Yale Center for Research Computing
HPC Policies

September 25, 2020

Table of Contents

Introduction

The Yale Center for Research Computing (YCRC) is a computational core facility under the Office of the Provost created to support the advanced computing needs of the research community. The YCRC is staffed by a team of research scientists, application specialists, and systems administrators with expertise in supporting high performance computing (HPC) and computationally dependent disciplines.

The YCRC advances research at Yale by administering a state-of-the-art cyberinfrastructure, providing sustainable research services, and facilitating an interdisciplinary approach to the development and application of advanced computing and data processing technology throughout the research community.

Access and Accounts

The YCRC operates high performance computing (HPC) resources in support of faculty research in a wide range of disciplines across the university. Upon request, the YCRC will provide a principal investigator (PI) group account on the YCRC's HPC resources to any member of the ladder faculty appointed at the level of Assistant Professor or above, or any member of the research faculty appointed at the level of Research Scientist/Scholar or above. Other university faculty or staff may apply for such an account, subject to additional administrative approval. Once a group account has been created, and with approval of the

PI, members of a PI’s research group may request individual user accounts under the group account. The PI is ultimately responsible for all use of YCRC resources by members of the PI’s group.

Because of its focus on computational research conducted by faculty, research staff, postdocs, and graduate students, the YCRC provides only limited access to undergraduate students, most often when they are part of a PI’s research group or enrolled in a class that makes use of YCRC resources. In unusual circumstances, undergraduate students may request use of YCRC resources for independent research under the oversight of a faculty sponsor/advisor. Such requests will be considered on a case-by-case basis, and approval will be at the discretion of the YCRC.

The YCRC encourages appropriate use of its resources for classes. Instructors planning use of YCRC resources in their classes must consult with the YCRC and obtain prior approval at least 60 days prior to the start of the class. Accounts created for class use will be disabled and removed once the class is over.

User accounts are personal to individual users and may not be shared. Under no circumstances may any user make use of another user’s account. Users are expected to follow standard security practices (see the Data & Security section below) to ensure the safety and security of their accounts and data.

Policies regarding account creation and access to HPC resources are subject to change.

Systems Administration

The YCRC’s engineering group administers all HPC systems, including machines (known as HPC clusters), storage facilities, and related datacenter networking. Since these systems are intended to support research applications and environments, they are designed and operated to achieve maximum levels of performance, rather than high availability or other possible goals. Therefore, the engineering group’s primary responsibility is to maximize system stability and uptime while maintaining high performance. To that end, all systems are subject to regular maintenance periods as listed on the System Status Page on the YCRC website. During these maintenance periods and, rarely, at other times, some or all of the YCRC HPC resources may not be available for use.

Compute & Storage Resources

In general, the YCRC will provide each PI group with access to compute and storage resources on one YCRC cluster, subject to reasonable limits and YCRC discretion. In some circumstances, the YCRC may provide a PI group or some of its members with access to additional resources as appropriate for their computational work. (Such additional access may be temporary.) Users are expected to do their best to use resources efficiently and to release idle resources.

All users will have access to limited amounts of storage in home, project, and short-term scratch directories, free of charge. Quotas are used to limit the amount of storage and the number of files per user and/or PI group. Users are expected to do their best to delete files they no longer need. (Files in short-term scratch directories are purged automatically after 60 days.) PIs are responsible for all storage usage by members of their groups. Upon request and subject to YCRC discretion, additional storage may be provided, which may or may not incur a cost.

Data & Security Considerations on YCRC Resources

Users are not permitted to use or store high or medium risk, sensitive, or regulated data of any sort (e.g., HIPAA data, PHI, or PII) at any time on any YCRC facilities other than those explicitly designated for such data. Regardless of the sensitivity/risk classification of the data, users are not permitted to use or store, on any YCRC facility, data that are covered by a data use agreement (DUA) unless the DUA has been approved by the Office of Sponsored Projects, and the YCRC has been informed of and agreed to meet all applicable computing-related requirements of the DUA, including, but not limited to, requirements for data encryption, access control, auditing, and special actions to be taken upon removal of the data.

Security of YCRC facilities and all data stored on them is extremely important, and the YCRC takes several steps to help provide a secure computing environment, including:

- Operating the clusters in a secure datacenter, with restricted and logged access;
- Using firewalls to allow access only from Yale computers or the Yale VPN, and restricting that access to only login and data transfer servers;
- Requiring ssh key pairs (not passwords) for ssh authentication;
- Keeping operating systems up to date, and regularly applying security patches;

In addition, users bear individual responsibility for the security of YCRC clusters and their data. Accordingly, users are expected to follow standard security practices to ensure the safety and security of their accounts and data. This includes:

- Using strong passphrases on ssh keys;
- Setting permissions appropriately on data files and directories;
- Never sharing private keys, passphrases, or other login information;
- Following the terms of any Data Use Agreement that covers the data.

Further information regarding Yale cybersecurity and data classifications can be found at http://cybersecurity.yale.edu.

Backups

Except as described on the YCRC website for specific clusters, only files in users' home directories are backed up, and then only for a short time (currently approximately 30 days on most clusters). No other user files are backed up at all. Backups are stored locally, so major events affecting the HPC data center could destroy both the primary and backup copies of user files. Users should maintain their own copies of critical files at other locations. YCRC cannot guarantee the safety of files stored on HPC resources.

Resource Scheduling and Jobs

The YCRC's HPC resources are shared by many users. The YCRC uses a workload management system (Slurm) to implement and enforce policies that aim to provide each PI group with fair, but limited, access to the HPC clusters. Users may not run computationally intensive jobs on the login nodes. Instead, users must submit such jobs to Slurm, specifying the amount of resources to be allocated for the jobs. Jobs running for longer than one week are discouraged. Jobs exceeding their requested resource amounts will be terminated by Slurm with little or no warning. In order to avoid loss of data if jobs terminate unexpectedly, users are strongly encouraged to checkpoint running jobs at regular intervals.

Research Support and User Assistance

The YCRC includes a number of research support staff who can help users with a variety of tasks, including education and training, software installation, and cluster usage. The YCRC offers a growing number of classes and workshops on a variety of topics relevant to research computing. New users are particularly encouraged to attend one of the introductory training workshops to learn about the HPC clusters and become familiar with the YCRC's standard operating procedures.

The YCRC procures, installs, and maintains a number of standard software tools and applications intended for use on YCRC facilities, including its HPC clusters. Among these are compilers and languages (e.g., Python, C, C++, Fortran), parallel computing tools (e.g., MPI, parallel debuggers), application systems (e.g., R, Matlab, Mathematica), and libraries (e.g., Intel Math Kernel Library, NAG, GNU Scientific Library, FFTW). Users requiring additional software for use on YCRC facilities are encouraged to install their own copies, though the YCRC's research support staff is available to assist as needed. For customizable systems such as Python and R, the YCRC has set up procedures to enable users to easily install their own modules, libraries, or packages.

The YCRC has set up its own "ticketing system" to help manage and address inquiries, requests, and troubleshooting related to the HPC clusters. Users may contact YCRC staff by sending email to researchcomputing@yale.edu. While the time required to resolve particular issues may vary widely, users may expect an initial communication from a YCRC staff member within a reasonable time (often within one business day). Users may also obtain assistance by visiting the YCRC to meet with the research support staff, either on a first-come-first-served basis during the open office hours posted on the YCRC website, or by appointment at other times.

Account Expiration

The YCRC audits cluster accounts annually on **November 1**, and all accounts not associated with a valid netid will expire at that time. In addition, it is essential for the YCRC to be able to contact all account holders for security and communication purposes. Therefore, logins will be disabled throughout the year for accounts that are found not to be associated with a valid email address. When accounts expire, PIs are ultimately responsible for ensuring that all files are properly managed or removed. However, the YCRC reserves the right to delete expired accounts and all files associated with them, if necessary. When possible, users should arrange to transfer file ownership before leaving the group(s) with which they have been working.

Hardware Acquisition and Lifecycle

At least once per year, the YCRC will offer PIs an opportunity to purchase dedicated HPC compute and storage resources for their groups. Often, such opportunities may be coordinated with the YCRC's regular refresh and upgrade cycles for its compute and storage hardware infrastructure, but, subject to YCRC approval, PIs may request hardware purchases at other times, as well. At its discretion, the YCRC may restrict the types and quantities of dedicated HPC resources purchased for PI groups and may require that purchased resources be compatible with the YCRC's datacenter and network infrastructure and with applicable policies.

All HPC resources will be purchased from vendors of the YCRC's choosing and will be purchased with warranties acceptable to the YCRC (currently for 5 years). HPC resources will have lifetimes consistent with their warranties, commencing upon delivery, after which the resources may be decommissioned, or the lifetimes may be extended, at the YCRC's discretion.

Each PI group may access its dedicated compute resources using a private Slurm partition restricted to use by group members. The YCRC reserves the right to allow other users to make use of any idle dedicated compute resources by submitting jobs to a Slurm scavenge partition. Any such scavenge job will be subject to nearly immediate termination should a member of the owning PI group request access (via its private partition) to the compute resources being used by the scavenge job.