Introduction to BioHPC Cloud

BioHPC Cloud Workshop

Jaroslaw Pillardy

Bioinformatics Facility
Institute of Biotechnology
Cornell University

http://biohpc.cornell.edu/lab/lab.aspx
BioHPC Cloud

BioHPC Cloud is a computational resource configured and optimized for the needs of computational biology and bioinformatics

- Available to Cornell and external users
- Provides access to computing, storage and software (over 700 titles)
- Located in Rhodes Hall and accessible remotely and on-site
Renting computing resources

Hosting computing resources

Storage

Consulting and education
BioHPC Cloud: computing

Two ways of using BioHPC Cloud computing resources:

• Buy hours and run computations on BioHPC Cloud workstations/servers
  Only pay for the hours you need, avoid problems with maintaining your own hardware, software and resource allocation.

• Buy your own hardware (or better ask us to buy it for you) and host it in BioHPC Cloud
  If you need a special hardware, or you expect to use it 24/7, you can still avoid problems related to maintaining your own hardware, software and resource allocation by hosting the machine with BioHPC
BioHPC Cloud: hosting

Buy your own hardware and host it in BioHPC Cloud

• Fully customizable hardware – buy just what needed
• Access to all software and storage – same as rental
• Machine usage restricted to group: usage may be regulated internally (all have access) or by reservations (users need to reserve machine)
• No root access – installations that require full admin rights may be done by us or in a customized Docker container
• Use BioHPC scheduler or use SLURM or both
• Storage servers can also be hosted
BioHPC Cloud: storage

• 1,440 TB of networked storage available in two volumes: 1,207 TB Lustre volume and 233 TB Gluster volume.

• Very robust and scalable architecture: cluster storage based on RAID6/RAID7 servers connected by Lustre or Gluster, new servers can be added to expand storage.

• Very affordable! $95.01 per TB per year

• Users get free storage allocations with the accounts
BioHPC Cloud: Backup service

• Users are able to choose backup options using BioHPC website (what to backup, how many versions etc)

• Backup servers (currently 459TB) are located in different building (Weill Hall)
BioHPC Cloud: software

• BioHPC Cloud is pre-configured for bioinformatics with software and related software infrastructure (libraries, development tools etc.).

• All software installed as a response to our or our users computing needs.
  If you need a program that is not installed, ask us. We may install it, it depends on possible usage level and time investment required to deploy.

• 735 titles as of 4/2/2020

• Common genomic data is available locally in the Cloud: sequence and annotation databases, preformatted for common programs

http://biohpc.cornell.edu/lab/labsoftware.aspx
<table>
<thead>
<tr>
<th>Type</th>
<th>Cores</th>
<th>RAM</th>
<th>HDD</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>interactive workstations (general)</strong></td>
<td>4</td>
<td>24GB</td>
<td>4TB HDD</td>
<td>cbsuwrkst1,2,3,4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(can be used directly in 625 Rhodes)</td>
</tr>
<tr>
<td><strong>general</strong> servers</td>
<td>8</td>
<td>16GB</td>
<td>1TB HDD</td>
<td>cbsum1c1b0NN, cbsum1c2b0NN</td>
</tr>
<tr>
<td><strong>medium gen1</strong> servers</td>
<td>16</td>
<td>24-core</td>
<td>128GB RAM, 4TB HDD, 1TB SSD</td>
<td>cbumm01-16</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>64GB</td>
<td>1TB HDD</td>
<td>cbumm02-05, cbumm07-10</td>
</tr>
<tr>
<td><strong>medium gen2</strong> servers</td>
<td>40</td>
<td>256GB RAM, 8TB HDD</td>
<td>cbumm21-32</td>
<td></td>
</tr>
<tr>
<td><strong>large gen1</strong> servers</td>
<td>64</td>
<td>512GB RAM, 12TB HDD</td>
<td>cbumm02-05, cbumm07-10</td>
<td></td>
</tr>
</tbody>
</table>
CPU: an integrated circuit (a “chip”) containing computational hardware. May be more than one per server, typically 2-4.

Core: a subunit of CPU, a part of CPU capable of independent computing. It shares CPU communication infrastructure and internal memory, but not computing hardware.

Hyperthreaded Core (sometimes Thread): a CPU core subunit that shares parts of computing hardware (FPU etc) with another similar subunit. Typically there are two hyperthreaded cores per core.
“large gen2” servers

2 96-core, 512GB RAM, 12TB HDD, 1TB SSD (cbsulm12-13)
4 112-core, 512GB RAM, 12TB HDD, 1TB SSD (cbsulm14-17)
2 80-core, 512GB RAM, 10TB HDD (cbsulm21-22)
3 88-core, 512GB RAM, 7TB NVMe SSD (cbsulm18-20)

“extra large” servers

1 64-core, 1024GB RAM, 9TB HDD, 1TB SSD (cbsuem01)
1 112-core, 1024GB RAM, 12TB HDD (cbsuem02)
1 88-core, 1024GB RAM, 7TB NVMe SSD (cbsuem03)

“gpu-equipped gen2” servers

2 32-core, 256GB RAM, 12TB HDD, 2x nVidia P100 (cbsugpu02-03)
BioHPC Cloud hardware infrastructure

“networked storage: total 1,743 TB available in two volumes
   233 TB  Gluster cluster (4 servers)
   1,207 TB  Lustre cluster (10 servers)

login machines (cbsulogin, cbsulogin2, cbsulogin3)
   3  12-core, 64-128GB RAM
BioHPC servers – big picture

**Infrastructure servers**
- Lustre, Gluster, management
- No direct user access
- No reservations
- Restricted network access

**Rental servers**
- general, medium, large, extra large, GPU
- Users must make reservations
- Any BioHPC user can make reservation
- Local storage cleaned after reservation ends
- Cornell campus network only

**Hosted servers**
- Purchased by groups or departments
- Users must be added to appropriate groups
- Only users approved by PIs can be added
- Local storage persistent – no cleaning
- Cornell campus network only

**Login servers**
- For data transfers and indirect access
- Any BioHPC user can login
- No local storage
- No computing allowed
- Open to the world network
Accessing BioHPC servers from outside Cornell campus

- Only login servers are accessible directly from outside network
cbsulogin.biohpc.cornell.edu, cbsulogin2.biohpc.cornell.edu, cbsulogin3.biohpc.cornell.edu

- All other BioHPC servers are behind firewall, and NOT accessible from outside

- Preferred method of outside access is Cornell VPN, available for all Cornelians from CIT

- BioHPC users without Cornell Netid should use login servers for outside access.

- Staff or faculty of Weill Cornell Medical College can request Cornell Netid
Setting up an account

• In order to get an account fill out account request form online
  https://biohpc.cornell.edu/NewUserRequest.aspx

• All Cornell employees, students and alumni are eligible

• Any collaborators of Cornell employees, students and alumni are eligible

• External user requests are considered on case-by-case basis, depending on BioHPC overall usage.
You can manage all your Lab resources and find answers to many questions on BioHPC website

http://biohpc.cornell.edu/lab/lab.aspx
There is lots of information in “User Guide”, it covers a lot of ground and answers common questions [http://biohpc.cornell.edu/lab/userrguide.aspx](http://biohpc.cornell.edu/lab/userrguide.aspx)
If you have an account, you can login here.

Forgot password? Reset it here – it will be sent to your registered e-mail. All you need to remember is your BioHPC id – usually same as your Cornell Netid.

Need an account – fill the form.
BioHPC Cloud Login

UserId: jarakpp2
Password: ********

Login  Cencel  Forget your password?
Keep your data, especially e-mail up to date! All communications depend on e-mail …

Change your password
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one. Or get access to a hosted server(s).

- Get extra storage if needed – all users get free storage allocations, but it may not be enough

- Verify that your software is available and read instructions

- Transfer data to your Lab storage

- Make reservation(s)

- Connect to reserved workstations

- *Compute!*
Credit Accounts

In order to reserve workstations you need to have hours available in your **BioHPC Credit Account**, or you need to have access to hosted server(s).

Multiple users can be assigned to the same Credit Account, but only one person (owner) can manage it: add hours, add/remove users, view usage etc.

If your group already has a Credit Account, you can ask the owner to be added to it and get instant access to its hours.

You can set up your own Credit Account and buy hours using Cornell Account, a credit card or a PO.
Any server can be reserved using any Lab Credit Account.

The hours from Lab Credit Account, called “purchase hours” will be used at different speed depending on server type.

http://biohpc.cornell.edu/Lab/Pricing.aspx
## Credit Accounts

<table>
<thead>
<tr>
<th>CLASS</th>
<th>EQUIVALENT TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>purchase hours</td>
<td>1.000</td>
</tr>
<tr>
<td>general</td>
<td>3.052</td>
</tr>
<tr>
<td>medium gen1</td>
<td>1.960</td>
</tr>
<tr>
<td>large gen1</td>
<td>1.078</td>
</tr>
<tr>
<td>gpu-equipped gen1</td>
<td>1.674</td>
</tr>
<tr>
<td>extra large</td>
<td>0.750</td>
</tr>
<tr>
<td>large gen2</td>
<td>0.833</td>
</tr>
<tr>
<td>medium gen2</td>
<td>1.691</td>
</tr>
<tr>
<td>gpu-equipped gen2</td>
<td>0.990</td>
</tr>
</tbody>
</table>
What is the best server type to use?

Depends very much on the task at hand!

You need large memory when doing assembly and similar tasks.

However, vast majority of tasks can be done using our medium memory gen2 machines, at the same time they are quite affordable presenting the best price/performance ratio. Best overall choice!
Credit Accounts

Up-to-date price list is always online. Prices are updated annually, if they change at all.

The more time you buy at a time the cheaper it gets.

BioHPC Cloud hours NEVER expire, so you can buy a large block cheap and use them in a long term.
medium gen2
This pricing applies to gen2 medium memory generally accessible workstations, i.e. cbsumm21-32 machines (40 cores; 256GB RAM; 8TB HDD).

<table>
<thead>
<tr>
<th>purchase unit</th>
<th>current unit</th>
<th>hours</th>
<th>unit cost (Cornell)</th>
<th>cost per hour (Cornell)</th>
<th>unit cost (external)</th>
<th>cost per hour (external)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>server</td>
<td>core</td>
<td>server</td>
<td>core</td>
</tr>
<tr>
<td>60 hours</td>
<td>101.5 hours</td>
<td>101.5 hours</td>
<td>$90.23</td>
<td>$0.89</td>
<td>$112.79</td>
<td>$1.11</td>
</tr>
<tr>
<td>200 hours</td>
<td>338.2 hours</td>
<td>338.2 hours</td>
<td>$300.78</td>
<td>$0.89</td>
<td>$375.97</td>
<td>$1.11</td>
</tr>
<tr>
<td>1 month</td>
<td>1.7 months</td>
<td>1,234.6 hours</td>
<td>$878.28</td>
<td>$0.71</td>
<td>$1,097.85</td>
<td>$0.89</td>
</tr>
<tr>
<td>6 months</td>
<td>10.1 months</td>
<td>7,407.5 hours</td>
<td>$4,215.75</td>
<td>$0.57</td>
<td>$5,269.69</td>
<td>$0.71</td>
</tr>
<tr>
<td>1 year</td>
<td>1.7 years</td>
<td>14,815.1 hours</td>
<td>$6,745.20</td>
<td>$0.46</td>
<td>$8,431.50</td>
<td>$0.57</td>
</tr>
</tbody>
</table>
Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (Computational Biology Service Unit, BRC). Bioinformatics Facility is Cornell University core facility for computational biology and bioinformatics.

This website integrates dynamic computational and training resources of the facility. For more information about the facility please go to the facility main website.

Workshops
Office Hours
BioHPC Computing Lab
BioHPC NGS Data

If you would like to receive notifications about facility events, services and new developments please join our mailing list.
BioHPC Cloud: My Lab Credit Accounts

Lab Credit Accounts can now be used for all machines, there is no need to create an account for each type of machine. Conversion factors between hours used on different machine types are in the table at the bottom of the page. Each type of hours is listed for each Lab Credit Account.

Filter by:

Name:  

Description:  

<table>
<thead>
<tr>
<th>#</th>
<th>Account Name</th>
<th>Owner</th>
<th>Type</th>
<th>Machine Class</th>
<th>Hours Available</th>
<th>Hours Used</th>
<th>Hours Reserved</th>
<th>Other Users</th>
<th>Active?</th>
<th>Created</th>
<th>Description</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>CBSU</td>
<td>jasek</td>
<td>hours</td>
<td>purchase</td>
<td>102155.00</td>
<td>0.00</td>
<td>2278.50</td>
<td>0.00</td>
<td>3</td>
<td>11/28/2011 4:34:45 PM</td>
<td>CBSU staff unlimited access for restricted machines</td>
<td>Reservations Add Hours</td>
</tr>
<tr>
<td>7</td>
<td>jasekpp, general</td>
<td>jasek</td>
<td>hours</td>
<td>purchase</td>
<td>440.61</td>
<td>7930.88</td>
<td>0.00</td>
<td>0.00</td>
<td>3</td>
<td>11/29/2011 2:49:15 PM</td>
<td></td>
<td>Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account</td>
</tr>
<tr>
<td>Account ID</td>
<td>Type</td>
<td>Hours</td>
<td>General Membership</td>
<td>Large Gen1</td>
<td>Large Gen2</td>
<td>Extra Large Gen2</td>
<td>GPU-Equipped Gen2</td>
<td>Expires</td>
<td>Hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------------</td>
<td>-------</td>
<td>--------------------</td>
<td>------------</td>
<td>------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>---------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2448</td>
<td>membership</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>3/23/2017</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2447</td>
<td>new test</td>
<td></td>
<td>99.92</td>
<td>188.93</td>
<td>119.90</td>
<td>99.92</td>
<td>85.12</td>
<td>2/20/2017</td>
<td>2:21:51 PM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2461</td>
<td>member2</td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7/7/2017</td>
<td>5:31:03 PM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7 Records found. Show 1-7

Conversion factors between hours used on different machine types.

<table>
<thead>
<tr>
<th></th>
<th>General Membership</th>
<th>Medium Gen1</th>
<th>Medium Gen2</th>
<th>Large Gen1</th>
<th>Large Gen2</th>
<th>Extra Large Gen2</th>
<th>GPU-Equipped Gen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>1.000</td>
<td>0.620</td>
<td>0.620</td>
<td>0.393</td>
<td>0.328</td>
<td>0.279</td>
<td>0.440</td>
</tr>
<tr>
<td>Medium Gen1</td>
<td>1.614</td>
<td>1.000</td>
<td>1.000</td>
<td>0.635</td>
<td>0.529</td>
<td>0.451</td>
<td>0.709</td>
</tr>
<tr>
<td>Medium Gen2</td>
<td>1.614</td>
<td>1.000</td>
<td>1.000</td>
<td>0.635</td>
<td>0.529</td>
<td>0.451</td>
<td>0.709</td>
</tr>
<tr>
<td>Large Gen1</td>
<td>2.543</td>
<td>1.576</td>
<td>1.576</td>
<td>1.000</td>
<td>0.833</td>
<td>0.710</td>
<td>1.118</td>
</tr>
<tr>
<td>Large Gen2</td>
<td>3.051</td>
<td>1.891</td>
<td>1.891</td>
<td>1.200</td>
<td>1.000</td>
<td>0.852</td>
<td>1.342</td>
</tr>
<tr>
<td>Extra Large Gen2</td>
<td>3.582</td>
<td>2.220</td>
<td>2.220</td>
<td>1.409</td>
<td>1.174</td>
<td>1.000</td>
<td>1.575</td>
</tr>
<tr>
<td>GPU-Equipped Gen2</td>
<td>2.275</td>
<td>1.409</td>
<td>1.409</td>
<td>0.895</td>
<td>0.745</td>
<td>0.635</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Website credentials: user: jarekpp [jarekpp@yahoo.com] [BioHPC Lab]
Choose type of account
Membership: general membership
This membership allows you to use continuously the types of machines listed in the table below with some limits.
The limits are 'rolling' which means that they have to be met any given time, but get updated over time. For example if you make the maximum length reservation now, you will be able to extend it by one day tomorrow at the same time.

<table>
<thead>
<tr>
<th>machine</th>
<th>maximum reservation length (rolling)</th>
<th>maximum combined reservation (rolling)</th>
<th>annual cost (Cornell)</th>
<th>annual cost (external)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unit</td>
<td>hours</td>
<td>unit</td>
<td>hours</td>
</tr>
<tr>
<td>general</td>
<td>1.0 weeks</td>
<td>168.0</td>
<td>1.0 weeks</td>
<td>168.0</td>
</tr>
<tr>
<td>medium gen1</td>
<td>1.0 days</td>
<td>24.0</td>
<td>1.0 days</td>
<td>24.0</td>
</tr>
</tbody>
</table>
**BIOHPC LAB: MY LAB CREDIT ACCOUNTS**

Filter by:

Name: * | Description: *

☐ Show inactive accounts

Order by: Created | Ascending

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Owner</th>
<th>Type</th>
<th>Hours Available</th>
<th>Hours Used</th>
<th>Other</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>jarrekkp_general</td>
<td>jarrekkp</td>
<td>general</td>
<td>624.96</td>
<td>142.04</td>
<td>0</td>
<td>Edit Credit Account, Reservations, Add Hours, Purchase History, Deactivate Account</td>
</tr>
<tr>
<td>11</td>
<td>cbuelm02_jarrekkp</td>
<td>jarrekkp</td>
<td>cbuelm02</td>
<td>115.70</td>
<td>0.00</td>
<td>0</td>
<td>Edit Credit Account, Reservations, Add Hours, Purchase History, Deactivate Account</td>
</tr>
<tr>
<td>154</td>
<td>cbuelm01</td>
<td>jarrekkp</td>
<td>cbuelm01</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td>Edit Credit Account, Reservations, Add Hours, Purchase History, Deactivate Account</td>
</tr>
</tbody>
</table>

3 Records found. Show: 1-3 | New Credit Account | Transfer hours between accounts

- manage users
- purchase hours
- transfer hours between accounts
add users, can be comma-separated list of Lab id’s

delete user
BioHPC Lab: Transfer hours between lab credit accounts

FROM: jarekpp_general_149.51 hrs left

TO: jarekpp_general_149.51 hrs left

100 hours
Getting started with a new account

× You need hours: create and fund your own Lab Credit Account or get added to one

• Get extra storage if needed – all users get free storage allocations, but it may not be enough

• Verify that your software is available and read instructions

• Transfer data to your Lab storage

• Make reservation(s)

• Connect to reserved workstations

• Compute!
### Choose Type of Machines

**Medium Gen 1:** 24 cores, 128GB RAM

These workstations can be only accessed remotely via an ssh client. Total of 17 workstations available, 3 enhanced security.

Display reservations from February 4, 2019

<table>
<thead>
<tr>
<th>Machine</th>
<th>Date</th>
<th>Reservation Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium Gen 1</td>
<td>Feb 04, 2019</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td></td>
<td>Feb 05, 2019</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td></td>
<td>Feb 06, 2019</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td></td>
<td>Feb 07, 2019</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td></td>
<td>Feb 08, 2019</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td></td>
<td>Feb 09, 2019</td>
<td>AVAILABLE</td>
</tr>
</tbody>
</table>

You can modify your reservations by clicking on them, you can also add a new reservation by clicking on the button `AVAILABLE` next to the box above.
choose type of machines

choose machine

choose credit account

click on “available” link to start in time slot you would like to use
<table>
<thead>
<tr>
<th>Slot Code</th>
<th>System Type</th>
<th>RAM</th>
<th>CPU Cores</th>
<th>Storage</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>cbsu1c2b007</td>
<td>CentOS 6.2 Dell PowerEdge M600</td>
<td>16GB RAM, 1TB HDD</td>
<td>8 cores</td>
<td>16GB RAM, 1TB HDD, VM supported</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td>cbsu1c2b009</td>
<td>CentOS 6.2 Dell PowerEdge M600</td>
<td>16GB RAM, 1TB HDD</td>
<td>8 cores</td>
<td>16GB RAM, 1TB HDD, VM supported</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td>cbsu1c2b012</td>
<td>CentOS 6.2 Dell PowerEdge M600</td>
<td>16GB RAM, 1TB HDD</td>
<td>8 cores</td>
<td>16GB RAM, 1TB HDD, VM supported</td>
<td>AVAILABLE</td>
</tr>
<tr>
<td>cbsu1c2b014</td>
<td>CentOS 6.2 Dell PowerEdge M600</td>
<td>16GB RAM, 1TB HDD</td>
<td>8 cores</td>
<td>16GB RAM, 1TB HDD, VM supported</td>
<td>AVAILABLE</td>
</tr>
</tbody>
</table>

- **My reserved slots are marked in blue.**
- **Available slots are marked in green.**
- **Reserved slots are marked in red.**
### New Reservation

**Proposed reservation is shown in black. You have to click “Reserve!” button in order to finalize reservation.**

**Manage all my active reservations**

### Display reservations from February 4, 2019

<table>
<thead>
<tr>
<th>Start Date</th>
<th>End Date</th>
<th>Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00 AM</td>
<td>10:00 PM</td>
<td>192.0</td>
</tr>
<tr>
<td><strong>Mon Feb 04, 2019</strong></td>
<td><strong>Tue Feb 05, 2019</strong></td>
<td><strong>available</strong></td>
</tr>
<tr>
<td>Wed Feb 06, 2019</td>
<td>Jan 29, 2019</td>
<td>10:00 AM - 12:00 PM</td>
</tr>
<tr>
<td>Thu Feb 07, 2019</td>
<td>Jan 30, 2019</td>
<td>ALL DAY</td>
</tr>
<tr>
<td>Fri Feb 08, 2019</td>
<td>Jan 31, 2019</td>
<td>ALL DAY</td>
</tr>
<tr>
<td>Sat Feb 09, 2019</td>
<td>Jan 32, 2019</td>
<td>ALL DAY</td>
</tr>
<tr>
<td>Sun Feb 10, 2019</td>
<td>Jan 33, 2019</td>
<td>ALL DAY</td>
</tr>
<tr>
<td>Mon Feb 11, 2019</td>
<td>Jan 34, 2019</td>
<td>ALL DAY</td>
</tr>
<tr>
<td>Tue Feb 12, 2019</td>
<td>Jan 35, 2019</td>
<td>ALL DAY</td>
</tr>
</tbody>
</table>

- **Adjust dates and times**
- **Preview new dates/times**
- **Current (edited) reservation in black**
Manage My Reservations

My active reservations (reservations starting in future are marked in red):

<table>
<thead>
<tr>
<th>Res#</th>
<th>Start</th>
<th>End</th>
<th>Computer</th>
<th>OS</th>
<th>System info</th>
<th>Other users</th>
<th>Credit Account</th>
<th>Action</th>
<th>VNC port#</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>2/17/2014 3:13:01 PM</td>
<td>2/18/2014 3:00:00 PM</td>
<td>chem1ch3600</td>
<td>Linux</td>
<td>Dell PowerEdge M600</td>
<td>jasheg</td>
<td>ChemBb Connect VNC Reserve 1205</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other active reservations I can access (reservations starting in future are marked in red):

<table>
<thead>
<tr>
<th>Res#</th>
<th>Start</th>
<th>End</th>
<th>Computer</th>
<th>OS</th>
<th>System info</th>
<th>Other users</th>
<th>Credit Account</th>
<th>Action</th>
<th>VNC port#</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>2/17/2014 3:13:01 PM</td>
<td>2/18/2014 3:00:00 PM</td>
<td>chem1ch3600</td>
<td>Linux</td>
<td>Dell PowerEdge M600</td>
<td>jasheg</td>
<td>ChemBb Connect VNC Reserve 1205</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

You can connect to your Linux reserved workstations using VNC protocol at 1200x800. For more on VNC please read "Access with VNC" in the Lab's User Guide.

Add user with labid [ ] to my reservation # [ ]

New reservation from [Date and Time] [Valid Dates and Times] for the first available computer in [Location]

Go To Main Reservations Page | Go To My Reservations History
Reservations

You can only make a reservation if you have enough hours. Your hours will be “reserved” when you make a reservation so you cannot use them twice – but not immediately deducted.

Only actually used hours are charged, if you cancel your reservation before it starts nothing is charged, if you cancel before it ends only the actual hours between start and cancel are charged.

NOTE: “Time used" is defined as the wall clock time elapsed when your reservation is active - it has nothing to do with how much you actually used the CPU/RAM during this time. "Time used" reflects the span for which you had the workstation reserved.
Reservations

What happens when a reservation ends, but you are still working?

• You will not be able to login to the machine anymore.

• If you are logged in you will stay logged in until the user from the next reservation logs in.

• Your programs will continue to run, as long as the machine is not used for 2 hours – this is to give you time to extend reservation if you need it.

• When a person that has the next reservation (now current) logs in all your programs and processes will be killed and you will be logged out. Same thing will happen after 2 hours past reservation.
What if I am running a program and need to end the reservation immediately when the program ends?

- Run the program from a script – more about scripting on “Linux for Biologists”.
- After a line with program name add the following command:
  
  ```bash
  /programs/bin/labutils/endres.pl
  ```

```bash
#!/bin/bash
/home/myid/myprogram [options]
/home/myid/labutils/endres.pl
```
What if I need a scheduler to submit jobs?

No problem, SLURM is installed on all servers and you can create a “cluster” out of your reserved and/or hosted servers

manage_slurm new machine1,machine2,...,machineN

You can also activate SLURM cluster on one server only and use it to submit jobs using this server cores.

https://biohpc.cornell.edu/lab/userguide.aspx?a=software&i=689#c
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one
  
  - Get extra storage if needed – all users get free storage allocations, but it may not be enough
  
  - Verify that your software is available and read instructions
  
  - Transfer data to your Lab storage

- Make reservation(s)
  
  - Connect to reserved workstations
  
  - Compute!
Storage

• **Networked storage**
  Very large storage (1,207TB+233TB), access limited by network speed, much slower than local storage. *Unsuitable for direct computing*. Very good for storing data long-term or sharing data between workstations – networked storage is the same on all machines.

• **Local storage**
  Fast storage, especially on large memory and medium memory workstations. Small – between 1TB (general) and 12TB (large memory). *Designed to be used for computations*. 
Linux directory structure is *continuous*, i.e. regardless of the physical location of storage it all seems to be part of one directory tree starting from root (/).

Each file or directory has a unique **path** starting from root (/):

/home/jarekp/tmp/tmpfile

/programs/bin/labutils/endres.pl
Storage

Not easy to tell which storage is local and which global just by a name.

/home/jarekp
/usr/local
/workdir/jarekp
/shared_data
/local_data
Storage

If in doubt “df -h” lists all storage devices

These are network devices – starting with “computername:/”
Storage

- **Networked storage**
  - /home
  - /shared_data
  - /programs

- **Local storage**
  - /workdir
  - /SSD
  - /local
Local Storage Note

Your local storage on a rental machine (in /workdir or /SSD) is **NOT** persistent!

It will be deleted some time after your reservation ends, so you must copy all the files back to your home directory if you want to keep them.

Local directories are cleaned at 3am every night.

If the machine is not used the files may stay up to 5 days.

If the machine is used by other users, the files may be deleted on the first night ...
What if my local storage /workdir is full on my reserved server?

It should not happen, every night at 3am a script is trying to clean it, but if there is a lot of local storage available it will try to avoid deleting until after 5 days past reservation.

You can run this script yourself:

/programs/config/clean_workdir

This way 5 day rule will be ignored and all data from previous reservations will be deleted.
Storage – big picture

Networked storage

/home
/programs
/shared_storage

cbsuXXX
/workdir
/local

cbsuXXX
/workdir
/local

cbsuXXX
/workdir
/local

cbsuXXX
/workdir
/local
Home directories

Each Unix (Linux) user has a personal storage space called home directory usually referred as /home/userid.

All home directories on BioHPC Cloud are networked and they are the same on all servers.

**DO NOT RUN ANY COMPUTATIONS IN YOUR HOME DIRECTORY!**
Copy your files to /workdir/mylabid first and run computations there!

Space available for each user in home directory is limited by a quota, which depends on type of the user and his resources.

You can always see your current limits and storage under “My Storage” menu. The storage info is updated daily during the night.
Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (Computational Biology Service Unit). Bioinformatics Facility is Cornell University core facility for computational biology and bioinformatics.

This website integrates dynamic computational and training resources of the facility. For more information about the facility please go to the facility main website.

Workshops
Office Hours
BioHPC Computing Lab
BioHPC NGS Data

If you would like to receive notifications about facility events, services and new developments please join our mailing list.
Home storage quotas

If user DOES NOT have access to paid storage

- User is associated with an active Lab Credit Account. Home directory storage limit is 200 GB.

- User is associated with an active hosted hardware resource. Home directory storage limit is 200 GB.

- User is NOT associated with an active Lab Credit Account or hosted hardware. Home directory storage limit is 20 GB.

Free storage quotas cannot be combined, added to purchased storage or used for multiple accounts. They are just to make sure users can carry out common computations without purchasing extra storage.
**BIOHPC CLOUD:**

**MY STORAGE**

This page provides detailed view of your storage. Your home directory or any of the group storage directories you belong to may in turn be a part of a storage group. For detailed explanation of BioHPC Cloud storage system please click here.

**HOME DIRECTORY**

Your home directory location is `/home/jarekpp2`

<table>
<thead>
<tr>
<th>Current disk usage:</th>
<th>0.0 GB [0.000 TB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage system:</td>
<td>Lustre</td>
</tr>
<tr>
<td>Number of files:</td>
<td>230</td>
</tr>
<tr>
<td>Number of directories</td>
<td>37</td>
</tr>
<tr>
<td>Disk usage updated:</td>
<td>1/30/2020 8:01:02 PM</td>
</tr>
<tr>
<td>Current disk quota:</td>
<td>200.0 GB [0.2 TB]</td>
</tr>
<tr>
<td>disk quota type:</td>
<td>automatic - user associated with a hosted equipment group</td>
</tr>
</tbody>
</table>

- **my home directory location**
- **current usage**
- **current limit**
- **limit type**
- **add more storage or change quota**
BioHPC Cloud: My Storage

This page provides a detailed view of your storage. Your data can be stored in your home directory or in any of the group storage directories you may have access to. Your home directory may in turn be a part of a storage group. For detailed explanation of BioHPC Cloud storage system please click here.

Storage purchase history

HOME DIRECTORY

Your home directory location is /home/jarekpp

Current disk usage: 24.8 GB [0.024 TB]
Storage system: Lustre
Number of files: 707
Number of directories: 205
Disk usage updated: 1/30/2020 8:01:02 PM
Current warning threshold: 512.0 GB [0.5 TB]
Warning threshold type: purchased storage
Purchased storage credit balance: 0.274 TB-years
Purchased storage expiration date: 5/25/2031 1:21:59 AM

Explore usage by file and directory

Update home directory storage info

Browse home directory usage  Add or modify home directory storage

You have access to the following storage groups:
BioHPC Cloud: Storage Usage Browser

/home/jarekp Storage

Show files/directories greater than 0 GB Filter by Size

- 492.7 GB /home/jarekp (44498 files / 3813 directories)
- 183.2 GB scanlon (452 / 29)
- 87.7 GB tmp (1781 / 191)
- 66.6 GB Kerry (4 / 1)
- 17.7 GB 2691-ellison_QTL4E_Lp_R2.fastq.gz
- 17.2 GB 2691-ellison_QTL4E_Lp_R1.fastq.gz
- 16.2 GB 2690-ellison_QTL4E_Lk_R2.fastq.gz
- 15.5 GB 2690-ellison_QTL4E_Lk_R1.fastq.gz
- 47.2 GB workdir.copy (75 / 3)
- 24.7 GB perl_bio.2013 (383 / 28)
- 17.0 GB Wojtek (23 / 7)
- 11.7 GB mouse (41 / 1)
- 9.9 GB blastdb (206 / 7)
- 5.2 GB wojtek_data (85 / 7)
- 4.4 GB teresa (4 / 1)
Extra storage can be purchased for $95.01 per TB per year.

Extra storage can be applied to user’s home directory.

Extra storage can be shared among group of users in a storage group.

Users can have their home directories placed in a storage group; their combined storage limit is then equal to the limit of the storage group.

Please contact us to create a storage group, once created it can be managed online under “My Groups” and “My Storage”
my home directory location is still under /home so it does not belong to a storage group

but now I have access to another directory, /home/imaging_share where I can store my files too
### Backup Credit Account Status

<table>
<thead>
<tr>
<th>DATE</th>
<th>Account</th>
<th>Purchased TB-Year</th>
<th>Used TB-Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edit</td>
<td>1-25-2017</td>
<td>BackupDefaultPool</td>
<td>1.00</td>
</tr>
<tr>
<td>Edit</td>
<td>1-25-2017</td>
<td>BackupDefaultPool5</td>
<td>1.00</td>
</tr>
</tbody>
</table>

### Backup Storage List

<table>
<thead>
<tr>
<th>Source Server</th>
<th>Backup Root</th>
<th>Retention</th>
<th>Frequency</th>
<th>MinSave</th>
<th>Current Backup Size(TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cbsubscb02</td>
<td>/local/storage/jarekp</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0.02</td>
</tr>
<tr>
<td>cbsublm06</td>
<td>/local_data</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0.00</td>
</tr>
<tr>
<td>Network Storage</td>
<td>/home/jarekp</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>0.35</td>
</tr>
</tbody>
</table>

[Learn more about Purchasing Backup Credit](#)

[Learn more about Managing Backup](#)
BIOHPC CLOUD: BACKUP STORAGE

Server: Network Storage
Backup Account Pool: BackupDefaultPool

Enter Backup Root: 
Add Directory to Backup

<table>
<thead>
<tr>
<th>Source Server</th>
<th>Backup Root</th>
<th>Retention</th>
<th>Frequency</th>
<th>MinSave</th>
<th>Account</th>
<th>Edit</th>
<th>Stop Backup</th>
<th>Manage Excludes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Storage</td>
<td>/home/imaging_share/CURRENT</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>BackupDefaultPool</td>
<td>Edit</td>
<td>Stop Backup</td>
<td>Manage Excludes</td>
</tr>
<tr>
<td>Network Storage</td>
<td>/home/jarekp</td>
<td>14</td>
<td>1</td>
<td>14</td>
<td>BackupDefaultPool</td>
<td>Edit</td>
<td>Stop Backup</td>
<td>Manage Excludes</td>
</tr>
</tbody>
</table>

Website credentials: user: jarekp
password: jpd6@cornell.edu
[BioHPC Cloud]

logout
Storage

Storage can be only purchased in 1 TB-year chunks, it needs to be done up front, used storage will be subtracted every day, until your storage credit is gone. Expiration date is estimated based on current storage.

**PAYING FOR STORAGE**

Storage is purchased in **1 TB-year increments**

**Example:** 1 TB-year buys 1 TB of storage for a year, or 2 TB for ½ year, or 0.5 TB for 2 years, etc.

Usage measured every day and subtracted from the purchased amount.

Usage after N measurements (typically – days)

\[ U_N = \sum_{i=1}^{N} S_{i-1} \times \Delta t_i \]

Subtracted daily from the originally purchased storage amount, displayed on the website.

\[ \Delta t_i = 1 \text{ day (typically)} \]

\[ S_0 = 0 \text{ (first day free!)} \]
Storage: TB-years

Storage in TB-years represents an *area* and it is always preserved, but either of the rectangle sides can be changed resulting in the other one adapting: lower quota extends expiration time, higher quota shortens time span.

Both rectangles represent the same purchase: 6 TB-years.
<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase #</td>
<td>TBD</td>
</tr>
<tr>
<td>Storage</td>
<td>genomics_share</td>
</tr>
<tr>
<td>Payment type</td>
<td>Cornell Account</td>
</tr>
<tr>
<td>Cornell Account Number</td>
<td>XXXX-XXXX-XXXX-XXXX-XXXX (account is required, other optional)</td>
</tr>
<tr>
<td>Cornell Account Owner Name</td>
<td></td>
</tr>
<tr>
<td>Cornell Account Owner Cornell NetID</td>
<td></td>
</tr>
<tr>
<td>Purchasing options:</td>
<td></td>
</tr>
<tr>
<td>Current used storage</td>
<td>111.593.1 GB [108.978 TB]</td>
</tr>
<tr>
<td>Purchased storage left</td>
<td>9,674.3 GB-year [9.448 TB-year]</td>
</tr>
<tr>
<td>Add 1 units of 1.0 TB-year at $93.75 each</td>
<td></td>
</tr>
<tr>
<td>1.0TB-year for $93.75</td>
<td></td>
</tr>
<tr>
<td>Warning threshold</td>
<td>112,640.0GB (110.0TB)</td>
</tr>
<tr>
<td>Expiration date with current usage</td>
<td>9/25/2018 8:35:30 AM</td>
</tr>
</tbody>
</table>

Adding 1TB-year

This is left over from past

Expiration date will be 9/25/2018
Sharing data with external collaborators

• Use Globus – you can share for reading any of your files or directories provided the other person has Globus account

• Request temporary BioHPC account for data sharing. This allows external collaborators both inbound and outbound transfer.

• Request permanent BioHPC account for external collaborator
BioHPC Lab: Request Temporary User

You can request a temporary access to BioHPC Lab for an external or internal collaborator. You need to choose for how long and type the collaborator e-mail in the box below. Make sure the e-mail is valid, without it the collaborator will not be able to access temporary account. Temporary account allows user to log in to cbslogin or cbslogin2 and transfer data to and from BioHPC Lab. Temporary account cannot be used to purchase hours or storage, reserve machines or carry out other common tasks, it is designed for data transfer only.

Currently used temporary accounts

<table>
<thead>
<tr>
<th>account</th>
<th>requested by</th>
<th>assigned to</th>
<th>expiration date</th>
<th>action</th>
</tr>
</thead>
<tbody>
<tr>
<td>cbsguest1</td>
<td>leg23</td>
<td></td>
<td>10/6/2017 1:26:43 PM</td>
<td></td>
</tr>
</tbody>
</table>

Request temporary account

Once you submit the request the collaborator will be notified by e-mail to set up password for the temporary account. You will be granted access to all files and directories on the temporary account, you will be able to go to the temporary account home directory and copy files from there to your destination, you will be able to copy your files to the home directory. Some of the temporary account. You will be notified by e-mail that your request has been processed. Sometimes files created by temporary user will not have write or read access to the requestor, use "reset file access" link to reclaim access to the files.

NOTE: Once you are done with the temporary account please delete all your files, otherwise the next user will gain full access to the files and directories you left over in the temporary account home directory.

You can terminate temporary account at any time, and so can the person you assigned to it. Once the temporary account is terminated the access password will be reset, your access to temporary account home directory will be removed and the account will be available for another user. Do not terminate temporary account before you copied and removed all your files.

Collaborator e-mail:

Account will be valid for: 1 day

Submit

Website credentials:
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one

- Get extra storage if needed – all users get free storage allocations, but it may not be enough
  - Verify that your software is available and read instructions

- Transfer data to your Lab storage

- Make reservation(s)

- Connect to reserved workstations
  - *Compute!*
Software

https://biohpc.cornell.edu/lab/userguide.aspx?a=software
https://biohpc.cornell.edu/lab/labsoftware.aspx

Please read the information provided and consult author’s website if needed.

Some programs require special setup to run, some need to be run with full path, some are installed only on a subset of workstations.
Please check detailed information before running any program, it may contain important instruction on how to properly use the software in BioHPC Cloud.

Non-tabular Lab Software page

Filter by:

Name: [*]  [Go]  [About: [*]  [Go]

[Show Windows]  [Show Linux]

Order by Name Ascending

583 Records found. Show 1-200

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>OS</th>
<th>About</th>
<th>Installed</th>
<th>Updated</th>
<th>More</th>
</tr>
</thead>
<tbody>
<tr>
<td>ngAssembler or ngMapper</td>
<td>2.8</td>
<td>Linux</td>
<td>Assembly and alignment software for 454 or other long reads.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>a5</td>
<td>20150522</td>
<td>Linux</td>
<td>A genome assembly pipeline for bacteria and archaea.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>ABBricate</td>
<td>v0.8</td>
<td>Linux</td>
<td>Mass screening of contigs for antimicrobial resistance or virulence genes.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>Abrajuj</td>
<td>20161221</td>
<td>Linux</td>
<td>Abrajuj is a de novo assembler for PacBio and Oxford Nanopore Technologies reads.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>ABYSS</td>
<td>1.9.0</td>
<td>Linux</td>
<td>Illumina short reads assembly tool.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>AdapterRemoval</td>
<td>2.1.1</td>
<td>Linux</td>
<td>Remove adapters from sequences in either single end or paired end experiments</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>Admixtools</td>
<td>5.1</td>
<td>Linux</td>
<td>The ADMIXTOOLS package implements 5 methods described in Patterson et al (2012) Ancient Admixture in Human History</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>Admixture</td>
<td>1.23</td>
<td>Linux</td>
<td>Software tool for maximum likelihood estimation of individual ancestries from multilocus SNP genotype datasets.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>agrep</td>
<td>3.41.5</td>
<td>Linux</td>
<td>approximate GREP for fast fuzzy string searching.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>albacore</td>
<td>2.3.4</td>
<td>Linux</td>
<td>Nanopore base caller.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>Alder</td>
<td>1.03</td>
<td>Linux</td>
<td>The ALDER software computes the weighted linkage disequilibrium (LD) statistic for making inference about population admixture.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>AlleleSeq</td>
<td>1.1</td>
<td>Linux</td>
<td>Detects SNVs from ChiP-seq or RNA-seq experiments.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>ALLMAPS</td>
<td>20150710</td>
<td>Linux</td>
<td>ALLMAPS is capable of computing a scaffold ordering that maximizes the colinearity to a collection of maps, including genetic, physical or comparative maps into the final chromosome build.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>ALLPATHS-LG</td>
<td>52415</td>
<td>Linux</td>
<td>Illumina short reads assembly tool.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>AMOS</td>
<td>3.1.0</td>
<td>Linux</td>
<td>AMOS is a collection of tools and class interfaces for the assembly of DNA reads.</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
<tr>
<td>AMPHORA</td>
<td>2</td>
<td>Linux</td>
<td>AMPHORA is an Automated Phylogenomic Inference Pipeline for bacterial sequences</td>
<td></td>
<td></td>
<td>[detailed information]</td>
</tr>
</tbody>
</table>
BioHPC Cloud Software

There is 583 software titles installed in BioHPC Cloud. The software is available on all machines (unless stated otherwise in notes), complete list of programs is below, please click on a title to see details and instructions. Tabular list of software is available here.

454 gsAssembler or gsMapper, a5, ABRicato, ABriuinn, ABYYSS, AdapterRemoval, Admixtools, Admixture, agrep, albacore, Alder, AlleleSeq, ALLMAPS, ALLPATHS-LG, AMOS, AMPHORA, analysis, ANGSD, Annovar, antismash, apollo, Arlequin, aspena, atac-seq-pipeline, athena_meta, Atlas-Link, ATLAS_GapFill, ATSAS, Augustus, AWS command line interface, axo, bamtools, bamUtil, Bassel, BayeScan, Bayescenv, BBmap, BCfTools, bcftools, BCP, Beagle, Beasts2, bedops, BEDTools, bfc, bgc, bgen, bigDF, bigWig, bioawk, biobambam, Bioconductor, biom-format, BioPerl, BioPython, Birdsuite, Bismark, blast, BLAST, blast2go, BLAT, BMGE, bmflagger, Boost, Bowie, Bowtie2, BPGA, BreedingSchemaLanguage, broseq, brocc, BSseeker2, BUSCO, BWA, bwameth, cactus, canu, CAP3, cBAr, CDSU RNAseq, CCTpack, cd-hit, CEGMA, CellRanger, centrifuge, CFSAN SNP pipeline, CheckM, chromatome, Circlel, Circos, Circuitscape, CLUMP, Clustal Omega, CLUSTALW, Cluster, cmake, CNVnator, compact, CONCOCT, Conda, copyNumberDiff, cortex_var, CRISPRecFinder, CRISPRosso, CrossMap, CRT, cuda, Cufflinks, cutadapt, dadi, dadi-1.6.3_modif, ddScript, DecomSeq, deepTools, defusion, delay, destroy, DETONATE, diamond, diplocSHIC, Discover, Discover de novo, distra, DIYABC, Docker, dReg, dReg-HD, dropseq, dropseqPipe, ds, ds-utils, ecopcr, ecoprimers, edgeR, edger, eons, EIGENSOFT, EMBASSY, entropy, ephem, emrined, etc3, exabayes, exonerate, express, FALCON, FALCON_unzip, Fast-QS1, fasta, fastclaster, FastME, FastML, fastp, fastq_species, fdetect, FastQC, fastsimcoal2, fastStructure, FastTree, FASTX, fineRADStructure, fineSTRUCTURE, filSNE, flash, flash2, flexbar, Flexible Adapter Remover, Flye, FMAP, FragGeneScan, FragGeneScan-freebayes, FunGene Pipeline, GAEA, Galaxy, GATK, gatk4, GBRS, gcc, GCTA, gdc-client, GEM library, GEMMA, geneid, GeneMark, GeneMarker, Genome STRIP, GenomeMapper, GenomeStudio (Illumina), GenomicConsensus, gensim, germline, gffread, gigliq, GMAP/GSNAP, GNU Compilers, GNU parallel, gradient-4.4, graphf, graphviz, Grinder, GROMACS, GSEA, GTTools, Gubbins, Hapcalc, HAPCUT, HAPCUT2, hapflk, haploMerger, haploMerger2, HapSeq2, HarvestTools, HiC-Pro, HiCExplorer, HISAT2, HMMER, Homer, HOTSPOT, HTSeq, HUMAnN2, hyperopt, HypHy, iAssembler, ibDLD, IDBA-UD, idp-deno, igBLAST, ICoR, IGV, Ima2, Ima2p, IMAGE, imageJ, immantation, impute2, IMSA-A, INDELseek, informal, InStruct, InteMap, InterProScan, ipyrad, IQ-TREE, iRep, jags, java, jbrowse, jellyfish, JoinMap, julia, jupyter, kallisto, Kent Utilities, keras, kmer, kmerFinder, kraken, kSNP, kseq, LACHESIS, lammmps, LAST, icMLkin, LDAP, leeHom, Lep-MA3, Lighter, LINKS, LocusZoo, longranger, LUCY, LUCY2, LUMPY, lyve-SET, MACS, MACs simulator, MACS2, MAFFT, maftools, Magic-BLAST, MAKER, MAQ, MASH, MaSuRCA, Mauve, MaxBin, mcctools, ml, megabalit, MeGaMerge, MEGAN, MELT, MEME Suite, MERLIN, MetaBAT, MetaCRAG, metaCISPR, metamos, MetaPathways, MetaFlAn, MetaVelvet, MetaVelvet-SL, MGmapper, Migrate-n, mikado, Minimac4, minimap2, mira, miRDeep2, MISO (misopy), MiToBim, MixMapper, MKTest, MMAP, MMSEQ, mosdepth, mother, MrBayes, mrqFAST, msdl, MSMD, msprime, MSRA Genotype Assessor, msstats, MSMTP, mugsy,
Name: bamtools
Version: 2.5.1
OS: Linux
About: BAM file processing and filtering.
Added: 12/13/2011 2:56:40 PM
Updated: 2/27/2018 4:25:37 PM
Link: http://seqanswers.com/wiki/BamTools
Platform: all

Notes: The latest version of the program is in your PATH, and can be used directly by typing its name at the prompt:

```
bamtools [options]
```

There are 2 previous version(s) available. Any version can be accessed either by typing full path, or by adding it to the PATH and then typing its name at the prompt. NOTE: you need to set PATH only once per login/session.

<table>
<thead>
<tr>
<th>version</th>
<th>access</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1 (latest)</td>
<td>full path: /programs/bamtools-2.5.1/bin/bamtools [options]</td>
</tr>
<tr>
<td></td>
<td>add to PATH: export PATH=/programs/bamtools-2.5.1/bin:$PATH</td>
</tr>
<tr>
<td>2.3.0</td>
<td>full path: /programs/bamtools-2.3.0/bin/bamtools [options]</td>
</tr>
<tr>
<td></td>
<td>add to PATH: export PATH=/programs/bamtools-2.3.0/bin:$PATH</td>
</tr>
<tr>
<td>2.2.3</td>
<td>full path: /programs/bamtools-2.2.3/bin/bamtools [options]</td>
</tr>
<tr>
<td></td>
<td>add to PATH: export PATH=/programs/bamtools-2.2.3/bin:$PATH</td>
</tr>
</tbody>
</table>
Details for **Trinity** *(hide)*

**Name:** Trinity

**Version:** 2.8.4

**OS:** Linux

**About:** De novo transcriptome assembly.

**Added:** 12/13/2011 3:22:56 PM

**Updated:** 10/25/2018 10:13:24 PM

**Link:** [http://trinityrnaseq.github.io/](http://trinityrnaseq.github.io/)

**Notes:**
You need to use full path to the binaries:

```
/programs/trinityrnaseq-Trinity-v2.8.4/Trinity [options]
```

You can also add the program to your PATH:

```
export PATH=/programs/trinityrnaseq-Trinity-v2.8.4:$PATH
```

and then use it directly by typing program name at the prompt.

It is recommended to run on medium memory nodes (e.g., cbsumm03). A large memory machines (e.g. cbsulm05) may be needed to process larger datasets.

If using the latest version of Trinity (2.8.4), run the following before launching the Trinity script:

```
export PATH=/programs/jellyfish-2.2.3/bin:/programs/salmon-0.11.3/bin:$PATH
```

(this needs to be done once per session, and is not necessary for earlier versions of trinity).

To launch Trinity, use the full path to the executable, like so:

```
/programs/trinityrnaseq-Trinity-v2.8.4/Trinity [options] &
```

If you wish to use eXpress software to estimate read abundance via Trinity's utility `align_and_estimate_abundance.pl`, add the location of the eXpress executable to the PATH:

```
export PATH=/programs/express:$PATH
```

TransDecoder is not part of Trinity release starting from 2.1.1.

Here is the instruction of running TransDecoder on BioHPC Lab: [https://cbsu.tc.cornell.edu/lab/userguide.aspx?u=software&f=209#c](https://cbsu.tc.cornell.edu/lab/userguide.aspx?u=software&f=209#c)
You can “lock in” the program version by using full path or prepending your version to the PATH.

The commands to do it are always listed on the Lab program page.
Want to install software yourself?

- Many software titles can be installed in user space, especially Python or Perl.
- You can use Conda to install software in a separate environment.
- You can download and use Docker image. You can install software inside Docker container.

⇒ “Software installation on Linux”
⇒ “Using Docker in BioHPC Cloud “
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one
- Get extra storage if needed – all users get free storage allocations, but it may not be enough
  - Verify that your software is available and read instructions
  - Transfer data to your Lab storage
- Make reservation(s)
  - Connect to reserved workstations
  - *Compute!*
Connecting to BioHPC Cloud machines

Text-based connection: ssh (Secure SHell)

GUI (graphical) connection: X-Windows or VNC
SSH

SSH is the simplest and most useful connection.

It is text based, no graphics.

Details will be explored in “Linux for Biologists” workshop, but we will cover the basics in this workshop.
SSH - Windows

• Install PuTTY – open source SSH package for Windows

• Start PuTTY

• Type fully qualified server name you want to connect to, e.g. cbsu1c2b007.biohpc.cornell.edu

• Click “Open”
SSH - Windows

```
jarekp@cbsumlc2b007:~
login as: jarekp
jarekp@cbsumlc2b007.biohpc.cornell.edu's password:
Access denied
jarekp@cbsumlc2b007.biohpc.cornell.edu's password:
Access denied
jarekp@cbsumlc2b007.biohpc.cornell.edu's password:
Last failed login: Thu Apr 2 17:18:56 EDT 2020 from vpnuser-10-41-251-4.cuvpn.cornell.edu on ssh:notty
There were 10 failed login attempts since the last successful login.
Welcome to the BRC Bioinformatics Facility BioRPC Cloud!
server: cbsumlc2b007, 8 cores, 16GB RAM, CentOS 7.6.1810
User: jarekp attempting to use machine cbsumlc2b007 at Thu Apr 2 17:19:13 2020
[jarekp@cbsumlc2b007 ~]$
```
SSH – Mac and Linux

• Open a terminal (command window)

• Type
  `ssh userid@cbsuxxx.biohpc.cornell.edu`

• Provide password when prompted

• Example:
  `ssh jarekp@cbsum1c2b007.biohpc.cornell.edu`
SSH – Mac and Linux

[root@obsulogin2 ~]# ssh jarekp@cbsumlc2b007.biohpc.cornell.edu
The authenticity of host 'cbsumlc2b007.biohpc.cornell.edu (128.04.181.158)' can't be established.
ECDSA key fingerprint is SHA256:5WbMFAc6CF00+Z52n2+wFyxRWkUEX5iYCM/s55TQh/v.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'cbsumlc2b007.biohpc.cornell.edu,128.04.181.158' (ECDSA) to the list of known hosts.
jarekp@cbsumlc2b007.biohpc.cornell.edu's password:
Last login: Thu Apr 2 17:24:18 2020 from obsulogin

Welcome to the BRC Bioinformatics Facility BioHPC Cloud!
server: cbsumlc2b007, 8 cores, 16GB RAM, CentOS 7.6.1810

[jarekp@cbsumlc2b007 ~]$
Logging in to CBSU machines from outside of Cornell

Two ways to connect from outside:

- Install and run the Cornell VPN software ([http://www.it.cornell.edu/services/vpn](http://www.it.cornell.edu/services/vpn)) to join the Cornell network, then proceed as usual

- Log in to cbsulogin.biohpc.cornell.edu (or cbsulogin2 or cbsulogin3):
  
  ```
  ssh jarekp@cbsulogin.biohpc.cornell.edu  # using PutTY or other ssh client program
  ```
  
  Once logged in to cbsulogin, ssh further to your reserved machine

  ```
  ssh jarekp@cbsum1c2b007.biohpc.cornell.edu
  ```

Backup login machine is cbsulogin2.biohpc.cornell.edu and cbsulogin3.biohpc.cornell.edu

[https://biohpc.cornell.edu/lab/doc/BioHPCLabexternal.pdf](https://biohpc.cornell.edu/lab/doc/BioHPCLabexternal.pdf)
SSH – connect from outside without VPN

```
[jarekpp@cbsumlc2b007 ~]$ ssh jarekpp@cbsumlogin.biohpc.cornell.edu
jarekpp@cbsumlogin.biohpc.cornell.edu's password:
Last login: Mon Apr 6 18:05:28 2020 from 23.133.160.63
******************************************************************************
  * Welcome to BioHPC Lab login server!                                     *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  * This machine is for remote logins and                                  *
  * data transfers only - NOT for computing.                              *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  * Software installed on BioHPC Lab machines                            *
  * is NOT supposed to be installed nor work                              *
  * on this machine. If you want to do computing                         *
  * please go to the compute nodes!                                      *
  * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
  * For more info on BioHPC Lab please go to                             *
  * http://cbsu.tc.cornell.edu/lab/lab.aspx                               *
  *----------------------------------------------------------------------*
[jarekpp@cbsumlc2b007 ~]$ ssh cbsumlc2b007
jarekpp@cbsumlc2b007's password:
Last login: Mon Apr 6 10:07:05 2020 from cbsumlogin
----------------------------------------------------------------------
Welcome to the BRC Bioinformatics Facility BioHPC Cloud!
server: cbsumlc2b007, 8 cores, 16GB RAM, CentOS 7.6.1810
----------------------------------------------------------------------
[jarekpp@cbsumlc2b007 ~]$ w
10:07:38 up 241 days, 22:24, 5 users, load average: 0.11, 0.06, 0.05
USER     ITY  FROM       LOGINT IDLE JCPU PCPU  WHAT
jarekpp pts/0  cbsumlogin 18:07  2.00s 0.08s 0.01s  w
[jarekpp@cbsumlc2b007 ~]$`
```
Logging **out** of a Linux machine

- While in terminal window, type `exit` or `Ctrl-D` - this will close the current terminal window.
Exercise: connect to your assigned workstations using ssh

• Find your assigned machine on the list on workshop page https://biohpc.cornell.edu/ww/machines.aspx?i=132

• Windows: open PuTTY program, enter your workstation name and connect. Provide your username and password when prompted.

• Linux or Mac: Open terminal window and type ssh command “ssh labid@server.biohpc.cornell.edu”. Provide your username and password when prompted.

• Once connected check storage configuration (df -h) and who else is there (w).
VNC is a remote desktop connection for Linux. It will open full graphical environment the same way one is created when logging in via a graphical console.

Please note that it is MUCH slower way to connect since lots of graphical data needs to be sent over the network. SSH is much faster and should be used when no graphics is needed.

VNC is available from Cornell campus network, from outside VPN or tunneling is required.

VNC stands for Virtual Network Computing
Logging in to a Linux workstation via VNC client

(GUI)

In web browser, navigate to http://biohpc.cornell.edu/, log in (if not yet logged in), click on **User:your_id**, select tab **My Reservations**

Click “Connect VNC”, to initialize VNC connection, or “Reset VNC” re-initialize existing connection

Select resolution you want

You can connect to your Linux reserved workstations using VNC protocol at **1280x800** from this page, for more on VNC please read “Access with VNC” in the Lab’s User Guide
Logging in to a Linux workstation (GUI)
Logging in to a Linux workstation (GUI)

You need software client to connect to your machine via VNC.

We recommend RealVNC VNC Viewer for all platforms.
Logging in to a Linux workstation (GUI)
Logging in to a Linux workstation via VNC client (GUI)

Right-click anywhere within desktop, select **Open Terminal** .... or .... click **Applications -> Accessories -> Terminal**
VNC sessions are *persistent*.

VNC sessions run even when the client is disconnected. Note: SSH can also do that with *screen* or *tmux*.

If you need to reset the session you need to use “Reset VNC” link.

Equivalent to Windows Remote Desktop.
Connecting with VNC form external network without VPN
Mac and Linux

• Enable your VNC connection first (see slide 82)

• Open local terminal window on your Mac or Linux computer

• Use the following command to connect to BioHPC. You can replace cbsulogin with cbsulogin2 or cbsulogin3, cbsuxxx with your server name, 5901 with your port no and biohpcid with your BioHPC userid.

  
  ssh -N -L 5901:cbsuxxx:5901 biohpcid@cbsulogin.biohpc.cornell.edu

• Now you can connect to your VNC by typing localhost:5901 in your VNC Viewer software.
Connecting with VNC form external network without VPN Windows

- Enable your VNC connection first (see slide 82). Note what is your VNC port.

- Open your PuTTY and fill out cbsulogin.biohpc.cornell.edu (or cbsulogin2 or cbsulogin3) as target server.

- On the left panel scroll down to Connection -> SSH -> Tunnels
Enable your VNC connection first (see slide 82). Note what is your VNC port. Type the port as shown below with the destinations server name and click Add. Now you can connect to your VNC by typing localhost:5901 in your VNC Viewer software.
Exercise: connect to your assigned workstation using VNC

• Go to “My Reservations” page
  http://biohpc.cornell.edu/lab/lab.aspx, log in, click on “My Reservations” menu link

• Choose resolution (depends on your monitor)

• Click on “Connect VNC”

• Follow prompts

• Open terminal window in the VNC desktop by right-click on the desktop background and choosing “Open Terminal”.

• Disconnect (close browser windows) and then reconnect. Is the session still alive?
Connecting to BioHPC Cloud

GUI (graphical) connection: X-Windows

Not persistent – programs will get killed when client disconnects, even when using screen or tmux.

Often faster than VNC
Connecting to BioHPC Cloud: X-Windows

• Install X-Windows software on your computer. We recommend MobaXterm free (http://mobaxterm.mobatek.net/download.html)

• Start MobaXterm

• Connect to BioHPC Cloud machine using PuTTY. Make sure X11 forwarding is enabled. X11 is a synonym for X-Windows
Connecting to BioHPC Cloud: X-Windows

Windows

• Start your GUI (graphical) software in terminal window. For testing you may try eog (Linux image viewer) or firefox.

• You may need to allow MobaXterm connections through Windows firewall.

• New window with your GUI program will appear. The program will physically run on the BioHPC Cloud machine, but it will display graphics on your local computer.
Connecting to BioHPC Cloud: X-Windows
Linux or Mac

• Connect to BioHPC Cloud machine using ssh with X11 forwarding:
  ssh -X userid@workstation.biohpc.cornell.edu

• Start your GUI program in remote terminal window. For testing you may try eog (Linux image viewer) or firefox.

• New window with your GUI program will appear. The program will physically run on the BioHPC Cloud machine, but it will display graphics on your local computer.

• NOTE: Mac machines no longer have X11 as a part of their system, it must be installed by user, XQuatrz is what we know works
Connect to BioHPC Cloud: X-Windows from external network no VPN Windows

• Make sure X11 forwarding is enabled (see slide 94).

• Connect to cbsulogin.biohpc.cornell.edu (or cbsulogin2 or cbsulogin3)

• From cbsulogin connect to your target server (cbsuxxx) with the following command

  ```
  ssh -X cbsuxxx
  (replace cbsuxxx with you target server name)
  ```

• Make sure MobaXTerm is running and start your X11 application
Open local terminal window on your Mac or Linux computer

Use the following command to connect to BioHPC. You can replace cbsulogin with cbsulogin2 or cbsulogin3, cbsuxxx with your server name, 5901 with your port no and biohpcid with your BioHPC userid.

```bash
ssh -X -t -t biohpcid@cbsulogin.biohpc.cornell.edu "ssh biohpcid@cbsuxxx -X"
```

NOTE: Mac machines no longer have X11 as a part of their system, it must be installed by user, XQuatrz is what we know works
Exercise: connect to your assigned workstation with X-Windows

- Connect to your workstations using ssh with X11 forwarding enabled
- Windows: start MobaXterm
- Start eog in remote terminal.

`eog /home/jarekp/picture.jpg`
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one
- Get extra storage if needed – all users get free storage allocations, but it may not be enough
  - Verify that your software is available and read instructions
  - Transfer data to your Lab storage
- Make reservation(s)
- Connect to reserved workstations
  - *Compute!*
Transferring data

• sftp or scp (secure file transfer protocol)
  Transfer can be done to and from BioHPC machines.

• Globus
  Transfer can be done to and from BioHPC machines.

• wget
  Transfer can be done to BioHPC machines only.
File transfer: wget
from web- and ftp sites to lab workstations

Option 1: run **wget** command on the workstation (if you know the URL of the file)

- Examples:

  
  ```bash
  ```

  (will download the file BLOSUM100 from the NCBI FTP site and deposit it in the current directory under the name BLOSUM100)

- another Example (the following should be typed on one line):

  ```bash
  wget -O e_coli_1000_1.fq "http://cbsuapps.biohpc.cornell.edu/Sequencing/showseqfile.aspx?cntrl=646698859&laneid=487&mode=http&file=e_coli_1000_1.fq"
  ```

  (the command above can be used to download files given by complicated URLs; note the ““” marks around the link and the –O option which specifies the name you want to give the downloaded file)
Exercise: download BLOSUM100 with wget

- Connect to your workstation using ssh
- Type wget command

```bash
```
File transfer: sftp
between PC or Mac and a BioHPC workstation

On Windows PC: install and use your favorite sftp client program, such as
• winscp: http://winscp.net/eng/index.php
• CoreFTP LE: http://www.coreftp.com/
• FileZilla (client): http://filezilla-project.org/
• ... others...
• When connecting to Lab workstations from a client, use the sftp protocol (port 22). You will be asked for your user name and password (the same you use to log in to the lab workstations).
• Transfer text file in text mode, binary files in binary mode (the “default” not always right).
• All clients feature
  • File explorer-like graphical interface to files on both the PC and on the Linux machine
  • Drag-and-drop functionality

On a Mac: file transfer program is fetch (recommended by Cornell CIT)
• https://it.cornell.edu/managed-servers/transfer-files-using-fetch
• graphical user interface
• Drag-and-drop functionality
But FileZilla has now also Mac version which I like best ...
Connecting to cbsslogin.biohpc.cornell.edu...
Connected to cbsslogin.
Retrieving directory listing...
Listing directory /home/jarekp
Directory listing of "/home/jarekp" successful
Exercise: download BLOSUM100 file from Lab machine to your local computer

- Connect to your workstation using sftp program (FileZilla)
- Download BLOSUM100 – you got it there before using wget
Transferring data: Globus

Globus Online is an online system based on GridFTP engine that supports fast and reliable data transfer and can be scheduled and controlled online in a web browser.

Data can be transferred between endpoints, which can be data servers or Globus Connect clients.

BioHPC Cloud Globus endpoint servers are is biohpc#cbsulogin, biohpc#cbsulogin2 and biohpc#cbsulogin3

You can transfer data to another server or to your computer (using Globus Connect client).

http://biohpc.cornell.edu/lab/doc/Globus_at_BioHPC_Lab.pdf
https://www.globusonline.org/quickstart/
Protected Data Support

Working with PHI, PII, or CUI?
Need to manage HIPAA-regulated data?
Globus has you covered!

LEARN MORE
REQUEST PRICING

Research data management simplified.

TRANSFER  SHARE  PUBLISH  BUILD
Log in to use Globus Web App

Use your existing organizational login
e.g., university, national lab, facility, project

Look-up your organization...

Didn’t find your organization? Then use Globus ID to sign in. (What's this?)

Continue

Or

Google Sign in with Google  ORCID Sign in with ORCID ID
Log in to use Globus Web App

Use your existing organizational login

e.g., university, national lab, facility, project

- Korean Access Federation
- Cornell University
- Weill Cornell Medical College
- University College Cork
- Texas A&M University-Corpus Christi

Sign in with Google  Sign in with ORCID ID
Click in the Collection field above to begin

Watch a two minute tour of what's new
Please authenticate to access this collection

Login Server

cbsulogin3.biohpc.cornell.edu

Username

jarekp

Password

Crowd123!

Advanced

Authenticate
### Collection Details

**Collection:** biohpc#cbsulogin3

**Path:** /~/

### File Manager Contents

<table>
<thead>
<tr>
<th>File Name</th>
<th>Last Modified</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_200prok</td>
<td>07/21/2011 01:05pm</td>
<td>548.99 KB</td>
</tr>
<tr>
<td>4913</td>
<td>05/29/2018 05:32pm</td>
<td>0 B</td>
</tr>
<tr>
<td>783_3_3944_N_PhiX_R1.fastq.gz</td>
<td>03/11/2013 03:50pm</td>
<td>1.37 KB</td>
</tr>
<tr>
<td>aaa</td>
<td>05/29/2018 05:32pm</td>
<td>24 B</td>
</tr>
<tr>
<td>aaa~</td>
<td>05/29/2018 05:32pm</td>
<td>0 B</td>
</tr>
<tr>
<td>aaaa</td>
<td>07/14/2014 03:01pm</td>
<td>5.10 GB</td>
</tr>
<tr>
<td>aaaaa</td>
<td>10/17/2012 01:56pm</td>
<td>9 B</td>
</tr>
<tr>
<td>addzero1.pl</td>
<td>01/30/2012 05:15pm</td>
<td>602 B</td>
</tr>
<tr>
<td>Home</td>
<td>01/20/2015 05:15pm</td>
<td>1.20 GB</td>
</tr>
</tbody>
</table>
Exercise: Login to Globus using your BioHPC account
Getting started with a new account

- You need hours: create and fund your own Lab Credit Account or get added to one
- Get extra storage if needed – all users get free storage allocations, but it may not be enough
- Verify that your software is available and read instructions
- Transfer data to your Lab storage
- Make reservation(s)
- Connect to reserved workstations

• Compute!
Exercise 1. Connecting with ssh

Exercise 2. Connecting with VNC

Exercise 3. Connecting with X11 (X-Windows)

Exercise 4. Download data with wget

Exercise 5. Download data with FileZilla (sftp)

Exercise 6. Connecting with Globus