Brief Overview: Research Computing Services and Shared Computing Cluster

Research Computing Services (RCS), a department within Information Services & Technology (IS&T), provides specialized hardware, software, training, and consulting for all areas of computational research at Boston University.

RCS resources are typically used in computational science and engineering, simulation, modeling, bioinformatics, genomics, data analysis, and other disciplines that require high-performance computing, massive storage, complex visualization or specialized research software. The main computational resource provided by RCS is a large Linux cluster, the Shared Computing Cluster (SCC). This cluster is available to all University faculty, their students, and their collaborators for research and for educational use in courses related to computation. The cluster comprises more than 6,000 Intel and AMD CPU cores, over 100,000 NVIDIA GPU CUDA cores, more than 50 TB of main memory, and over 2 PB of disk for research data. The compute nodes are interconnected with a variety of fabrics: QDR and FDR InfiniBand for multi-node parallel jobs, 10 GigE for data intensive jobs and GigE 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 storage systems are protected through hardware RAID and snapshots. All of the project storage spaces can be offered with dbGaP compliance for storing human genomics data.

The SCC and other Research Computing resources are located at the Massachusetts Green High-Performance Computing Center (MGHPCC) data center. 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 nearly 400 software application packages installed on the SCC. Researchers may request that additional software be installed on the cluster. Assistance in acquiring and installing site-licensed, common desktop applications such as MATLAB and SPSS is also available.

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 resources are typically used in computational science and engineering, simulation, modeling, bioinformatics, genomics, data analysis, and other disciplines that require high-performance computing, massive storage, complex visualization or specialized research software.

Resources are managed in close consultation with the Research Computing Governance Committee, the Shared Computing Cluster Faculty Advisory Committee, the BU Center for Computational Science, and the Rafik B. Hariri Institute for Computing and Computational Science & Engineering.

In conjunction with the development of Boston University's Information Technology 2015-2020 Strategic Plan, the Research Computing Governance Committee has developed and approved a high-level

[Cyberinfrastructure Plan](#) for the University. This plan augments the overall technology plan with additional initiatives to build integrated and scalable research computing and networking capabilities. The Research Computing Governance Committee is co-chaired by Gloria Waters, Vice President for [Research](#), and Ronald Corley, Associate Provost for Research for the [Boston University Medical Campus](#).

Computational Resources

The main computational resource provided by RCS is a large Linux computational cluster, the Shared Computing Cluster (SCC). This cluster is available to all University faculty, their students, and their collaborators for research and for educational use in courses related to computation. The cluster comprises more than 6,000 Intel and AMD CPU cores, over 100,000 NVIDIA GPU CUDA cores, more than 50 TB of main memory, and over 2 PB of disk for research data. The compute nodes are interconnected with a variety of fabrics: QDR and FDR InfiniBand for multi-node parallel jobs, 10 GigE for data intensive jobs and GigE for general purpose compute jobs running within a single node.

The cluster is composed of fully *shared* resources and *buy-in* resources. The *shared* resources, fully funded by the University, are available on a fair-share, no-cost basis to all faculty-sponsored research groups. The *buy-in* resources are directly funded by researchers for their own priority use. Excess *buy-in* capacity, typically about 30% overall, is returned to the shared pool for general use. Buy-in nodes may be purchased from our Web site at any time at specially negotiated rates with our vendors.

All shared and buy-in resources are managed by Research Computing at no charge to researchers. A *dedicated* service model provides fee-based management for other research computing systems that don't align with the shared or buy-in models. Finally, a *co-location* model offers a no-cost, stand-alone service providing rack space, power, cooling, and network access in our data center for use by faculty wishing to have complete control in managing their own resources.

Research Computing partners with the NSF funded TeraGrid through the Campus Champions program and can facilitate access to those resources. Please contact Research Computing staff for more information.

Data Storage

A variety of storage options are provided on the Shared Computing Cluster and through other IS&T services. On the SCC all accounts are provided with home directory space and well as scratch storage for running jobs. 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 our Storage-as-a-service offering, both at highly subsidized rates. Additionally, researchers may request storage space on our back-up storage system, sited at another location, to make a duplicate copy of critical data. This secondary storage, called STASH, is obtainable through the same service models and with the same cost structure as primary storage. All storage systems are protected through hardware RAID and snapshots. All of the project storage spaces can be offered with dbGaP compliance for storing human genomics data.

For longer term storage, IS&T offers an archive storage system. 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 for over 100,000 devices to all institutional information, communication, and computational facilities, along with the Internet, regionally aggregated resources, and advanced networks such as Internet2. On campus, tens of thousands of ports and wireless access points are interconnected via optical fiber and a robust hierarchy of high-speed routers and switches. The University has two full Class B IPv4 address space assignments.

Encrypted 802.1x wired and wireless network access is available throughout campus, including all residence halls and classrooms. The University recently invested several million dollars to ensure the entire wireless network conforms to the most current 802.11n standards. Redundant connections to the Internet and Internet2 operate at 10G with automatic failover.

Software

Research Computing offers a wide range of software packages installed on the SCC, as well as assistance in acquiring and installing site-licensed, common desktop applications such as MATLAB and SPSS. The SCC has nearly 400 application software packages installed. Researchers may request our assistance in installing additional application software packages through our Web site.

Consulting

Research Computing staff are available to assist researchers in numerous areas. These include basic use of the system, accounts and allocations, optimizing CPU and/or I/O performance, programming in a variety of languages, statistics programming and applications, code porting, code parallelization, performance measurement, and numerical methods. Short term consulting is offered at no cost. Longer term or dedicated staff time can be arranged for a fee.

Training

Research Computing offers a number of training services in a wide range of topics, including Linux basics, programming, high performance computing, data analysis and visualization.

Each semester the RCS staff offer a series of tutorials consisting of one to three hours of classroom instruction. Most of the sessions are hands-on and are designed to help researchers make effective use of the Boston University Shared Computing Cluster and its related scientific visualization resources. All the tutorials are free and open to all members of the Boston University community.

A full list of the RCS tutorials and a schedule of upcoming sessions is available on the Web. From the Web site, the slides for the most recent version of each tutorial are available.

In addition to the regularly scheduled tutorials, Research Computing staff can offer extra sessions and/or customize tutorials for a particular course, seminar, lab or research group. These customized tutorials

can be combinations of our regular materials or other similar content of specific interest to your group. Please contact us for more details by email to help@rcs.bu.edu.

Massachusetts Green High-Performance Computing Center (MGHPCC)

Boston University is a partner in the Massachusetts Green High-Performance Computing Center (MGHPCC), a collaboration of universities, industry, and the Massachusetts state government. The group has built and operates a research data center in Holyoke, Massachusetts to take advantage of the abundant source of clean renewable energy from Holyoke Gas & Electric's hydroelectric power plant on the Connecticut River. MGHPCC partners include university consortium members Boston University, Harvard University, Massachusetts Institute of Technology, Northeastern University, and the University of Massachusetts; industry partners Cisco and EMC; and the Commonwealth of Massachusetts.

MGHPCC is a world-class, high-performance computing center with an emphasis on green, sustainable computing. The green aspects of the project range from the use of clean, sustainable electric generation to power the data center to fostering research collaborations in energy, climate and the environment. The continuing development of this center is creating unprecedented opportunities for collaboration between research, government, and business in Massachusetts.

The MGHPCC has become the first university research data center to receive a LEED® Platinum certification, the highest green building ranking. The MGHPCC is also one of only 13 data centers in the country to receive a Platinum certification.

The MGHPCC is designed to support the growing scientific and engineering computing needs at five of the most research-intensive universities in Massachusetts – Boston University, Harvard University, Massachusetts Institute of Technology, Northeastern University, and the University of Massachusetts. 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 pair of 10 GigE connections, providing an aggregate capacity of 40 Gb/s, from its campus to its resources located in the Holyoke facility.