BioHPC: A High Performance Computing System for Life Sciences

Bioinformatics Facility
Biotechnology Resource Center
Cornell University
The BioHPC Cloud has 4 types of computers

1. **Hosted Servers**: Computers owned by individual labs but managed by the BioHPC team;
2. **Rental Servers**: Computers available for hourly rent by BioHPC users;
3. **BioHPC Network Storage**: Central storage accessible from each BioHPC server;
4. **Login Nodes**: Computers used for accessing BioHPC data storage without reservation, and for BSCB users to access their cluster. The login nodes can also be used for accessing compute servers from outside campus without VPN. No computing is allowed on these machines.

![Diagram of BioHPC Cloud with 4 types of computers: Hosted Servers, Rental Servers, BioHPC Network Storage, and Login Nodes.](image)
On each server,

- Your home directory `/home/<userID>` is located on the BioHPC network storage. You can access the same home directory from any BioHPC computer. However, it is slower to read/write files in the home directory than the local storage. **Do not compute directly on files in home directory.** You also have `/home2/<userID>` network directory which is fast – you can compute directly on files in home2, you can access the same home2 directory from any BioHPC computer.

- Local storage: There is a `/workdir` on each server which is mounted on the local hard drives inside each server. It is much faster and much more reliable to access data files in the `/workdir`. **Make sure to copy data files from your home directory to the `/workdir` before computing.**

cbsumm01
Limit on storage size

- **Your home and home2 directory:** Up to 200 GB free storage for each account. You can purchase more storage if needed. The data files in your home directory are kept as long as your account is active.

- **The working directory */workdir***: There is no restriction on size, up to the local storage capacity of each server. On **rental machines**, the */workdir* is cleared after your reservation ends. On **hosted machines**, files on */workdir* are kept permanently.
Data backup:

• By default, **there is no second copy or backup** for files in your home directory. As the BioHPC storage system uses RAID, the risk of data loss due to hard drive failure is very low. However, if files are deleted due to human errors, or your account gets compromised, they cannot be recovered.

• For important data files, we recommend that you keep a second copy. BioHPC does provide optional data backup service on user-specified directories. More information on the backup service can be found at [https://biohpc.cornell.edu/lab/userguide.aspx?a=backupguide](https://biohpc.cornell.edu/lab/userguide.aspx?a=backupguide).
The Firewall

• The three login nodes (cbsulogin, cbsulogin2, cbsulogin3) are open to the external network. You can access these three machines from anywhere in the world.

• The rest of BioHPC servers are behind the Cornell firewall. On the Cornell campus, they can be accessed directly. From outside Cornell, you can either connect through Cornell VPN (https://it.cornell.edu/cuvpn), or “ssh” via the login nodes.
Use the BioHPC Website to Manage User Access

https://biohpc.cornell.edu

Reserve a Rental Machine

Change Password

Make Payment

Mange VNC

Storage & Backup

Book Office Hours
Quick start

Step 1. Transfer data files to BioHPC

You can use one of the login nodes or any other BioHPC machines to transfer data files to your home directory. Any sftp client software will work, e.g. FileZilla (https://filezilla-project.org/download.php?show_all=1)
Quick start  Step 2. Reserve a rental machine
(skip this step if you are using a hosted machine)

1. Login to the BioHPC website (https://biohpc.cornell.edu) and click “Reservations”;
2. Reserve a computer from one of the categories. Each category is priced differently
   (https://biohpc.cornell.edu/lab/Pricing.aspx);
3. You can only access a machine while your reservation is active.
4. Reserve time longer than what is needed to run your job. You can adjust or cancel a reservation at
   any time through the “My Reservations” page.

BioHPC Cloud:
: Reservations
Quick start  Step 3. Login to the computer and run software

1. Login to the compute server through ssh (e.g., Putty on Windows, Terminal on Mac);
2. Create your working directory under `/workdir` and copy your data files there or use `/home2/userid` directory;
3. You can find all pre-installed software on this page: https://biohpc.cornell.edu/lab/labsoftware.aspx (Use “*” as wildcard when searching for a software name). Follow instructions on each software detail page;
4. If you do not find the software you want to use, email brc_bioinformatics@cornell.edu or install the software by yourself in your home directory;
5. If you are using a rental machine, make sure to copy results back to home directory after the work is done. You can always access your home directory through the login nodes.
Miscellaneous information

• When using software like Python, R, Java, GCC, Docker, Tensorflow, etc. for the first time, please consult software details page under https://biohpc.cornell.edu/lab/labsoftware.aspx (use the search box to find the entry)

• Please read our user guide to find more information about data file sharing, using GUI software, etc.: https://biohpc.cornell.edu/lab/userrguide.aspx
Miscellaneous information

• Use “screen” when running software;

• Parallelization – running jobs on multiple cores:
  ▪ perl_fork_univ.pl: perl_fork_univ.pl jobList 5
  ▪ GNU parallel
  ▪ Local scheduler (SLURM)
How to get help?

Where to find more information?
• User guide https://biohpc.cornell.edu/lab/userguide.aspx
• Workshops https://biohpc.cornell.edu/workshops.aspx

Prefer to talk to a real person?
Book an office hour https://biohpc.cornell.edu/lab/office1.aspx

More questions?
Email: support@biohpc.cornell.edu
Happy Computing!

The BioHPC Team

support@biohpc.cornell.edu