Purdue University Campus Cyberinfrastructure Plan

The NSF "CI2030: Future Advanced Cyberinfrastructure" report[1] describes *"An integrated cyberinfrastructure that reaches from university and college campuses to the national centers is needed; this will require coordinated investments by all of the stakeholders….*

*… the vast majority of capacity-class computing activities will be carried out **on campuses**.*

*... communication and collaboration across research silos offers the possibility of building and deploying an **integrated cyberinfrastructure** that effectively and efficiently supports a broad range of scientific and engineering research.*

Research[2] reveals that *"modest, but **consistent investment** in HPC results in measurable increases in research competitiveness, and a corresponding increase in research funding and publication counts."* and "*investment in high performance computing as measured by entries on the Top 500 list is a good predictor of research competitiveness*"

In this context, Purdue University operates a campus cyberinfrastructure under the following vision and principles:

Vision

- To be the one-stop provider of choice for research computing and data services at Purdue
  - Delivering cost-effective, powerful, reliable, easy-to-use computing and data services to Purdue researchers.
  - *"High Performance Computing at the Highest Proven Value"*

General Principles

- Governed jointly with the faculty – "*A coalition of the willing*":
- **Consistent investment** annually in current-generation cyberinfrastructure resources
- Delivering a broadly-usable cyberinfrastructure to a wide cross-section of scientific disciplines.
- With self-service interfaces to empower faculty to control their resources

Institutional Cyberinfrastructure Background

Since the 1960s, Purdue University has operated central computing infrastructure in support of research. In 2001, this research computing organization was combined with administrative IT to form ITaP (Information Technology at Purdue). Reporting to the Purdue System CIO, Purdue Research Computing has operated large-scale "community clusters" since 2004.

---

[1] https://www.nsf.gov/cise/oac/ci2030/ACCI_CI2030Report_Approved_Pub.pdf

[2] http://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1001&context=computing_pubs

Today, Research Computing operates community clusters, research storage, science gateways, a visualization facility, high-speed networks, and provides cyberinfrastructure expertise for Purdue researchers.

Governance

Research computing at Purdue is governed by Research Computing leadership and the faculty partners. Twice annually, governance meetings are held – one meeting open to the entirety of the cluster community, and one smaller meeting of the governance committee consisting of staff from Research Computing, and the Executive Vice President for Research and Partnerships, academic IT, and top faculty partners. These meetings are used to present the state of the program, discuss architecture decisions and technology deployments, and to gather requirements for upcoming research computing needs by the campus.

Areas of Focus (2018-2020)

- Computing to support Data Science, AI, and Machine Learning
    - Composable computing infrastructures
    - Create an upgrade path for growing from laptop to HPC
- Provide robust instructional resources for HPC and Data Science education
- Provide computing facilities and expertise to support Life Sciences, Plant Science, and Drug Discovery
- Secure computing to support regulated research (export control and HIPAA), national defense, and aerospace

Computation

The core of Purdue's cyberinfrastructure strategy, Purdue has operated since 2004 a world-class community cluster program - each year deploying an HPC system on the order of 550 nodes, with approximately 70 faculty groups investing in each system. In total, over 200 faculty groups from 60 academic departments and every academic college invest funds in the program, with over 1200 active users. Research Computing and the faculty **share the capital costs** of the program, with faculty purchasing compute nodes with grant or institutional funds; and Research Computing centrally funding shared network and storage infrastructure.

Professional administrative and computational science staff are centrally funded.

A testament to the power of this shared model, when the 2013 Conte cluster was deployed, it ranked at #28 in world on the November 2013 Top 500 list, the highest ranked system deployed in the community cluster program.

Purdue supports export controlled research and controlled unclassified information (CUI) on "Weber", deployed in 2019. Efforts to support CUI and regulated research are supported by NSF CICI award # 1840043.

Data

Since the 1990s, Purdue has provided the Fortress archive system to all researchers at Purdue, at no cost to the individual researcher. This large-scale archive has grown to hold over 17PB of research data as of Spring 2020.

In fall 2014, the Research Data Depot entered production, providing a highly-redundant, highly-reliable 2.5PB storage service available for purchase by research labs. Building upon an institutional investment, faculty can invest funds to purchase TBs of storage per year, and utilize high-capability storage for actively-used, shared datasets, applications, or other uses.

The Research Data Depot is well-suited as a storage target for instruments, or for collaborative data sharing using the campus Globus endpoint. Over 600 labs have purchased over 2 PB of storage in the Data Depot, 60% of which are not yet using HPC. Providing an on-ramp to computing for data-only researchers remains a strategic goal.

Finally, ITaP and the Purdue University Libraries jointly develop and maintain the Purdue University Research Repository (PURR) for creating data management plans, share data with collaborators, and publish and describe finished datasets. ITaP and Libraries personnel are engaged with Purdue researchers to train on best practices for managing and working with their research data.

Science Gateways

Science gateways provide access to advanced resources for science and engineering researchers, educators, and students. Through streamlined, online, user-friendly interfaces, gateways combine a variety of cyberinfrastructure components in support of a community-specific set of tools, applications, and data collections. [3]

The campus cyberinfrastructure is complemented by an active research programs in science gateways. Research Computing scientists provide science gateways supported by NSF DIBBS (*"DEEDS"*, award # 1724728), and CSSI (*"GeoEDF",* award # 1835822).

The NanoHUB (award # 1227110) provides a library of 3,000 learning resources to 195,000 users worldwide, each year. Its 232 simulation tools, free from the limitations of running software locally, are used in the cloud by over 10,800 users annually.

Visualization

The Envision Center for Data Perceptualization provides expertise and professional services to assist, support, and collaborate with faculty, students, and industry in scientific visualization, virtual and augmented reality and media creation.

The Envision center facility provides facilities for motion capture, augmented and virtual reality, and a large visualization wall for interacting with large datasets.

Expertise

[3] Wilkins‐Diehr, Nancy. "Special issue: science gateways–common community interfaces to grid resources." Concurrency and Computation: Practice and Experience 19, no. 6 (2007): 743-749.

Research computing staff provide domain expertise, technology solutions, software development to Purdue researchers. Staff offer open office hours around campus, training sessions in data science and cluster usage, and instruct in academic courses. Expert staff are available to partner with faculty on projects, and provide cyberinfrastructure/computational expertise.

Networking

All Purdue centrally managed compute and storage systems are connected to a central research network core in the Purdue Math Building Data Center at 400 Gbps. This central research core connects to Purdue's campus core at 400 Gbps, which then connect to the internet edge wide-area network routers at 400 Gbps. At the campus peering point for research networks, Purdue operates an intrusion detection system funded by an NSF CICI award (# 1738981). Purdue maintains a 200 Gbps connection to the Indiana GigaPOP in Indianapolis, Indiana, with an additional 200 Gbps path (shared with Indiana University) to Chicago, Illinois. From there, systems can access a variety of research networks, including the NSF-supported XSEDE network, at 100 Gbps.

As a long-time LHC site (a CMS Tier-2), Purdue has infrastructure in place to monitor network performance with Perfsonar, and supports IPv6 in a dual-stack mode, to selected computing resources and data transfer nodes.

In 2018, Purdue was awarded an NSF CC* networking grant (award # 1827184) to add high-speed Science DMZ connections to selected big data facilities enabling high-volume, high-velocity, or interactive science data flows to both the campus research cyberinfrastructure and off campus facilities.

Physical Infrastructure

Most research computing assets reside in Purdue University's Mathematical Sciences Building on the West Lafayette campus. The building, built in 1966 and expanded in 1982, has received multiple renovations to its 6,900 square feet of datacenter space. Most recently, an NSF ARI award provided funding for significant upgrades to power and cooling, providing 2.5 MW of power and 525 tons of cooling capacity.

Some resources, including the secondary site for the Fortress archive and the Research Data Depot are located in a datacenter in Felix Haas Hall. Built in 1908 and remodeled in 1984, Haas hall provides 1985 square feet of space, 400 KW of power, and 60 tons of cooling capacity.

Three research systems are housed in HP POD modular datacenters. The POD facilities provide an additional 42 racks of capacity, 850 KW of power, and 350 tons of cooling capacity.

National Cyberinfrastructure Community

Purdue University is a partner in the NSF-funded national advanced computational cyberinfrastructure XSEDE since 2011. Purdue is also a Level 2 Service Provider (SP) to XSEDE and a member of the XSEDE Service Provider Forum since its inception. Purdue's research computing systems are connected to the XSEDEnet via 200 Gbps network links. Purdue research computing staff is part of the XSEDE ECSS organization and provide in-depth consulting services to help researchers nationwide to effectively utilize the XSEDE resource. Their expertise areas include application and workflow optimization, scientific application development, and science gateway development and operations. Purdue staff also play a key role in the XSEDE

Campus Champion program by leading and coordinating activities for the campus champion community to get help from experts and also help connect researcher needs with XSEDE resources and other national cyberinfrastructures.

Purdue is a resource provider to the Open Science Grid, providing community cluster computing and storage resources to the OSG to support the CMS (Compact Muon Solenoid) project as a Tier-2 site as well as other OSG virtual organizations.

Purdue is a member of InCommon, using the federated identity infrastructure to easily allow Purdue researchers to access national resources like XSEDE, OSG, or Globus.