Research Computing Services and the Shared Computing Cluster (SCC)

Updated 9/1/2024

Brief Overview:

Research Computing Services (RCS), a department within Information Services & Technology (IS&T), provides advanced computing facilities, software, training, and consulting for computational research and academic courses at Boston University.

For 40 years, RCS has provided consulting, training, and infrastructure support to thousands of researchers and students on the Charles River and Medical Campuses. RCS supports a wide range of disciplines including Physical Sciences, Engineering, Biological and Medical Sciences, Public and Global Health Research, Business and Economic Research, Social Sciences, Mathematics, and Statistics, as well as Data Sciences and Artificial Intelligence. RCS serves 3,700 researchers in 1,200 projects from 90 departments and centers at the University. In addition, 70 courses from 20 academic departments use the SCC.

RCS manages the University's [Shared Computing Cluster (SCC)](#), a large heterogeneous Linux cluster with over 28,000 CPU cores, 400 GPUs, and 14 PB of disk for research data. The compute nodes are interconnected with a variety of InfiniBand fabrics for multi-node parallel jobs, 25 GbE for data intensive jobs,

 and 10 GbE for general purpose compute jobs running within a single node.

A variety of storage options are provided on the SCC and through other IS&T services. All SCC storage systems are protected through hardware RAID and snapshots. A portion of the project disk storage space is approved for storing data with more stringent compliance requirements, such as some forms of HIPPA-limited data sets or NIH data from the Database of Genotypes and Phenotypes (dbGaP). No Restricted Use or HIPPA data can be stored on any part of the SCC.

The SCC and other Research Computing resources are located at the LEED Platinum certified Massachusetts Green High-Performance Computing Center (MGHPCC) data center in Holyoke, MA. Connectivity between the MGHPCC and the BU campus is provided by two pair of 10 GigE (10 Gb/s) fiber loops, providing redundancy, as well as a total capacity of 40 Gb/s.

RCS supports a wide range of over 900 software application packages installed on the SCC. Researchers may request additional software to be installed.

Additional Details (as Needed)

Research Computing Services

Research Computing Services (RCS), a department within Information Services & Technology, provides specialized hardware, software, training, and consulting for all areas of computational research at BU. RCS supports a wide range of disciplines including Physical Sciences, Engineering, Biological and Medical Sciences, Public and Global Health Research, Business and Economic

Research, Social Sciences, Mathematics, and Statistics, as well as Data Sciences and Artificial Intelligence. RCS staff have diverse scientific backgrounds, many with advanced degrees in their area of specialty.

Resources are managed in close consultation with the Research Computing Governance Committee, the Shared Computing Cluster Faculty Advisory Committee, and the Rafik B. Hariri Institute for Computing and Computational Science & Engineering. The Research Computing Governance Committee is co-chaired by Tom Bifano, Interim Vice President for Research, and Anita DeStefano, Professor of Biostatistics and Neurology as well as Graduate Affair Faculty Fellow for Diversity and Inclusion.

Computational Resources

RCS manages the Shared Computing Cluster (SCC), a large heterogeneous Linux cluster available to all University faculty, their collaborators for research, and their computational academic courses. The cluster comprises approximately 28,000 CPU cores, 400 GPUs, and over 14 PB of disk for research data. The compute nodes are interconnected with a variety of InfiniBand fabrics for multi-node parallel jobs, 25 GbE for data intensive jobs, and 10 GbE for general purpose compute jobs running within a single node.

The SCC is composed of shared and buy-in resources. The shared resources, fully funded by the University, are available on a fair-share, no-cost basis to all faculty-led research groups. The buy-in resources are directly funded by researchers for their own priority use. Excess buy-in capacity is returned to the shared pool for general use. Buy-in Program offerings are available at rates specially negotiated with our vendors and detailed on the Buy-in Program web-page. All shared and buy-in resources are professionally managed by RCS at no charge to researchers.

Other service models include dedicated and co-location services. The dedicated service model provides fee-based management for other research computing systems that do not align with the shared or buy-in models. For those faculty wishing to have complete control managing their own resources, the co-location model provides rack space, power, cooling, and network access without RCS staff support.

RCS staff can facilitate access to regional and national advanced computing facilities for projects that require resources beyond the scope of the SCC. Resources and services include the NSF funded Advanced Cyberinfrastructure Coordination Ecosystem (ACCESS) program and the National Artificial Intelligence Research Resource (NAIRR) pilot program as well as regional services including the New England Research Cloud (NERC). Please contact RCS staff for more information on these programs and resources.

Data Storage

A variety of storage options are provided on the Shared Computing Cluster and through other IS&T services. All SCC accounts are provided with home directory space. Each research project may request up to 1 TB of high-performance project storage space, 20% of which may be backed up. Additional project space is available for purchase through the Buy-in program or rental through the Storage-as-a-Service program. For those researchers who need to maintain a second copy of data, an off-site storage system called STASH is available with the same cost structure as primary storage. All storage systems are protected through hardware RAID and snapshots. A portion of the project disk

storage space is compliant for storing Confidential data and NIH dbGaP (human genomics) data. No Restricted Use or HIPPA data can be stored on any part of the SCC.

IS&T's Network File Storage service provides a secure, centrally-managed, storage environment for University data. Data on this system can be easily copied to and from the SCC's high-performance storage system for computation. All research projects are entitled to 1 TB of storage at no cost with additional storage available at a subsidized annual rate.

Networking

The SCC and other Resource Computing resources are located at the MGHPCC data center. Connectivity between the MGHPCC and the BU campus consists of two pairs of 10 GigE (10 Gb/s) fiber loops, providing redundancy, as well as a total capacity of 40 Gb/s.

The BU campus network provides high-speed access to campus networks, all institutional information, communication, and computational facilities, along with the Internet, regionally aggregated resources, and advanced networks such as Internet2.

Software

The SCC has over 900 general and domain specific software application packages installed and available through the module system. Many compilers and libraries common in high-performance computing are also supported. Researchers may request additional software to be installed.

Consulting

Research Computing Services' broadly trained staff assist researchers on a range of topics. These include introductory use of the Shared Computing Cluster (SCC), the various programming tools and packages, data management, program parallelization, performance measurement, and code tuning, numerical methods, computational techniques. Short-term training and consulting is offered at no cost. Longer-term or dedicated staff time can be arranged for a fee.

Training

Research Computing provides training on a wide range of topics designed to help researchers make effective use of the Shared Computing Cluster. A semesterly tutorial series provides hands-on tutorials including Linux basics, a variety of programming languages, high-performance computing, data analysis, and data visualization. Additionally, RCS staff can offer extra sessions and/or customize tutorials for a particular course, seminar, lab or research group. The RCS website provides current semester training offerings and video training resources. All training resources are free and open to all members of the Boston University community.

Massachusetts Green High-Performance Computing Center (MGHPCC)

Boston University is a founding member of the Massachusetts Green High-Performance Computing Center (MGHPCC), a collaboration of universities, industry, and the Commonwealth of Massachusetts. The consortium built and operates a world-class, LEED Platinum certified, high-performance computing center to support the expansion of university research computing and collaboration. MGHPCC consortium members include Boston University, Harvard University,

Massachusetts Institute of Technology, Northeastern University, the University of Massachusetts, and Yale University; industry partners include Cisco and Dell EMC; and the Commonwealth of Massachusetts.

The MGHPCC was the first university research data center to achieve LEED Platinum Certification. There are only 13 LEED Platinum certified data centers in the United States. The design and engineering of the datacenter focus on maximizing energy efficiency while minimizing negative environmental impacts. The data center is supplied with economical clean renewable energy by Holyoke Gas & Electric's hydroelectric power plant located on the Connecticut River.

The MGHPCC, its continual evolution and regular consortium engagement, encourages collaboration between members and regional parties. The unique facility infrastructure has allowed for large collaborative projects that developed regional cyberinfrastructure resources and services. The continuing development of this center is creating unprecedented opportunities for collaboration between research, government, and business in the Northeast.

The MGHPCC is designed to support the growing scientific and engineering computing needs at six of the most research-intensive universities in Massachusetts – Boston University, Harvard University, Massachusetts Institute of Technology, Northeastern University, the University of Massachusetts, and Yale University. The computing infrastructure in the MGHPCC facility includes 33,000 square feet of computer room space optimized for high performance computing systems, a 19MW power feed, and a high efficiency cooling plant that can support up to 10MW of computing load. The on-site substation includes provisions for expansion to 30MW and the MGHPCC owns an 8.6 acre site, leaving substantial space for the addition of new floor space. The communication infrastructure includes a dark fiber loop that passes through Boston and New York City and connects to the NoX, the regional education and research network aggregation point. Boston University is connected to the MGHPCC through two pairs of 10 GigE connections, providing an aggregate capacity of 40 Gb/s, from its campus to its resources located in the Holyoke facility.

MOC Alliance/NERC/NESE

Hosted by the Rafik B. Hariri Institute for Computing and Computational Science & Engineering and housed at the MGHPCC, The **Mass Open Cloud (MOC) Alliance** is a partnership between higher education, government, and industry to create an open production cloud that will provide domain researchers with predictable low-cost services while enabling innovation by a broad community of academic researchers and industry collaborators. To achieve these goals, the MOC Alliance supports and coordinates interrelated projects, including the New England Research Cloud (NERC), a production cloud service supported institutionally by BU and Harvard University, the Northeast Storage Exchange (NESE), the Open Cloud Testbed (OCT) for cloud researchers, the Center for Systems Innovation at Scale (i-Scale), NSF award 2333320, currently comprised of a site at Boston University and a site at Northeastern University, and the $20M Red Hat Collaboratory at BU. A bare metal service enables computers to be securely shifted between MOC Alliance services as needed.

In 2024, the MOC Alliance has continued to grow, most recently coordinating a partnership with Lenovo to add 64 NVIDIA A100 GPUs to the NERC and becoming a founding member of the AI Alliance to host research and education use cases and to enable developers of tools and hardware accelerators to expose their innovation to users.

The **New England Research Cloud (NERC)** provides a secure on-premise private cloud solution that is integrated with the Massachusetts Green High Performance Computing Center's (MGHPCC) networking facilities and the Northeast Storage Exchange's (NESE) storage services. Using self-service portals or public interfaces, users can request cloud resources for testing, development, and production purposes. NERC provides the following types of services: Virtual Machine (VM) orchestration and object storage (Swift) services through RedHat's OpenStack platform, Container orchestration services through RedHat's OpenShift platform, and a data science platform through RedHat's OpenShift Data Service (RHODS). Currently, NERC consists of over 5,200 virtual CPU cores, 27TB of memory, and 150 GPUs. NERC currently supports 90 principal investigators and 880 users from 10 regional institutions. NERC has received two rounds of funding, totaling $2M from MassTech Collaborative Matching Grant Awards.

NERC storage services leverage Ceph, a distributed and high-performance storage solution supported by the **Northeast Storage Exchange (NESE)**, to offer block and object storage for OpenStack and persistent volumes for OpenShift. The main goals of the NESE, originally funded by a $3.8M NSF Data Infrastructure Building Blocks (DIBBSs) program, are to meet the storage needs of the data revolution for science, engineering, education, and technology, particularly for researchers in the northeastern part of the U.S. NESE provides two types of storage technologies: Disk and Tape. NESE Disk provides real-time access through multiple technologies to a multi-petabyte backend storage cluster, while tape provides cold storage for inactive data or staging large-scale instrument data to be processed later. Currently, NESE consists of 50PB of disk storage and 70PB of tape storage.