Purdue University
Purdue e-Pubs

University General Facility Boilerplate Descriptions | Office of Research and Partnerships

5-6-2019

Rosen Center for Advanced Computing: Facilities, Equipment, and Resources

X Carol Song
Purdue University, carolxsong@purdue.edu

Purdue University Office of Research and Partnerships

Follow this and additional works at: https://docs.lib.purdue.edu/gendes

Recommended Citation
Song, X Carol and Purdue University Office of Research and Partnerships, "Rosen Center for Advanced Computing: Facilities, Equipment, and Resources" (2019). *University General Facility Boilerplate Descriptions*. Paper 4.
https://docs.lib.purdue.edu/gendes/4

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Research Computing and Information Technology at Purdue
(Last updated: March 22, 2019)

Please contact Dr. X. Carol Song for more information regarding support for your proposal:
Dr. Carol Song, Director of Scientific Solutions, RCAC, 496-7467, carolxsong "at" purdue.edu

The RCAC (Rosen Center for Advanced Computing, http://www.rcac.purdue.edu) is the research computing arm of the Information Technology at Purdue (ITaP), the University's central IT organization. RCAC provides a large and diverse set of high-performance computing, data-intensive computing and cloud computing resources, high-speed network connections to national research wide-area networks, and large data storage and archival systems to the broad research communities at Purdue and elsewhere. RCAC also provides expertise, services and software development effort to the national research community through partnerships with national cyberinfrastructure projects such as the XSEDE (formerly TeraGrid), Open Science Grid (OSG), and the Science Gateways Community Institute (SGCI).

RCAC offers a broad range of technical support services to Purdue researchers, including user support, consulting, training, and documentation; software installation; computation optimization and scaling; large-scale data management; capacity planning and cluster deployment. RCAC researchers and software engineers also collaborate with faculty and students in creating solutions and tools through software development, integration and hosting services. The RCAC staff are provided with office space in Young Hall, Purdue University.

The paragraphs below provide additional details of the Purdue centrally supported research computing resources.

Computing and Data Resources

Purdue's RCAC operates several large, state-of-the-art computing clusters as well as data storage and other resources to serve the research and education community at Purdue campuses and nationwide. These systems support scientific computation and data analysis in many research domains ranging from natural sciences, engineering and social sciences. The RCAC operated systems include:

High-performance computing

- *Gilbreth* is Purdue's newest Community Cluster, built through a partnership with Dell and Nvidia in December 2018. Gilbreth nodes consists of various compute nodes with 128 to 192 GB of system memory. Each node has two GPUs per node, with a total of 56 P100 GPUs and 44 V100 GPUs in the system. Finally, for training large models, 3 nodes are available with 768GB of RAM, 8 TB of local flash storage, and 4 32 GB V100 GPU accelerators with Nvlink. Gilbreth nodes share access to a 2.5 PB Lustre filesystem and a 150TB all-flash shared storage system.
- *Brown* is designed and optimized for communities running traditional, tightly-coupled science and engineering applications. Brown was built through a partnership with Dell and Intel in October 2017. Brown consists of Dell compute nodes with two 12-core Intel Xeon Gold "Sky Lake" processors (24 cores per node) and 96 GB of memory. Brown currently has 550 nodes, totaling 13,200 cores on the cluster. All nodes have 100 Gbps EDR Infiniband interconnect and a 5-year warranty.
- *Halstead* is a community cluster built through a partnership with HP and Intel in November 2016. It consists of 508 HP compute nodes, each with two 10-core Intel Xeon E5 processor (20 cores total per node) and 128 GB memory. All nodes are connected through the EDR Infiniband interconnect and to

a Lustre parallel file system built on Data Direct Networks' SFA12KX EXAScaler storage platform. This cluster will be in operation from 2016 through December 2021.

- *Rice* is a community cluster built through a partnership with HP and Intel in April 2015. Rice consists of HP compute nodes, each with two 10-core Intel Xeon-E5 processors (20 cores per node) and 64 GB of memory. Rice has a total of 13,200 processor cores, and in testing, clocked in with a sustained, measured maximum speed of 407.1 teraflops and a peak performance of 549.1 teraflops. All nodes have Mellanox 56 Gb FDR Infiniband interconnect and a Lustre parallel file system built on Data Direct Networks' SFA12KX EXAScaler storage platform.
- *Snyder* is a big-memory system optimized for data intensive applications requiring large amounts of shared memory per node, such as those used in life sciences research. Snyder consists of more than 70 HP compute nodes, each with two 10-core Intel Xeon-E5 processors (20 cores per node) and various memory configurations of 256 GB, 512 GB and 1 TB. All nodes have 40 Gbps Ethernet connections.
- *Hammer* is a high-throughput computing system that is optimized for scientific applications running loosely-coupled computation tasks. Hammer was initially built through a partnership with HP and Intel in April 2015, consisting of HP DL60 compute nodes with two 10-core Intel Xeon-E5 processors (20 cores per node) and 64 GB of memory. All nodes have 10 Gbps Ethernet connections. Hammer was expanded in late 2016 with 40 new nodes, each with two hyper-threaded 10-core Intel Xeon E4-2660 v3 processors (40 cores per node) and 128 GB of memory. All nodes have 25 Gbps Ethernet connections. The number of nodes in Hammer will be expanded annually.
- *Scholar* is a cluster designed for use instructors and students. This system is open to Purdue instructors from any field whose classes include assignments that could make use of supercomputing or data science. Account management in Scholar is integrated with the university's registration system. The system hardware consists of HP compute nodes with two 8-core Intel Xeon-E5 processors (16 cores per node), 32 GB of memory, and a 500 GB system disk. All nodes have 56 Gbps FDR Infiniband connections. Scholar also provides 7 large interactive frontends to teach data science, GPU frontends, and a queue providing access to Nvidia V100 GPUs.
- *Diagrid*.org is a web-based system that delivers computation and data resources to scientists via a web browser. Built on the HUBzero™ Scientific Collaboration Platform software package, DiaGrid.org provides "Science-as-a-Service", online applications delivered to user's web browser with graphical user interfaces while accessing large-scale computing and data resources behind the scene. It is also an open platform for researchers and students to share their own scientific applications following a built-in, do-it-yourself process. Current applications on DiaGrid include tools for bioinformatics, molecular dynamics, structural biology, climate modeling, and statistical analysis and parameter sweeps using R language. DiaGrid.org also provides many social networking capabilities built for research and learning, including projects, groups, discussion forums, review, tagging, citation, and so on.
- *REED* (Research Environment for Encumbered Data) is designed for use with security protocol compliance data, beyond those generally used within an open academic setting. REED is implemented within Amazon's GovCloud service with access portals and control points positioned within Purdue's Research Computing Infrastructure. Current REED resources use Microsoft Windows Server, accessed through Remote Desktop protocols, and backed with an expandable pool of AWS compute nodes.
- *Data Workbench* is an interactive computing environment for non-batch big data analysis and simulation. The Data Workbench consists of HP compute nodes with two 8-core Intel Xeon processors (16 cores per node), and 256 GB of memory. All nodes are interconnected with 10 Gigabit Ethernet. Available tools include Jupyter, R Studio, and Remote Desktop. This resource is part of the Purdue research computing cyberinfrastructure, with high-speed network access to the storage systems.

Network connectivity

All Purdue clusters and storage systems are connected with redundant 160 Gigabits per second (Gbps) Ethernet links to the research network core. The campus systems' wide-area connection to the Indiana GigaPOP and major research networks including XSEDEnet, OmniPOP, and ESNET is 100 Gbps.

Public-facing research computing servers and data transfer nodes are accessible via IPv6. All research computing systems support outbound IPv6 in a dual stack mode.

The dedicated campus research network is built in the Science DMZ model featuring a high-speed core, Globus data transfer nodes, a set of PerfSONAR monitoring nodes at key points around the network, and distinct security policies from the campus at large. Currently, PerfSONAR is in production at the core of the research network, but 10 PerfSONAR nodes are being deployed at key points of the network infrastructure, including the research core, the primary campus routers, and the wide area network entry points.

Data Storage

Purdue operates and maintains a multi-tier storage system to meet the needs of researchers conducting computation, data analysis, data management and sharing, and data archiving.

- *Research Data Depot*: The Data Depot is a high-capacity, fast, reliable and secure data storage service designed, configured and operated for the needs of Purdue researchers in any field and shareable with both on-campus and off-campus collaborators. The Data Depot uses an enterprise-class GPFS storage solution with an initial total capacity of over 2 PB and will continue to expand to meet growing demands. This storage is non-purging, redundant and reliable, features regular snapshots, and is accessible from all Purdue research systems. Built on Data Direct Networks' SFA12k storage platform, the Data Depot has redundant storage arrays in multiple campus datacenters for maximum availability. Multiple data transfer methods are supported, including SCP/SFTP, CIFS/SMB (as Windows drive), and Globus data transfer for fast and easy unattended transmission of large amounts of data between local systems or to/from national resources and labs.
- *Scratch Storage*: Each computing cluster has a default Lustre parallel file system to provide work-area storage optimized for a wide variety of job types, and are designed to perform well with data-intensive computations, while scaling well to large numbers of simultaneous connections. Scratch storage currently consists of 1.4 to 3.4 PB of redundant, high-availability disk space. All scratch tier resources are high-performance, large capacity, and subject to scheduled purging of old files. User quota on the scratch storage is typically 100 TB and 1 million files.
- *Home storage*: Designed for use by individual researchers to hold files permanently. It is being served from the DDN GS7KX filesystem appliance and uses a quota system for space management. Snapshots of user data are regularly taken and accessible from all Purdue research systems.
- *Archive storage*: The Fortress Archive (HPSS) system is a large, long-term, multi-tiered file caching and storage system utilizing both online disk and robotic tape drives. This archival system consists of 400 TB of disk cache, 1.2 TB of SSD for file metadata storage, and two tape libraries currently totaling 11PB in storage capacity which can be expanded to provide additional archival storage according to user demands.
- *Globus data transfer*: RCAC supports the use of Globus for transferring large amounts of research. Globus offers the ability to reliably and securely transfer a large amount of data between systems. It

also allows researchers to easily share large datasets with collaborators through shared endpoints. Globus is available on all RCAC storage systems, including Research Data Depot and Fortress archive.

Visualization and Graphics

RCAC is home to the Purdue Envision Center, a facility that provides novel solutions to effectively communicate complex research concepts through computer graphics, advanced visualization, auditory (sound), haptic (touch), and multimodal interaction integrated with state-of-the-art high-performance computation. The Envision Center is staffed by project manager, technical lead, developers and students to assist researchers, instructors, and leaders in their quest for new knowledge and innovative products. Recent and ongoing research supported by center staff includes projects in the life sciences, audiology, engineering, communications, theatre, computer science, geology, and management.

National Cyberinfrastructure Resource Provider

Purdue University is a resource provider to the NSF-funded national high-end computational cyberinfrastructure XSEDE (http://xsede.org) and a member of the XD Service Provider Forum (https://www.xsede.org/web/sp-forum/spf-membership). Purdue's research computing systems are connected to the XSEDE resources via 100 Gbps network links. Purdue's Condor pool, a distributed computation resource for high-throughput computing, is available to XSEDE users via the Open Science Grid. Purdue's DiaGrid, a HUBzero web-enabled platform with online scientific applications backed up by high-performance and high-throughput computing resources, supports XSEDE researchers through hosting of their scientific applications and gateways. DiaGrid is also available to the broader research community, educators and students. As a partner to the XSEDE project, Purdue staff provides in-depth advanced consulting services to help researchers nationwide to effectively utilize the XSEDE resource. Their expertise areas include optimization, scientific application development, and science gateway development and operations. Purdue staff also participate in the XSEDE Campus Champion program, helping to connect campus research needs to advanced digital resources on XSEDE and other national cyberinfrastructures. Purdue is a resource provider to the Open Science Grid with its Condor pool for high-throughput computing applications. Purdue provides computing and storage resources to the OSG, supporting the CMS (Compact Muon Solenoid) project as a Tier-2 site and other OSG virtual organizations.