Managing Projects on the Shared Computing Cluster (SCC)
Topics for Today

- Research Computing Services
- Resources
- Projects
- Storage
- Service Units (CPU Time)
- Buy-In Program
- SCC Management Page
- Tips and Tricks

Mostly Show & Tell

High Level Overview

Please Ask Questions
Research Computing Services
Research Computing Services (RCS)

A group within Information Services & Technology at Boston University provides computing, storage, and visualization resources and services to support research that has specialized or highly intensive computation, storage, bandwidth, or graphics requirements.

Three Primary Services:

- Research Computation
- Research Visualization
- Research Consulting and Training
Me

- Research Facilitator and Administrator

- Background in biomedical engineering, technology, and bioinformatics

- Offices on both CRC and BUMC
  - Most of our staff on the Charles River Campus, some dedicated to BUMC

- Contact: help@scc.bu.edu
Resources
The Linux Virtual Lab (scc-lite)

BU’s *general purpose* Linux interactive time-sharing environment. Provides a linux environment, software packages, and low volume computation to *anyone* at of Boston University.

- Single multi-user system
- Entire BU community
- No project required
- Home directory (10GB)
The Shared Computing Cluster (SCC)

A heterogeneous Linux compute cluster for research at Boston University. Provides high-performance computing, storage, and project management for researchers at Boston University.

- 700 nodes, 15,000 CPU cores
- 250,000 GPU cores
- 3 Petabytes of disk space
- Projects to facilitate collaborative research
Projects
Projects on the SCC

Projects on the SCC allow BU Faculty, known as “Lead Project Investigator” (LPI), to separate and manage work requirements for different goals. Depending on the LPI, this can be different labs, workgroups, grants, projects or project aims.

Internally, projects allow RCS to handle resource allocation and reporting.

A project provides:

- User Control / Access Lists
- Storage Allocation
- Service Unit Allocation
- Reporting
Project Creation

A BU Faculty member can create a project by completing the webform on our website. Requests are typically processed the same day.

- A few required fields
- Many optional fields
- Can (and should) be updated later

http://www.bu.edu/tech/support/research/account-management/create-project/
Projects: User Control / Access Lists

Ability to grant or remove a person's access to project spaces and project allocations. Management of both internal and external users with optional individualized quotas.

**Account Support**
- Internal Researchers (BU)
- External Collaborators

**Optional Controls**
- Delegation Roles
- Account Expiration Dates
- User Quotas (SU, Disk)
Projects: Administrative Roles

Projects must have an LPI (Faculty), but the LPI may delegate/allow management by another BU employee, postdoc, or trusted graduate student.

- **Lead Project Investigator (LPI)**
  The BU faculty member that initiates and has full control of the project. This is typically the PI of a specific research effort.

- **IT/Administrative Contact**
  A person designated with full administrative control equivalent to the Lead Project Investigator (LPI). A second manager.
Projects: Informational Roles

An LPI may include contact information for additional staff members. Appropriate for more complex projects and initiatives (grants or buy-in).

- **Grant Principal Investigator (GPI)†**
  The person in whose name grants are awarded that support the work being carried out by this project. Useful for internal metrics.

- **Financial Contact †**
  The person we would contact if there are any payments needed regarding the project. Primarily for Storage-as-a-Service or Buy-in Program.

† This designation does not automatically create an account on the SCC.
Projects: Storage Allocation

LPIs and projects on the SCC are provided a base amount of storage for their research at no-cost. The no-cost allocations are intended to enable research at any level. Additional storage can be purchased.

Project Limits

- 1TB of storage per project
  - Up to 200GB backed up

LPI Limit

- Up to 3TB across all projects
Projects: SU Allocation

A Service Unit ("SU") is a measure of compute time on the SCC. Accounting for SU allows us to adequately provision the cluster and fairly distribute shared resources across the university.

- Each cycle, a project is allocated SU for all members to use.
- We track usage, enforce quotas, and provide reports to LPIs.
- There is no monetary charge for SU.
Projects: Responsibilities

The LPI is directly responsible for all activity on the project. RCS provides forms to automate these tasks and reports to monitor usage. You can email our Service Management Team for assistance (scfacct@bu.edu).

- Authorizing Users
- Storage Usage
- SU Usage
- Monitoring Usage (Monthly Reports)
- Annual Renewal
Projects: Lifecycle and Renewal

Projects have a life-cycle and need to be renewed until completion. Our website makes managing and updating all active projects as easy as submitting a form.

- A project lasts for 1 year.
- Can be renewed as long as the project is active.
- Can be closed at anytime.
- Archive space is available.
Storage
Storage on the SCC

RCS provides three primary storage locations and supports other IS&T services to facilitate research and collaboration.

Primary Filesystems
- Home Directory
- Backed-up project space
- Not backed-up project space

Auxiliary Filesystems (special case)
- Scratch
- STASH
- Archive
Primary Filesystems

The primary storage locations are served from a distributed filesystem located near compute infrastructure and available across the SCC.

- **Home directories**
  Individual space for long-term storage of small personal files.

- **Backed-up project space**
  Project space for long-term storage of moderate size files that are not replaceable or reproducible.

- **Non-backed-up project space**
  Project space for moderate-term storage of any size.

† Can be "restricted"
No-cost Storage Allocations

LPIS and projects on the SCC are provided a base amount of storage for their research at no-cost. The no-cost allocations are intended to enable research at any level. Additional storage can be purchased.

Project Limits

- 1TB of storage per project
  - Up to 200GB backed up

LPI Limit

- Up to 3TB across all projects
<table>
<thead>
<tr>
<th>Storage</th>
<th>Primary Filesystems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project</td>
<td></td>
</tr>
<tr>
<td>&quot;NB&quot; = &quot;Not Backed-up&quot;</td>
<td>/projectnb/{}</td>
</tr>
<tr>
<td>Project Group</td>
<td>RAID 6</td>
</tr>
<tr>
<td>Redundant File Servers</td>
<td>Daily: 30 Days</td>
</tr>
<tr>
<td>No</td>
<td>$$ $$</td>
</tr>
<tr>
<td>Project</td>
<td>Backed up</td>
</tr>
<tr>
<td>/project/{}</td>
<td></td>
</tr>
<tr>
<td>Project Group</td>
<td>RAID 6</td>
</tr>
<tr>
<td>Redundant File Servers</td>
<td>Daily: 180 Days</td>
</tr>
<tr>
<td>Yes</td>
<td>200 GB</td>
</tr>
<tr>
<td>Home Directory</td>
<td></td>
</tr>
<tr>
<td>Home</td>
<td>Individual</td>
</tr>
<tr>
<td>RAID 6</td>
<td>Daily: 180 Days</td>
</tr>
<tr>
<td>Yes</td>
<td>10 GB</td>
</tr>
<tr>
<td>Colloquial Name</td>
<td>Path</td>
</tr>
<tr>
<td>Hardware Protection</td>
<td></td>
</tr>
<tr>
<td>Snapshot Retention</td>
<td></td>
</tr>
<tr>
<td>Off-site Backup</td>
<td></td>
</tr>
<tr>
<td>Allocation Maximum</td>
<td></td>
</tr>
</tbody>
</table>
Auxiliary Filesystems (special case)

Auxiliary filesystems provide special purpose storage locations.

- **Scratch**
  A `/scratch` directory is available on every node for high-performance access while running jobs. The space is *temporary* and typically deleted upon job completion.

- **STASH**
  Second Tier Ancillary Storage Heap (STASH) allows projects to maintain an off-site copy of data. Maintained by project members. *Manual backup*.

- **Archive**
  A service provided by BU IS&T Data Archiving for secure, long-term storage of large quantities of *infrequently accessed data*. 

Restricted Data and dbGaP

An additional filesystem is available for restricted data.

- Created for dbGaP data (may comply with other data access agreements)
- Parallel and identical in policy to the /project and /projectnb filesystems
- BU network only
  - Not Internet-facing; Accessible from scc4 and compute nodes
- Not HIPAA compliant!

```
/project/projectname    →    /restricted/project/projectname
/projectnb/projectname  →    /restricted/projectnb/projectname
```
Monthly Reports

Storage Usage Summary

PROJECT: example
TITLE: Example Project
LPI: Charles Jahnke
EMAIL: cjhahnke@bu.edu
PROJECT YEAR TOTAL ALLOCATION (SUs): 2500
PROJECT END/RENEWAL DATE: 5-10-2018
BUY-IN GROUPS PROJECT IS IN: linga
URL OF THIS REPORT: https://acct.bu.edu/SCF/Projects/example/201707/example.html

#SU REPORT HERE#

********************************************************************************
Project Disk and STASH space information for project: example
The disk snapshot below was taken at approximately 4AM on August 3, 2017.

Usage numbers are rounded up to the nearest Gigabyte.

<table>
<thead>
<tr>
<th>Disk Quota</th>
<th>PROJECT</th>
<th>USAGE</th>
<th>#_OF_FILES</th>
<th>PROJECTNB</th>
<th>USAGE</th>
<th>#_OF_FILES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PROJECT</td>
<td>21</td>
<td>686</td>
<td>3</td>
<td>1652</td>
<td></td>
</tr>
<tr>
<td>cjhahnke</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>822</td>
<td></td>
</tr>
<tr>
<td>user2</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1017</td>
<td></td>
</tr>
</tbody>
</table>

********************************************************************************

Note that if any Users are identified above by a number rather than a login name, this indicates a past user of the project who no longer has an account but still has files in your project’s Project Disk and/or STASH partition(s).

You can also use the command ‘pquota’ on the system to see this data for any project. Usage is:

```
sc11% pquota [-au] [project1 project2 ...]
-u display usage information about individual users.
-a display information for ALL projects on the system
```
Service Units

“SU”
Service Units ("SU") on the SCC

A Service Unit ("SU") is a measure of *compute time* on the SCC roughly equivalent to 1 core-hour on a shared resource. Accounting for SU allows RCS to adequately provision the cluster and fairly distribute shared resources across research groups at the university.

- Service Unit ("SU")
- Allocation and Policy
- Limits
- Large Allocation Committee
- User Quotas
- Reporting
SU Measurement

A Service Unit ("SU") equals 1 hour multiplied a CPU normalization factor.

$$1 \text{ SU} = 1 \text{ hour} \times N$$

In the heterogenous cluster, some nodes are faster than others and 1 hour does not accurately measure computation across nodes. Benchmarks are run to determine CPU power and CPU time is normalized to SU using this factor.

Shared: $N = 0.7 - 1.0$  |  Buy-In: $N=0$
Allocation and Policy

Project Allocation:
The project allocation is decided by the LPI of a project. It should include the estimated compute time needed on the cluster until the renewal (1 year).

Exceeding Allocation:
It is OK! The LPI is notified via email and provided a link to request more. If no action is taken after 2 weeks, the project will be blocked from submitting jobs.

Requesting SU:
Try to accurately estimate needs until next renewal and submit the form. (It’s free for you, but we need the information for provisioning.)
Service Unit Limits

Service Unit allocations provide shared resources to projects at no-cost, but they are not unlimited and must be provided fairly across the university.

Requests exceeding the following “limits” result in review:

- Individual allocation requests greater than 10,000 SU.
- Cumulative allocation requests greater than 30,000 SU / project / year.
- Cumulative allocation requests greater than 180,000 SU / all LPI projects / yr.
Large Allocation Review Committee

The Large Allocation Review Committee (“LARC”) is composed of Faculty representing both the CRC and BUMC campuses as well as Research Computing staff. This committee reviews requests for large allocations of shared resources and grants SU based on the material provided in the application for resources.

- Funding and/or intellectual merit
- Technical efficiency study
- Resource availability review
User Quotas (Optional)

In addition to project level allocations, LPIs can set user level quotas!
Monthly Reports

Balance Summary

Breakdown: Node Type

Breakdown: User Split

PROJECT: example
TITLE: Example Project
LPI: Charles Jahnke
EMAIL: cjahnke@bu.edu
PROJECT YEAR TOTAL ALLOCATION (SU): 2500
PROJECT END/RENEWAL DATE: 5-10-2018
BUY-IN GROUPS PROJECT IS IN: linga
URL OF THIS REPORT: https://acct.bu.edu/SCF/Projects/example/201707/example.html

STARTING BALANCE (in SUs (service units)): 2447.4499
CREDITS THIS MONTH (in SUs (service units)): + 0.0000
TOTAL CHARGE (in SUs (service units)): - 0.2839

ENDINGS PROJECT BALANCE (in SUs (service units)): 2447.1660

BREAKDOWN OF USAGE BY BUY-IN GROUP (if applicable) AND SHARED NODE TYPE

BUY-IN GROUP/NODE TYPE | USAGE (Hours) | CHARGE (SU) | BUYIN?
-------------------------|--------------|------------|--------
linga                    | 0.5189       | 0.00       | 0.0000 | YES
node su factor 0.70      | 0.0031       | 0.70       | 0.0021 |
node su factor 0.85      | 0.1028       | 0.85       | 0.0074 |
node su factor 1.00      | 0.1944       | 1.00       | 0.1944 |
TOTAL                    | 0.8192       | 0.2839     |

BREAKDOWN OF USAGE BY USER (SPLITS OUT BUY-IN (if applicable) AND SHARED USAGE)
Only users who used CPU time during this period are shown.

USER LOGIN | USAGE (hours) | CHARGE (SU) | BUYIN?
------------|--------------|------------|--------
cjahnke     | 0.3922       | 0.0000     | YES
user2       | 0.1267       | 0.0000     | YES
user2       | 0.0083       | 0.0000     | YES
user3       | 0.1272       | 0.1225     | YES
TOTAL       | 0.8192       | 0.2839     |

<storage report>
Buy-In
Buy-In Overview

Researchers can purchase **compute** and **storage** resources to support their research projects. Resources are integrated into and augment the SCC, but allocation, usage and policy is set by the owner.

- Low Cost
- Professional management
- Rapid deployment

Storage Buy-In
Storage Overview

Researchers can purchase storage to extend the not-backed-up project space through the **Storage-as-a-Service** or **Buy-in Programs**.

- Utilize the SCC’s resilient, high-performance, parallel file-system
- Immediate availability for project usage
- Creation of a Storage Pool (allocated to projects)
- Exclusive access and control.

http://www.bu.edu/tech/support/research/computing-resources/service-models/buy-in/
Storage Options

Storage as a Service ("SaaS")

- Fast
- Flexible
  - Add/Remove anytime.
  - Adjust in small increments (1TB)
- Annual rates
- Cost: BLOCK FOR PRINT

Rent

Buy-In Program

- Fast *
- Structured
  - Only Add
  - 5 TB increments, 10 TB minimum
- Capital Purchase
- Cost: BLOCK FOR PRINT

Purchase

* We loan you the space immediately, purchase takes time.
Storage Pools

Result of any storage purchase is a Storage Pool, from which storage can be allocated to any projects on SCC.

- Creation of a Storage Pool
- Exclusive access and control.
- Allocate and reallocate to projects anytime
Compute Buy-In
Compute Overview

Researchers can purchase compute infrastructure (i.e. servers) and have priority access to these nodes for their projects. Currently, ~60% of the SCC is “Buy-In”.

- The ability to scale-out to additional compute nodes for their projects.
- Can purchase nodes to suit research needs (RAM, GPU, etc).
- Designate a “priority access” policy for specific research projects.
- Professional staff management of the hardware, software and security.
Compute Purchase Options

Offer eight different node types to meet specific compute requirements of researchers at BU. Please consult with us for configurations and pricing.

- **Value Node**: 28 cores, 256GB RAM, 10GbE, 1TB Disk
- **Memory**: Options offering either 512GB or 1TB RAM
- **Infiniband**: Options offering Infiniband nodes w/ either 256 or 512GB RAM
- **GPU**: Options offering either 2 or 4 NVIDIA Pascal P100 GPU cards
- **VirtualGL**: Smaller system w/ 1 NVIDIA Quadro M2000 Video Card

Priority Access and Queue Setup

When your node(s) are installed, we create a designated queue on the SCC to control usage of your nodes according to a policy you create. If you are not using the node it can become a “public” queue for shared use.

Typical Policy:

- Queue Name: yourqueuename
- Priority Access Projects: project1, project2, etc
- Priority Runtime: 720 hours
- Public Runtime: 12 hours
- Email Report: login@bu.edu (anyone)
SCC Management Page
The SCC Management Page

This new “one-stop-shopping” page designed to combine and functionally link the various pages and forms required to manage a project on SCC.

- Manage Personal User Information
- Manage Research Projects
- Manage Buy-in Storage Pools
- Manage Buy-in Compute Resources

https://acct.bu.edu/SCF/SCCMgmt/username
Manage Personal User Information

**Update Personal Information**
- Name: Charles Jahnke
- Email: cjahnke@bu.edu
- Organization: Boston University IS&T
- Address:
- Phone Number: 6176387727

**Change your default Research Computing Project**
- scv

**Project Data** for each project you are on:
- CPU Usage and Balance (during the current project year)
- /project Disk Usage and Balance
- /projectnb Disk Usage and Balance
- /restricted/project Disk Usage and Balance
- /restricted/projectnb Disk Usage and Balance
- /stash Disk Usage and Balance
- /restricted/stash Disk Usage and Balance

You can also Change your Shell, Change your Kerberos Password, or Adjust your Email Forwarding. Additional information on this is available here.
Manage Research Projects

### Manage SCC Research Projects

<table>
<thead>
<tr>
<th>adsp - Alzheimer Disease Sequencing Project</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>End/Renewal date:</strong> April 1, 2018; <strong>Project SU Balance:</strong> 29,923</td>
</tr>
</tbody>
</table>

- [Add internal (BU affiliated) user to Project](#)
- [Add external user to Project](#)
- [Manage Project Administrative Roles (IT/Administrative Contact, Grant Principal Investigator, and/or Financial Contact)](#)
- [Remove/modify users from/on your Project](#)
- [Request additional processor time (in SUs)](#)
- [Manage/request Project Disk Space](#)
- [Rent a Storage-as-a-Service Project Disk Space allocation](#)
- [Examine past monthly usage reports for this project](#)
DEMO
Manage Buy-in Storage Pools

- BUMC
- loaner
Manage Buy-in Storage Pools

DEMO
Manage Buy-in Compute Resources

- rcs
- linga
Tips and Tricks
Running jobs on specific projects

- Accounting is only as good as the data you provide.
- If working on multiple projects, specify a project.
- The qsub “-P” option
- Mandatory for BUMC

$ qsub -P projectname script.qsub
Email Forwarding

- SCC users will receive important emails
  - Job notifications: Optional
  - Project notifications (renewal/overages): Annual (+)
  - Tutorial Schedule Announcements: 3/year
  - Downtime or System Changes: Annual

- Email is sent to your BU email address.

- “External Collaborators” and “Guests” don’t have BU email mailboxes.

https://www.bu.edu/tech/services/cccs/email/forwarding/
Best Practices for Project Space

Plan Ahead! Collaborative workspaces get messy without structure.

Top Level:

- `/project (backed-up) space`: Irreproducible source/input data, code.
- `/projectnb space`: Intermediate data and results

Structure:

- Source Data
- Software/Code Directory
- User Directories
Managing Software - Global Modules

- Many centrally installed packages.
- The “module” system
  - module avail
  - module load
- Module/Software Listing
  [http://rcs.bu.edu/software](http://rcs.bu.edu/software)
- Don’t see what you need? Just ask us.
Managing Software - Personal

- Personal installs
  - Users/Projects can install and manage their own software.

- Manage Personal Libraries
  - R: `install.packages('newlibrary')`
  - Python: `pip install --user newlibrary`

- Private modules
  - Users/Projects can create “modules” for personal installs, just like the global modules.
Best Practices for Code

- **Standard software isn’t enough?**
  - You can build your own or construct pipelines.
- **Maintain code in the /project (backed-up) space.**
- **Intermediate data and results should be stored in the /projectnb space.**
- **Multiple developers?**
  - See “Version Control” tutorial and “git”
Summary

Resources
- Projects
- Storage
- Service Units
- Buy-In Program
- SCC Management Page
- Tips and Tricks

Linux Virtual Lab (scc-lite)
- General purpose
- Single multi-user system
- Entire BU community

Shared Computing Cluster (SCC)
- University Research
- Faculty sponsored “project” required
- Many nodes, reserved resources
Summary

Resources
Projects
Storage
Service Units
Buy-In Program
SCC Management Page
Tips and Tricks

BU Faculty sponsored
- Can have multiple for different purposes

Provides management features
- User Control / Access Lists
- Storage Allocation
- Service Unit Allocation (CPU-Time)
- Reporting
Summary

Resources
Projects
Storage
Service Units
Buy-In Program
SCC Management Page
Tips and Tricks

Primary Filesystems
- Home Directory (10GB)
- Backed-up project space (200GB)
- Not backed-up project space (unlimited)

Auxiliary Filesystems (special case)
- Scratch (temporary)
- STASH (off-site, $)
- Archive (IS&T Service, $)
Summary

Resources
Projects
Storage
**Service Units**
Buy-In Program
SCC Management Page
Tips and Tricks

Service Unit ("SU") \(\sim\) 1 CPU hour

Service Unit Project Allocation
- LPI requests annual allocation at renewal
- Can request more anytime
- But there limits and the LARC.
Summary

- Resources
- Projects
- Storage
- Service Units
- **Buy-In Program**
  - Storage Buy-In
    - Storage-as-a-Service ("SaaS")
    - Buy-In
    - Storage Pool
  - Compute Buy-In
    - Multiple models, different resources
    - Personal compute queue
- SCC Management Page
- Tips and Tricks
Summary

Resources
Projects
Storage
Service Units
Buy-In Program

SCC Management Page
Tips and Tricks

https://acct.bu.edu/SCF/SCCMgmt/username

One-Stop-Shopping-Management:
- Account Information
- Research Projects
  - SU, Storage, Users, etc
- Buy-in Storage Pools
- Buy-in Compute Resources
Summary

Resources
Projects
Storage
Service Units
Buy-In Program
SCC Management Page

Tips and Tricks

Planning and Accounting
- SU, charts
- Run jobs on specific projects ("-P" qsub option)

Email forwarding
- Make sure you get email from us.
- Especially for collaborators.

Best practices
- Managing project spaces
- Software: Global and Personal
- Code and Pipeline Development (git)
Questions?

RCS Tutorial Evaluation

http://rcs.bu.edu/survey/tutorial_evaluation.html

Research Computing Services Website

http://rcs.bu.edu