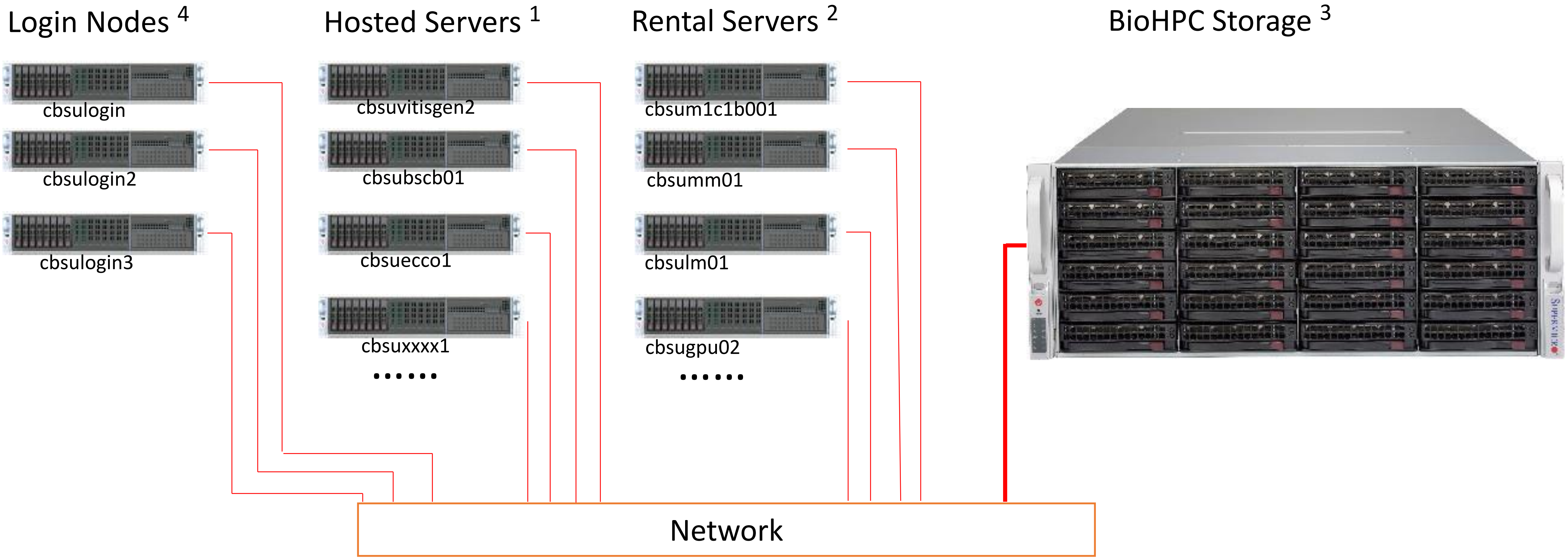
BioHPC: A High Performance Computing System for Life Sciences

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.



Login Nodes [4]

cbsulogin

cbsulogin2

cbsulogin3

Hosted Servers [1]

cbsuvitisgen2

cbsubscb01

cbsuecco1

cbsuxxxx1

……

Rental Servers [2]

cbsum1c1b001

cbsumm01

cbsulm01

cbsugpu02

……

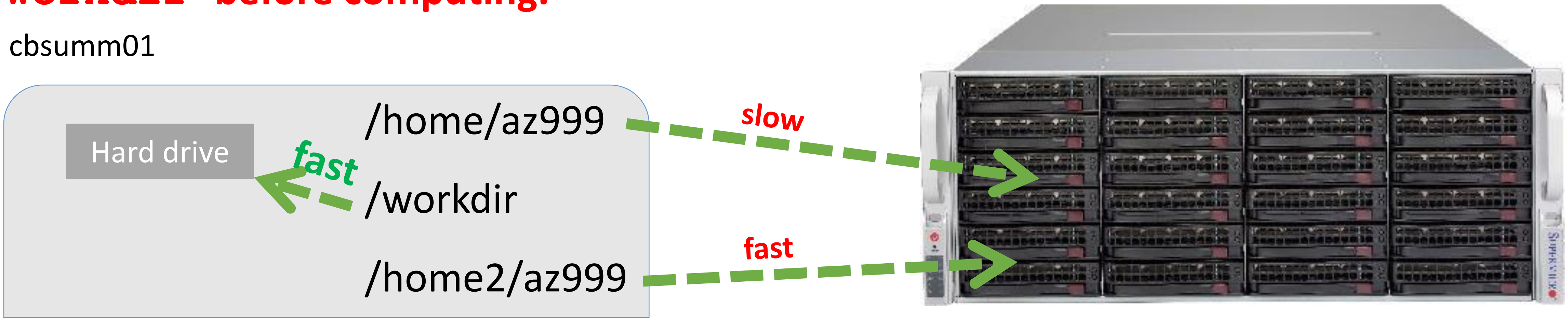
BioHPC Storage [3]

Network

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

Hard drive

fast

/home/az999

/workdir

/home2/az999

slow

fast

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 .

The Firewall

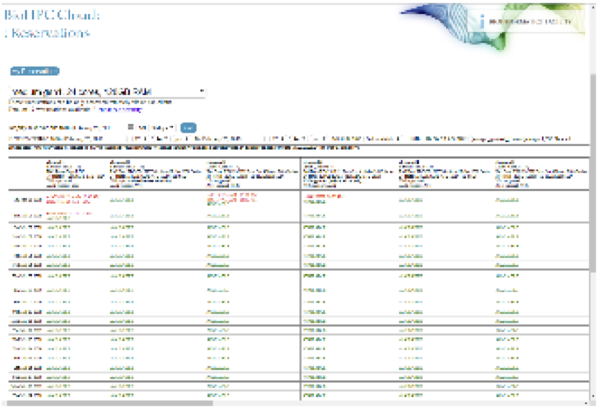
- The three login nodes (cbsulogin, cbsulogin2, cbsulogin3) are open to the external network. You can access these three machines from anywhere in the world.

- 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.

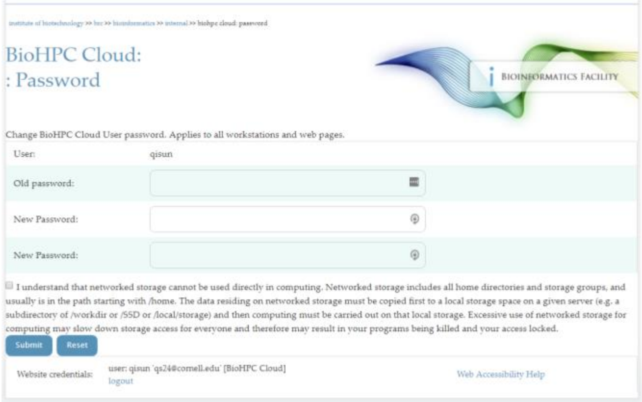
Use the BioHPC Website to Manage User Access

https://biohpc.cornell.edu

Reserve a Rental Machine


BioHPC Cloud:
: Reservations

Change Password


institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: password

BioHPC Cloud:
: Password

Change BioHPC Cloud User password. Applies to all workstations and web pages.

User:	qisun
Old password:	
New Password:	
New Password:	

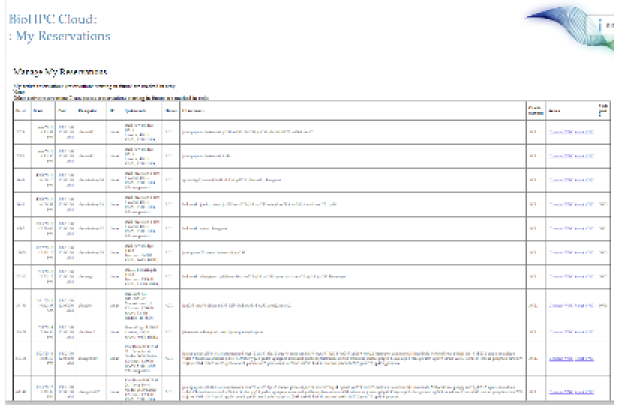
☐ I understand that networked storage cannot be used directly in computing. Networked storage includes all home directories and storage groups, and usually is in the path starting with /home. The data residing on networked storage must be copied first to a local storage space on a given server (e.g. a subdirectory of /workdir or /SSD or /local/storage) and then computing must be carried out on that local storage. Excessive use of networked storage for computing may slow down storage access for everyone and therefore may result in your programs being killed and your access locked.

Website credentials: user: qisun 'qs24@cornell.edu' [BioHPC Cloud]
logout

Make Payment


BioHPC Cloud:
: Add Hours

Mange VNC


BioHPC Cloud:
: My Reservations

Manage My Reservations

Storage & Backup

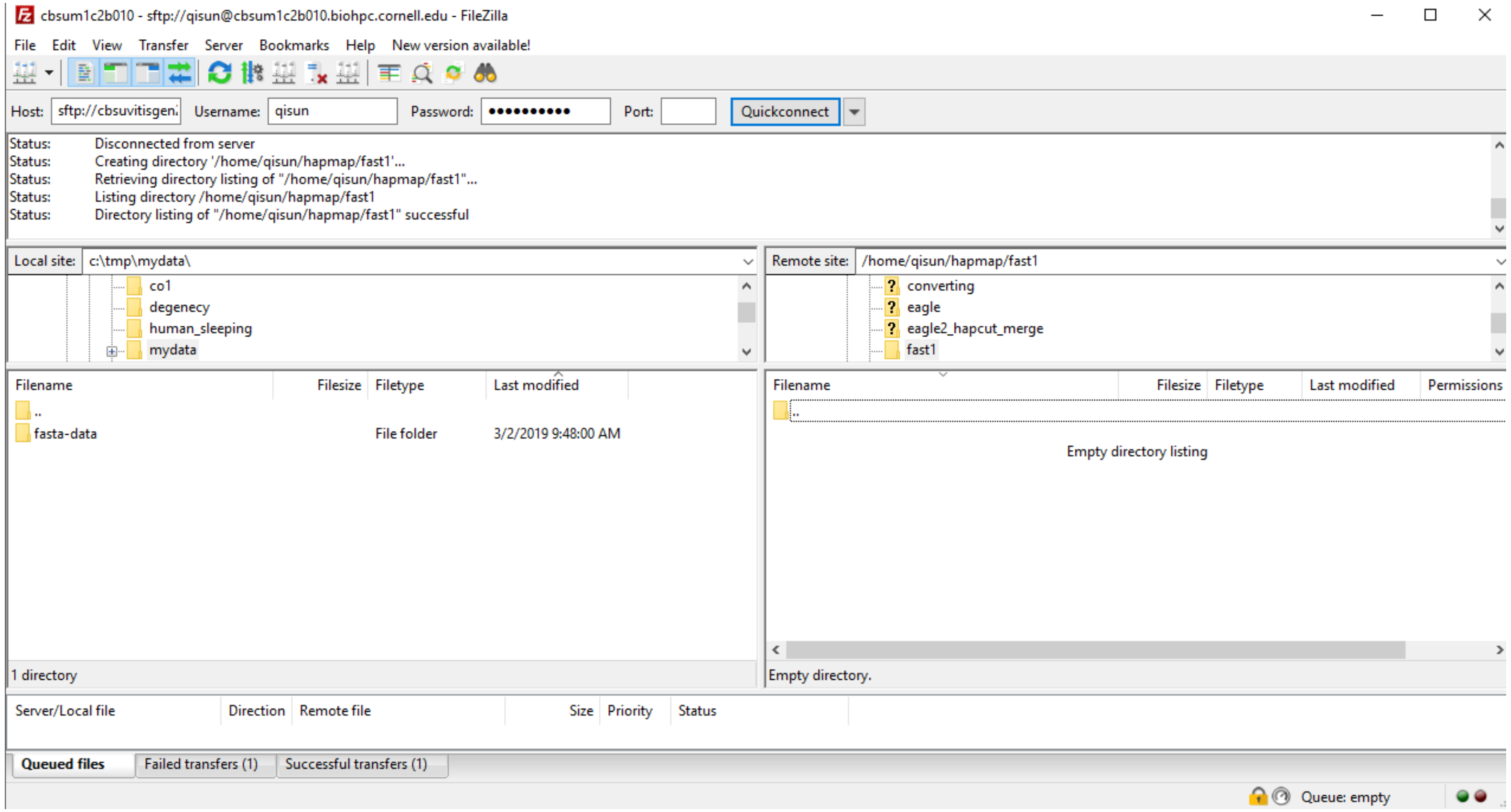

BioHPC Cloud:
: Backup Storage

Book Office Hours


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)



Host: sftp://cbsuvitisgen
Username: qisun
Password: ••••••••••
Port:

```
Status:	Disconnected from server
Status:	Creating directory '/home/qisun/hapmap/fast1'...
Status:	Retrieving directory listing of "/home/qisun/hapmap/fast1"...
Status:	Listing directory /home/qisun/hapmap/fast1
Status:	Directory listing of "/home/qisun/hapmap/fast1" successful
```

Local site: c:\tmp\mydata\

co1
degenecy
human_sleeping
mydata

Filename	Filesize	Filetype	Last modified
..			
fasta-data		File folder	3/2/2019 9:48:00 AM

1 directory

Remote site: /home/qisun/hapmap/fast1

converting
eagle
eagle2_hapcut_merge
fast1

..

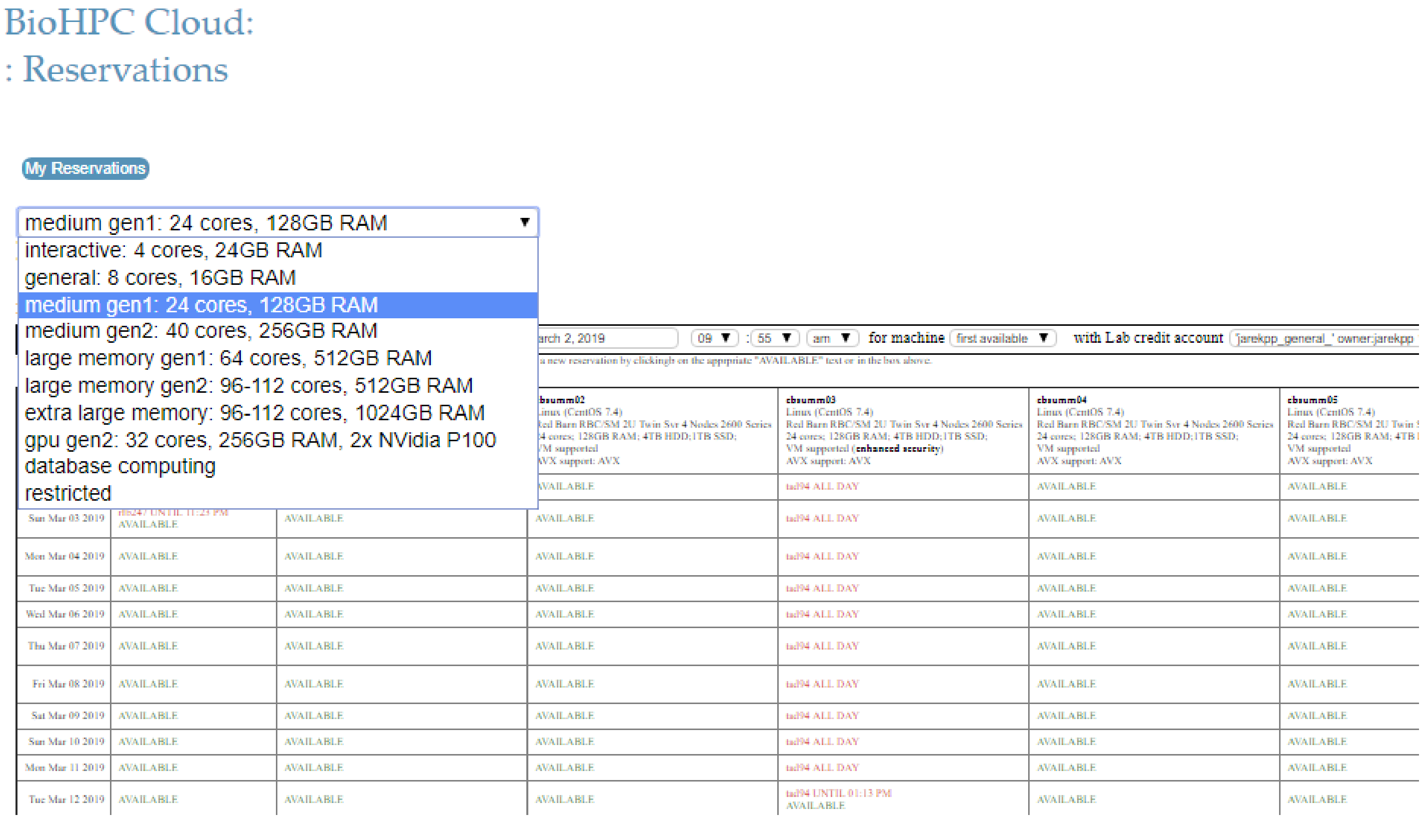
Empty directory listing

Empty directory.

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

medium gen1: 24 cores, 128GB RAM ▼

- interactive: 4 cores, 24GB RAM
- general: 8 cores, 16GB RAM
- medium gen1: 24 cores, 128GB RAM
- medium gen2: 40 cores, 256GB RAM
- large memory gen1: 64 cores, 512GB RAM
- large memory gen2: 96-112 cores, 512GB RAM
- extra large memory: 96-112 cores, 1024GB RAM
- gpu gen2: 32 cores, 256GB RAM, 2x NVidia P100
- database computing
- restricted

arch 2, 2019 | 09 ▼ : 55 ▼ | am ▼ for machine first available ▼ with Lab credit account 'jarekpp_general_' owner:jarekpp

a new reservation by clickingh on the appropriate "AVAILABLE" text or in the box above.

			bsumm02 inux (CentOS 7.4) ed Barn RBC/SM 2U Twin Svr 4 Nodes 2600 Series 4 cores; 128GB RAM; 4TB HDD;1TB SSD; M supported VX support: AVX	cbsumm03 Linux (CentOS 7.4) Red Barn RBC/SM 2U Twin Svr 4 Nodes 2600 Series 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX	cbsumm04 Linux (CentOS 7.4) Red Barn RBC/SM 2U Twin Svr 4 Nodes 2600 Series 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX	cbsumm05 Linux (CentOS 7.4) Red Barn RBC/SM 2U Twin S 24 cores; 128GB RAM; 4TB VM supported AVX support: AVX
			VAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Sun Mar 03 2019	[illegible] UNTIL 11:23 PM AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Mon Mar 04 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Tue Mar 05 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Wed Mar 06 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Thu Mar 07 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Fri Mar 08 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Sat Mar 09 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Sun Mar 10 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Mon Mar 11 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 ALL DAY	AVAILABLE	AVAILABLE
Tue Mar 12 2019	AVAILABLE	AVAILABLE	AVAILABLE	tad94 UNTIL 01:13 PM AVAILABLE	AVAILABLE	AVAILABLE

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/userguide.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.ornell.edu

Happy Computing !

The BioHPC Team
support@biohpc.cornell.edu