for the commercial cloud on Nautilus will be developed in due time; use Matrix to ask NRP colleagues in the meantime.

## *Accessing Support Resources for Research and Instruction*

### *1.7 How do I get support?*

NRP is using the innovative federated communication system (https://matrix.org) for all communications regarding the project. Matrix is deployed in Nautilus, and you can create an account in it using any compatible client. NRP provides support through Matrix with various channels (topics), one of which is dedicated to JupyterLab instructors and another to JupyterLab admins. There are also channels for support to sysadmins contributing resources and grant writers looking to use or contribute to the NRP (see https://nationalresearchplatform.org/updates/matrix-chat-for-nautilus-users/).

“Easy buttons” for accessing Matrix and JupyterLab “Hello World!” examples are under construction; the documentation at NRP right now is mainly targeted to campus IT support professionals and experienced computer scientists.

You will be informed how to get onto Matrix once you contact the NRP via join/contact at NRP. NRP leadership approves each addition to the mailing list. Instructors, after joining, create a namespace (Nautilus nomenclature for a project name); once they request and are granted administrator privileges, students’ namespaces can then be added and managed by the instructor.

---

# Connecting Your Institution to CENIC AIR

## *Connection Bandwidth and Science DMZ Models*

### *2.1 Can my institution connect compute/storage nodes to CENIC AIR/Nautilus with a DC or HPR connection?*

Nautilus requires a CENIC/CalREN-HPR 10G network connection for compute nodes, 40/100G for optional shared storage nodes, and a public Internet 1G connection with security enabled for maintenance. If your institution is listed at https://cenic.org/network/operations/maps/sites-by-tier as HPR-connected, there are several engineering options to create a Science DMZ (your campus may already have one; see section 2.5).

If your institution is not on HPR, please contact CENIC. You may also wish to place equipment at an institution with an HPR-connected Science DMZ and available rack space. CENIC AIR maintains a list of such places, although it does not negotiate the costs of hosting—that's a discussion to have between the two institutions.

### *2.2 How can I assess whether my institution presently has enough bandwidth from CENIC to support a Science DMZ as part of this program?*

There is a list of campuses with HPR connections at https://cenic.org/network/operations/maps/sites-by-tier. CENIC can help your institution with bandwidth provisioning, but if this is not currently possible, you may consider placing compute and storage nodes at a friendly campus with available rack space, power, and CalREN-HPR networking. CENIC AIR maintains a list of such campuses.

### *2.3 What is the minimum recommended physical space and power requirements to house a Science DMZ compute/storage nodes?*

A Science DMZ is often implemented as a rack (or part of a rack) in the properly air-conditioned computer room that also holds the campus border router—though it can be set up anywhere with fiber access and sufficient power/AC. Network traffic to the computers in the Science DMZ is split and diverted so it does not go through the main campus network. Usually, a top-of-rack switch is provided by the campus for the nodes in the Science DMZ.

Storage nodes do not take a lot of power, but the GPU nodes do. Nodes with eight 300-Watt GPUs can consume about 3kW each. Some GPUs (H100s) are now up to 700W. Nodes are generally 2 rack units or 4 rack units high. A few nodes can usually slip in easily; however, a whole rack of high-end GPUs needs very specialized power and cooling.

# *Adding Resources to CENIC AIR*

## *2.4 Where do Nautilus compute and data storage resources reside?*

CENIC AIR resources mostly reside on the campuses of California non-profit higher education institutions that physically host computers with specific network connections and meet Nautilus requirements for hardware and software. CENIC has installed storage servers at their Los Angeles and Sunnyvale backbone node sites to provide additional high-speed data sharing resources.

## *2.5 How does my institution add computers and storage to Nautilus?*

As shown in the maps in the **Introduction**, many institutions have added resources to Nautilus, from a single compute node to several racks of computers and storage. Reasons to do this are:

- Researchers and their institutions can plan for average needs, share any unused compute resources when unused, and similarly get more resources when needed.
- Successfully operating a system like Nautilus in support of an academic community requires a wide range of skills from hardware & system administration, to application and user support and user training. The Nautilus technical staff remotely manages the combined resource; there are community user support channels for all common topics of relevance that are monitored by professional staff to correct misconceptions, and answer questions that remain unanswered. There are also training events held, including an annual event for the entire NRP community.
- Dashboards and spreadsheets on usage are provided to help with capacity planning.
- Knowledge sharing about current and future features (software, hardware, and applications) is plentiful in the highly engaged Nautilus community.

To add computers and storage to Nautilus, you need to create a Science DMZ on your campus, which requires networking expertise and CENIC is ready to help. CENIC engineers offer multiple paths to help CENIC-connected campuses construct and connect a Science DMZ via CalREN-HPR. Although a Science DMZ is physically located on a campus, for speed and security reasons, it operates as if it is outside the campus from a networking point of view.

The term Science DMZ was first used by network engineers at the Energy Science Network (ESnet) and is defined by them as follows:

> *The Science DMZ is a portion of the network, built at or near the campus or laboratory's local network perimeter that is designed such that the equipment, configuration, and security policies are optimized for high-performance scientific applications rather than for general-purpose business systems or "enterprise" computing.*

*Developed by ESnet engineers, the Science DMZ model addresses common network performance problems encountered at research institutions by creating an environment that is tailored to the needs of high-performance science applications, including high-volume bulk data transfer, remote experiment control, and data visualization.*

A campus Science DMZ is isolated from the campus' own internal networks so traffic between computers in the Science DMZ and a campus' main network still goes through the campus firewalls or other security mechanisms. The resulting isolation effectively allows jobs to be remotely scheduled and permits Nautilus operations staff to maintain the software via highly automated scripts, a critical enabling technology for the NRP.

The NRP team can help you choose/specify compatible equipment, operating systems, and software to install resources in your Science DMZ. Resources can be as simple as a single compute node or several racks of them, plus storage.

## *2.6 How much computing and storage should I put in my campus Science DMZ?*

The general Nautilus idea is to encourage you to design your data science compute capacity for your average needs, sharing your excess capacity when you can, and using others' excess capacity when you need it on occasion. You can also participate in developing regional computing consortia connected (or not) to Nautilus. Data Science is a rapidly growing field, involving not only computer scientists but increasing numbers of domain researchers with digital data.

## *2.7 Will my institution's CENIC AIR equipment and support be interim assets as part of a project, or are they part of a long-term program in support of AI R&D?*

CENIC AIR and NRP are designed and intended to be persistent, long-term projects, but they are subject to continued funding. The bulk of the current system administration and user support funding for NRP comes from the NSF via the PNRP project to UC San Diego. The terms of the award are such that the San Diego Supercomputer Center expects this funding to continue until 2033. Any institutional or individual participation in CENIC AIR, including sharing compute and storage resources with other participants, is voluntary, and participants may choose to opt out of CENIC AIR at any time (although CENIC's Managed Science DMZ solution is subject to contractual terms). Some grant-purchased equipment owned by one campus might be placed on loan into a Science DMZ on a different campus for hosting, in which case either campus should be able to request the return of the equipment at any time. Any of these ownership/hosting arrangements should be bilateral between the institutions.

Nautilus' 95% or better uptime is generally appropriate for data science researchers; faculty with students who have instructional computer lab projects due on strict deadlines may need to consider other options, such as local resources, commercial clouds, or simply be a bit more flexible. Involvement with CENIC AIR/Nautilus computing is entirely optional and aspirational rather than contractual at this point, it should be noted.

### *2.8 What happens if compute resources become unavailable or are terminated?*

Generally speaking, Kubernetes manages Nautilus so compute jobs automatically shift to available resources. Unless there is some key data-generating instrument on a particular campus that is disconnected, you probably won't notice.

## *Technical Requirements for Participant Nodes*

### *2.9 We already have some GPU servers running Kubernetes. Must we still conform to NRP's Nautilus equipment and software requirements?*

If you want to make them accessible as part of Nautilus, you need to conform so that NRP's highly automated system administration can work efficiently.

## *Roles and Responsibilities*

### *2.10 If we provide hardware, will CENIC and NRP provide us with configuration, installation, and ongoing technical support?*

CENIC AIR/NRP, as a project, primarily operates Nautilus from a software point of view. Hardware choice, installation, and maintenance, as well as networking (switches, fiber, etc.), are your campus' responsibility. NRP leaders and staff are happy to advise you on the latest hardware configurations and make recommendations. One advantage to Nautilus is that you can test drive various hardware options before you commit to purchasing your own configurations.

### *2.11 Who is responsible for administering and maintaining devices that are part of Nautilus in Science DMZs?*

Nautilus technical staff members perform system administration for software and operating system maintenance purposes. Staff from the institution where a CENIC AIR Science DMZ is located provide hands-on support for things like hardware replacement and basic troubleshooting of computers and switches.

### *2.12 Who is responsible for configuring and maintaining my campus gateway router or router port that adjacently connects a CENIC AIR Science DMZ net?*

The campus is responsible for all networking configurations. CENIC Engineering can assist with network support at various levels within the CENIC AIR program, some contractual – such as the Managed Science DMZ solution – CENIC has created for this purpose.

### *2.13 Will my campus IT group have to support any CENIC AIR or Science DMZ devices that we physically host here? If not, how will outside support be provided for these devices?*

Besides providing power, cooling, networking, and occasional hands-on interventions, the local IT staff should be involved with the Science DMZ (once it is set up) only to the extent they are interested, especially with CENIC's Managed Science DMZ solution in force. Switches and routers should be serviced according to a campus's standard enterprise procedures. When a compute or data server goes down, Kubernetes takes it out of the available resource pool, and notifies Nautilus staffers, who then discuss repair/replacement with the equipment owner. We advise campuses to acquire hardware with 5-year warranty and plan for the equipment to be replaced soon after the warranty expires to best achieve sustainable operations. Kubernetes was designed to smoothly isolate dysfunctional compute elements for repair or end-of-life disposal.

### *2.14 What are the respective roles of CENIC Engineering and my institution's tech support staff for CENIC AIR or my institution's Science DMZ?*

Various levels of networking technical support are proposed as part of the CENIC AIR program. CENIC Engineering will define responsibility demarcation points in conjunction with the campus IT staff based on the level of CENIC network engineering support that the hosting institution has requested.

## *Funding Opportunities for Participants*

### *2.15 Is grant funding available to support my institution's involvement in CENIC AIR?*

Besides campus funds (often offered to faculty as part of startup packages, or as part of programs for increasing departmental or institutional computing capacity), there are many recurring Federal grant opportunities. For instance, the National Science Foundation's Campus Cyberinfrastructure Program (called CC* and pronounced "CC star") provides funding for campus-level cyberinfrastructure improvements, innovation, integration, and engineering for science applications and distributed research projects. Certain CC* awards include an obligation to share 20% of your funded facility with a wider community beyond your campus, for which CENIC AIR is an acceptable means to do so.

NSF's CISE/OAC has a variety of other tracks for equipment funding. NSF MRI and NIH S10 are other programs that support campus hardware infrastructure.

NSF's CISE/CNS CIRC Program is another one: see: https://new.nsf.gov/funding/opportunities/community-infrastructure-research-computer/nsf23-589/solicitation

The Department of Defense gives DURIP equipment awards. See, for example: https://www.nre.navy.mil/education-outreach/sponsored-research/university-research-initiatives/durip

CENIC AIR can help you choose suitable equipment, and you are encouraged to use Nautilus to test-drive various GPU configurations, for example, to optimize your procurement strategies.