Introduction to BioHPC Cloud

BioHPC Workshop

Jaroslaw Pillardy

Bioinformatics Facility
Institute of Biotechnology
Cornell University

http://cbsu.tc.cornell.edu/lab/lab.aspx
BioHPC

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 440 titles)

• Located in Rhodes Hall and accessible remotely and on-site
BioHPC Cloud

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 the Lab
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 install scheduler per request
• Storage servers can also be hosted
BioHPC Cloud: storage

• 988TB of networked storage available in two volumes: 755TB Lustre volume and 233TB 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! $93.75 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 270TB) 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.

• 510 titles as of 8/21/2018

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

http://cbsu.tc.cornell.edu/lab/labsoftware.aspx
BioHPC Cloud hardware infrastructure

interactive workstations with nice consoles (“general”):
4 4-core, 24GB RAM, 4TB HDD
(cbsuwrkst1,2,3,4 – can be used directly in 625 Rhodes)

“general” servers:
32 8-core, 16GB RAM, 1TB HDD
(cbsum1c1b0NN, cbsum1c2b0NN)

“medium gen1” servers
1 16-core, 64GB RAM, 1TB HDD
16 12-core, 128GB RAM, 4TB HDD, 1TB SSD
(cbsumm01-16)

“medium gen2” servers
12 40-core, 256GB RAM, 8TB HDD
(cbsumm21-32)

“large gen1” servers
8 64-core, 512GB RAM, 12TB HDD
(cbsulm02-05, cbsulm07-10)
<table>
<thead>
<tr>
<th>Server Type</th>
<th>Count</th>
<th>Description</th>
<th>Model Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>“large gen2” servers</td>
<td>6</td>
<td>96-core, 512GB RAM, 12TB HDD, 1TB SSD</td>
<td>cbsulm12-13, cbsulm14-17</td>
</tr>
<tr>
<td>“extra large” servers</td>
<td>2</td>
<td>64-core, 1024GB RAM, 9TB HDD, 1TB SSD</td>
<td>cbsuem01</td>
</tr>
<tr>
<td>“gpu-equipped gen2” servers</td>
<td>2</td>
<td>32-core, 256GB RAM, 12TB HDD, 2x nVidia P100</td>
<td>cbsugpu02</td>
</tr>
<tr>
<td>Networked storage</td>
<td></td>
<td>Total 988TB available in two volumes</td>
<td></td>
</tr>
<tr>
<td>Gluster cluster</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lustre cluster</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Login machines</td>
<td>3</td>
<td>12-core, 64GB RAM, 1.5TB HDD</td>
<td>cbsulogin, cbsulogin2, cbsulogin3</td>
</tr>
</tbody>
</table>
Setting up an account

- In order to get an account fill out account request form online
  https://cbsu.tc.cornell.edu/NewUserRequest.aspx

- All Cornell employees, students and alumni are eligible

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

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

http://cbsu.tc.cornell.edu/lab/lab.aspx
There is lots of information in “User Guide”, including some useful workflows (like “RNA-Seq” workflow)

http://cbsu.tc.cornell.edu/lab/userguide.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 Lab id – usually same as your Cornell Netid.

Need an account – fill the form
Change your password

Keep your data, especially e-mail up to date! All communications depend on e-mail ...
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!*
Lab Credit Accounts

In order to reserve workstations you need to have hours available in your Lab Credit Account for rental machines or access to hosted server(s).

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

If your group already has a Lab 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 Lab Credit Account and buy hours using Cornell Account or a credit card.
Lab Credit Accounts

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://cbsu.tc.cornell.edu/Lab/Pricing.aspx
# Lab 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>3.081</td>
</tr>
<tr>
<td>medium gen1</td>
<td>1.972</td>
</tr>
<tr>
<td>large gen1</td>
<td>1.082</td>
</tr>
<tr>
<td>gpu-equipped gen1</td>
<td>1.683</td>
</tr>
<tr>
<td>extra large</td>
<td>0.752</td>
</tr>
<tr>
<td>large gen2</td>
<td>0.835</td>
</tr>
<tr>
<td>medium gen2</td>
<td>1.700</td>
</tr>
<tr>
<td>gpu-equipped gen2</td>
<td>0.993</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, most 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!
Lab Credit Accounts

Up-to-date price list is always online. Prices are updated at the end of June, 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>60 hours</td>
<td>102.0 hours</td>
<td>102.0 hours</td>
<td>$90.23</td>
<td>$0.88</td>
<td>$112.79</td>
<td>$1.11</td>
</tr>
<tr>
<td>200 hours</td>
<td>340.0 hours</td>
<td>340.0 hours</td>
<td>$300.78</td>
<td>$0.88</td>
<td>$375.97</td>
<td>$1.11</td>
</tr>
<tr>
<td>1 month</td>
<td>1.7 months</td>
<td>1,241.1 hours</td>
<td>$878.28</td>
<td>$0.71</td>
<td>$1,097.85</td>
<td>$0.88</td>
</tr>
<tr>
<td>6 months</td>
<td>10.2 months</td>
<td>7,446.4 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>14,892.8 hours</td>
<td>$6,745.20</td>
<td>$0.45</td>
<td>$8,431.50</td>
<td>$0.57</td>
</tr>
</tbody>
</table>
manage your Lab Credit Accounts
**BIOHPC LAB: MY LAB CREDIT ACCOUNTS**

Lab Credit Accounts can be now 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 are 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>7</td>
<td>jarekapp_general</td>
<td>jarekapp</td>
<td>hours</td>
<td>purchase 1</td>
<td>149.61</td>
<td>93.52</td>
<td>0.00</td>
<td>3</td>
<td>yes</td>
<td>11/29/2011 2:49:15 PM</td>
<td></td>
<td>[Edit Credit Account]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>general</td>
<td>456.51</td>
<td>285.36</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>[Credit Account Users]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>medium gen1</td>
<td>282.89</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>[Reservations]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>medium gen2</td>
<td>282.89</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>[Add Hours]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>large gen1</td>
<td>179.53</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>[Purchase History]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>large gen2</td>
<td>149.61</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>[Deactivate Account]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>extra large</td>
<td>127.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gpus-equipped gen2</td>
<td>200.70</td>
<td>0.00</td>
<td>0.00</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 11| chsoult02_jarekapp | jarekapp  | hours  | purchase 1    | 1157.00        | 0.00       | 0.00          | 0           | yes    | 11/29/2011 5:34:55 PM   |             | [Edit Credit Account]     |
|   |                    |            |        | general       | 3536.37        | 0.00       | 0.00          | 0           |        |                        |             | [Credit Account Users]     |
|   |                    |            |        | medium gen1   | 2187.68        | 0.00       | 0.00          | 0           |        |                        |             | [Reservations]             |
|   |                    |            |        | medium gen2   | 2187.68        | 0.00       | 0.00          | 0           |        |                        |             | [Add Hours]                |
|   |                    |            |        | large gen1    | 1388.40        | 0.00       | 0.00          | 0           |        |                        |             | [Purchase History]         |
|   |                    |            |        | large gen2    | 1157.00        | 0.00       | 0.00          | 0           |        |                        |             | [Deactivate Account]       |
|   |                    |            |        | extra large   | 985.59         | 0.00       | 0.00          | 0           |        |                        |             |                           |
|   |                    |            |        | gpus-equipped gen2 | 1552.12 | 0.00       | 0.00          | 0           |        |                        |             |                           |
|   |                    |            |        | purchase      | 60.00          | 0.00       | 0.00          | 0           |        |                        |             |                           |
|   |                    |            |        | general       | 183.98         | 0.00       | 0.00          | 0           |        |                        |             |                           |
Conversion factors between hours used on different machine types.

<table>
<thead>
<tr>
<th></th>
<th>general</th>
<th>medium gen1</th>
<th>medium gen2</th>
<th>large gen1</th>
<th>large gen2</th>
<th>extra large</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</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>
Choose type of account

<table>
<thead>
<tr>
<th>Credit Account</th>
<th>Type</th>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD</td>
<td>Hours</td>
<td>Active</td>
<td></td>
</tr>
<tr>
<td></td>
<td>General membership</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Lab Credit Accounts

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 an 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>
**Manage Users**

**Purchase Hours**

**Transfer Hours between Accounts**
**BioHPC Lab: Credit Account Users Management**

Credit Account 'jarekpp_general'

**Filter by:**
- Lab ID: _____________________________
- Last name: ___________________________
- Cornell ID: __________________________
- Created >= MM/DD/YYYY and <= MM/DD/YYYY
- Email: ______________________________

**Order by:** Created, Descending

<table>
<thead>
<tr>
<th>First Name</th>
<th>Last Name</th>
<th>Lab ID</th>
<th>Cornell ID</th>
<th>Email</th>
<th>Institution</th>
<th>Department</th>
<th>Created</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jarek</td>
<td>Pillardy</td>
<td>jarek</td>
<td>jarekpp</td>
<td><a href="mailto:jarekpp@yahoo.com">jarekpp@yahoo.com</a></td>
<td>Cornell University</td>
<td>CBSU</td>
<td>6/22/2011 1:32:12 PM</td>
<td></td>
</tr>
<tr>
<td>Qi</td>
<td>Sun</td>
<td>qi</td>
<td>qi24</td>
<td><a href="mailto:qi24@cornell.edu">qi24@cornell.edu</a></td>
<td>Cornell University</td>
<td>CBSU</td>
<td>5/28/2011 3:08:00 PM</td>
<td>Delete</td>
</tr>
<tr>
<td>Robert</td>
<td>Bukowski</td>
<td>Robert</td>
<td>rob299</td>
<td><a href="mailto:rob299@cornell.edu">rob299@cornell.edu</a></td>
<td>Cornell University</td>
<td>Biotec</td>
<td>11/10/2010 5:14:47 PM</td>
<td>Delete</td>
</tr>
<tr>
<td>Jarek</td>
<td>Pillardy</td>
<td>jarek</td>
<td>jjp86</td>
<td><a href="mailto:jjp86@cornell.edu">jjp86@cornell.edu</a></td>
<td>Cornell University</td>
<td>Biotec</td>
<td>11/18/2010 1:11:33 PM</td>
<td>Delete</td>
</tr>
</tbody>
</table>

4 Records found. Show 1-4

Add users, can be comma-separated list of Lab id’s

Delete user
BIOHPC LAB:
TRANSFER HOURS BETWEEN LAB CREDIT ACCOUNTS

FROM: jarekpp_general_ 149.61 hrs left

TO: jarekpp_general_ 149.61 hrs left

100 hours

Transfer
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!
If you have an account you can login here
Go here to make a reservation

Go here to manage existing reservations
choose type of machines

choose credit account

choose machine

click on “available” link to start in time slot you would like to use
<table>
<thead>
<tr>
<th>Reservation Type</th>
<th>Cores</th>
<th>RAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>medium gen1</td>
<td>24</td>
<td>128GB</td>
</tr>
<tr>
<td>medium gen2</td>
<td>40</td>
<td>256GB</td>
</tr>
<tr>
<td>large memory gen1</td>
<td>64</td>
<td>512GB</td>
</tr>
<tr>
<td>large memory gen2</td>
<td>96-112</td>
<td>512GB</td>
</tr>
<tr>
<td>extra large memory</td>
<td>96-112</td>
<td>1024GB</td>
</tr>
<tr>
<td>gou gen2</td>
<td>32</td>
<td>256GB</td>
</tr>
<tr>
<td></td>
<td>2x NVidia P100</td>
<td></td>
</tr>
</tbody>
</table>

**Database Computing:**
- **restricted**

**My Reservations**

<table>
<thead>
<tr>
<th>Date</th>
<th>Status</th>
<th>Reserved By</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Aug 21</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Mon Aug 22</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Tue Aug 23</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Wed Aug 24</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Thu Aug 25</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Fri Aug 26</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Sat Aug 27</td>
<td>available</td>
<td></td>
</tr>
<tr>
<td>Sun Aug 28</td>
<td>available</td>
<td></td>
</tr>
</tbody>
</table>

**Reserved Machines:**
- **cherry2:**
  - Linux 7.4
  - Red Hat RISCO SUN Twin Star 4 Nodes 1400 Series
  - 24 cores, 126GB RAM, 1TB HDD, 1TB SSD
  - LVM supported

- **cherry3:**
  - Linux (CentOS 7.4)
  - Red Hat RISCO SUN Twin Star 4 Nodes 1400 Series
  - 24 cores, 126GB RAM, 4TB HHD, 1TB SSD
  - LVM supported

- **cherry4:**
  - Linux (CentOS 7.4)
  - Red Hat RISCO SUN Twin Star 4 Nodes 1400 Series
  - 24 cores, 126GB RAM, 1TB HHD, 1TB SSD
  - LVM supported

- **cherry5:**
  - Linux (CentOS 7.4)
  - Red Hat RISCO SUN Twin Star 4 Nodes 1400 Series
  - 24 cores, 126GB RAM, 4TB HHD, 1TB SSD
  - LVM supported
<table>
<thead>
<tr>
<th>My reserved slots are marked in blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available slots are marked in green</td>
</tr>
<tr>
<td>Reserved slots are marked in red</td>
</tr>
</tbody>
</table>
**BIOHPC LAB: MY RESERVATIONS**

**Change Reservation**

Proposed reservation is shown in black. You have to click "Reserve!" button in order to finalize your reservation. If you click "Cancel" button then your reservation will be saved as history.

**Manage all my active reservations**

**Display reservations from February 17, 2014.**

**Modify reservation #29076:**

- From: February 20, 2014
- To: February 27, 2014
- 12:00am
- for
- cbuim09 in remote workstations (large memory) with Lab credit account

**You can delete your reservations by clicking on [X], you can modify your reservations by clicking on [X]. You can also add a new reservation by clicking on the appropriate "AVAILABLE" text in the box above.**

**Current (edited) reservation in black**

**Adjust dates and times**

**Preview new dates/times**
change dates/times or account

cancel reservation
**BIOHPC LAB: MY RESERVATIONS**

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>2078</td>
<td>2/17/2014 8:30:00 AM</td>
<td>2/20/2014 3:00:00 PM</td>
<td>dell_PowerEdge_M600</td>
<td>Linux</td>
<td>Vmware16GB Ram, 1TB HDD</td>
<td>user1</td>
<td></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>Owner</th>
<th>Other users</th>
<th>Credit account</th>
<th>Action</th>
<th>VNC port #</th>
</tr>
</thead>
<tbody>
<tr>
<td>2078</td>
<td>2/17/2014 8:30:00 AM</td>
<td>2/20/2014 3:00:00 PM</td>
<td>dell_PowerEdge_M600</td>
<td>Linux</td>
<td>Vmware16GB Ram, 1TB HDD</td>
<td>user1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

You can connect to your Linux reserved workstations using VNC protocol at 1280x900 from this page. For more on VNC please read "Access with VNC" in the Lab's User Guide.

Add user with [email] to my reservation #.

New reservation from February 17, 2014 3:00pm to February 17, 2014 3:00pm for the first available computer in the Lab.

**reservations I own**

**other users’ reservations I am allowed to use**

**allow other users to use my reserved machine**
Reservations

You can only make reservation if you have enough hours. Your hours will be “reserved” when you make the 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 – 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.
Reservations

What if I am running a program and need to end the reservation immediately when the program ends?

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

```bash
#!/bin/bash
/home/myid/myprogram [options]
/programs/bin/labutils/endres.pl
```
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!*
Storage

• **Networked storage**
  Very large storage (700TB+313TB), 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*. 
Storage

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 (/).

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
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.

**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.
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.
<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>my home directory location</td>
<td>/home/jarekpp</td>
</tr>
<tr>
<td>current usage</td>
<td>24.6GB</td>
</tr>
<tr>
<td>current limit</td>
<td>automatic - user associated with an active Lab Credit Account</td>
</tr>
<tr>
<td>limit type</td>
<td></td>
</tr>
<tr>
<td>add more storage or change quota</td>
<td></td>
</tr>
</tbody>
</table>
## My Home Directory Location

<table>
<thead>
<tr>
<th>Current Usage</th>
<th>Current Limit</th>
<th>Limit Type</th>
<th>Add More Storage or Change Quota</th>
<th>Expiration (For Purchased Storage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current disk usage: 242.5GB</td>
<td>Storage system: Lustre</td>
<td>Number of files: 494</td>
<td>Number of directories: 92</td>
<td>Disk usage updated: 10/24/2016 7:01:01 PM</td>
</tr>
<tr>
<td>Current disk quota: 512.0GB</td>
<td>Quota type: purchased storage</td>
<td>Purchased storage expiration date: 10/25/2018 5:07:40 PM</td>
<td>Update home directory storage info (may take long time)</td>
<td>Add or modify home directory storage</td>
</tr>
</tbody>
</table>

You are not part of any storage groups.
Storage

Extra storage can be purchased for $97.35 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/qisun_c4 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>1-25-2017</td>
<td>BackupDefaultPool</td>
<td>1.00</td>
<td>0.2336</td>
</tr>
<tr>
<td>1-25-2017</td>
<td>BackupDefaultPool5</td>
<td>1.00</td>
<td>0.0116</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>cbsubm06</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>

- [Purchase Backup Credit](#)
- [Manage Backup](#)
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)

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

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.
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 estimate 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)

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

subtracted daily from the originally purchased storage amount, displayed on the website

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

\[ S_0 = 0 \text{ (first day free!)} \]
Local Storage Notes

Your local storage on a 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 will be deleted on the first night ...
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!*
Connecting to BioHPC Cloud machines

Text-based connection: ssh (Secure SHell)

GUI (graphical) connection: X-Windows or VNC
Logging in to a Linux machine

- On any Linux machine, you need
  - network name of the machine (e.g. cbsumm10.tc.cornell.edu)
  - an account, i.e., user ID and password
  - on your local computer: remote access software
    (typically: ssh client)

- Linux is a multiple-access system: multiple users may be logged in and operate on one machine at the same time
Logging in to a Linux machine

- Remotely from a PC (Windows) via ssh client
  
  - Install and configure remote access software (PuTTY).
  
  - Use PuTTY to open a terminal window on the reserved workstation using ssh protocol;
  
  - You may open several terminal windows, if needed.
Logging in to a Linux machine

- Remotely from other Linux machine or Mac via native ssh client

  - Launch the Mac’s or Linux terminal window. Type

    ```
    ssh labid@cbsuXXXX.tc.cornell.edu
    ```

    (replace the “cbsuXXXX” with the server that you just reserved, and “labid” with your own user ID). Enter the lab password when prompted.

  - You may open several terminal windows, if needed, and log in to the workstation from each of them.
Logging in to CBSU machines from outside of Cornell

Two ways to connect from outside:

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

- Log in to cbsulogin.tc.cornell.edu (or cbsulogin2.tc.cornell.edu):

  `ssh jarekp@cbsulogin.tc.cornell.edu`  
  (using PuTTy or other ssh client program)

Once logged in to cbsulogin, ssh further to your reserved machine

  `ssh jarekp@cbsuwrkst3.tc.cornell.edu`

Backup login machine is cbsulogin2.tc.cornell.edu

https://cbsu.tc.cornell.edu/lab/doc/BioHPCLabexternal.pdf
login as: jarekp
jarekp@cbsulogin's password:
******************************************************************************
* 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 *
******************************************************************************
[jarekp@cbsulogin ~]$
Terminal window

- User communicates with the machine via **commands** typed in the terminal window
  
  - Commands are interpreted by a program referred to as **shell** – an interface between Linux and the user. Usually shell called **bash** is used (another popular shell is **tcsh**).

  - Typically, each command is typed in one line and “entered” by hitting the **Enter** key on the keyboard.

  - Commands deal with **files** and **processes**, e.g.,
    - request information (e.g., list user’s files)
    - launch a simple task (e.g., rename a file)
    - start an application (e.g., Firefox web browser, BWA aligner, IGV viewer, …)
    - stop an application
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://cbsu.tc.cornell.edu/ww/machines.aspx?i=99

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

• Linux or Mac: Open terminal window and type ssh command “ssh labid@workstation.tc.cornell.edu”. Provide your user name and password when prompted.
Connecting to BioHPC Cloud

GUI (graphical) connection: VNC
Logging in to a Linux workstation via VNC client (GUI)

In web browser, navigate to [http://cbsu.tc.cornell.edu/](http://cbsu.tc.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
- Select resolution you want
Logging in to a Linux workstation

(GUI)

---

BIOHPC LAB: VNC CONNECTION

Your VNC session at chbsdesktop01.tc.cornell.edu has been successfully initialized!

Preferable method of connecting to your you machine with VNC is to use an external VNC client.
Windows users can use Real VNC VNC Freeware, which you can download here (choose free option).
Mac users can use Chicken of the VNC. Real VNC Viewer is also available for MAC (here).

To connect, launch your VNC client, type in machine name (chbsdesktop01) and port number (5901) and connect!
Typically machine name and port number are used together: chbsdesktop01.tc.cornell.edu:5901.

Due to security restrictions Java VNC viewer may not work in your browser. In most cases, you will have to lower security settings to "Minimum" in order to start the applet (see this link for details). You can access the applet using this link http://chbsdesktop01.tc.cornell.edu:5901.

Return to My Reservations.
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**
Exercise: connect to your assigned workstation using VNC

• Go to “My Reservations” page
  http://cbsu.tc.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: VNC

VNC sessions are *persistent*.

They run even when the client is disconnected.

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

Equivalent to Windows Remote Desktop.
Connecting to BioHPC Cloud

GUI (graphical) connection: X-Windows

Not persistent – programs will get killed when client disconnects.
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:
  \texttt{ssh -X labid@workstation.tc.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.

• \textbf{NOTE}: Mac machines no longer have X11 as a part of their system, it must be installed by user.
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.
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!
Transferring data

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

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

• wget
  Transfer can be done to Lab 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:


(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):

  `wget -O e_coli_1000_1.fq "http://cbsuapps.tc.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
  
  ```
  ```
File transfer: sftp
between PC or Mac and a Lab workstation

On Windows PC: install and use your favorite sftp client program, such as
- **winscp**: [http://winscp.net/eng/index.php](http://winscp.net/eng/index.php)
- **CoreFTP LE**: [http://www.coreftp.com/](http://www.coreftp.com/)
- **FileZilla** (client): [http://filezilla-project.org/](http://filezilla-project.org/)
- ... others...
- When connecting to Lab workstations from a client, use the sftp protocol. 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](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 ...
File transfer: scp
between a lab workstation and another Linux machine

Suppose we want to transfer a file /home/jarekp/test.fa from cbsulogin.tc.cornell.edu (another Linux machine; substitute “your” Linux machine here) and cbsum1c2b007 lab workstation /workdir/jarekp directory.

scp is a Unix command to copy files between servers

#copying file1 to dir2 on server2
scp user1@server1:/dir1/file1 user2@server2:/dir2

#copying dir2 to dir3 on server2
scp -r user1@server1:/dir1/dir2 user2@server2:/dir3

If you are logged in on cbsulogin

scp /home/jarekp/test.fa cbsum1c2b007:/workdir

If you are logged in on cbsum1c2b007

scp cbsulogin:/home/jarekp/test.fa /workdir

Option: Filezilla is installed on Lab workstations. Connect to Lab workstations with GUI support (VNC or X-Windows), type filezilla and connect to the other servers using sftp protocol.
Exercise: download BLOSUM100 file from Lab machine to your local computer

- Connect to your workstation using sftp program (FileZilla)
- Download BLOSUM100
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 and biohpc#cbsulogin2

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

http://cbsu.tc.cornell.edu/lab/doc/Globus_at_BioHPC_Lab.pdf
https://www.globusonline.org/quickstart/
Transferring data: Globus

- Sign up for Globus online account if you don’t have one (use the link above).
- Sign in to Globus online using your id and password.
  
  Go to File Transfer, in the left panel “Endpoint” field type `biohpc#cbsulogin` and click go. Login window will pop up, you need to use your BioHPC Cloud user id and password to connect to the endpoint, click Authenticate.

- When authentication is successful files from your home directory will show up in the left panel.

- If you would like to connect to another server endpoint you can just type it in the right side “Endpoint” field and connect.

- If you want to transfer to the local laptop (which is not a server) you need to install Globus Connect (if it is already installed, just start it).
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

Sign in with Google

Sign in with ORCID ID
Transferring data: Globus

• Sign up for Globus online account if you don’t have one (use the link above).
• Sign in to Globus online using your id and password
  • Go to File Transfer, in the left panel “Endpoint” field type `biohpc#cbsulogin` and click go. Login window will pop up, you need to use your BioHPC Cloud user id and password to connect to the endpoint, click Authenticate
  • When authentication is successful files from your home directory will show up in the left panel.
• If you would like to connect to another server endpoint you can just type it in the right side “Endpoint” field and connect.
• If you want to transfer to the local laptop (which is not a server) you need to install Globus Connect (if it is already installed, just start it).
BioHPC Cloud endpoint name
cbsulogin is our server

enter you Lab id and password

BioHPC Cloud endpoint name
Transferring data: Globus

- Sign up for Globus online account if you don’t have one (use the link above).
- Sign in to Globus online using your id and password
- Go to File Transfer, in the left panel “Endpoint” field type `biohpc#cbsulogin` and click go. Login window will pop up, you need to use your BioHPC Cloud user id and password to connect to the endpoint, click Authenticate

  - When authentication is successful files from your home directory will show up in the left panel.

  - If you would like to connect to another **server** endpoint you can just type it in the right side “Endpoint” field and connect.

  - If you want to transfer to the local laptop (which is not a server) you need to install Globus Connect (if it is already installed, just start it).
Transfer Files

Endpoint: biocore:Fasfolk
Path: ~/.

Start by selecting an endpoint.
Transferring data: Globus

• Sign up for Globus online account if you don’t have one (use the link above).
• Sign in to Globus online using your id and password
• Go to File Transfer, in the left panel “Endpoint” field type biohpc#cbsulogin and click go. Login window will pop up, you need to use your BioHPC Cloud user id and password to connect to the endpoint, click Authenticate
• When authentication is successful files from your home directory will show up in the left panel.
• If you would like to connect to another server endpoint you can just type it in the right side “Endpoint” field and connect.
• If you want to transfer to the local laptop (which is not a server) you need to install Globus Connect (if it is already installed, just start it).
Your local computer endpoint name. Usually account#computer
Globus Connect must be running and local endpoint registered

See this pdf for details
http://cbsu.tc.cornell.edu/lab/doc/Globus_at_BioHPC_Lab.pdf
Exercise: transfer BLOSUM100 file to your laptop using Globus
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 chologin or chologin2 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>cbi physic1</td>
<td>feg23</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 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.
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!
Software

https://cbsu.tc.cornell.edu/lab/userguide.aspx?a=software
https://cbsu.tc.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.
<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>454 gsAssembler or gsMapper</td>
<td>2.8</td>
<td>Linux</td>
<td>Assembly and alignment software for 454 or other long reads.</td>
<td>12/13/2011</td>
<td>3/24/2014</td>
<td>[details]</td>
</tr>
<tr>
<td>a3</td>
<td>20130522</td>
<td>Linux</td>
<td>A genome assembly pipeline for bacteria and archaea</td>
<td>10/14/2013</td>
<td>6/3/2016</td>
<td>[details]</td>
</tr>
<tr>
<td>Albireo</td>
<td>20161221</td>
<td>Linux</td>
<td>Albireo is a de novo assembler for PacBio and Oxford Nanopore Technologies reads.</td>
<td>12/21/2016</td>
<td></td>
<td>[details]</td>
</tr>
<tr>
<td>ABYSS</td>
<td>1.9.0</td>
<td>Linux</td>
<td>Illumina short reads assembly tool.</td>
<td>12/13/2011</td>
<td>5/9/2016</td>
<td>[details]</td>
</tr>
<tr>
<td>AdaptersRemoval</td>
<td>2.1.1</td>
<td>Linux</td>
<td>Remove adapters from sequences in either single end or paired end experiments</td>
<td>9/16/2015</td>
<td></td>
<td>[details]</td>
</tr>
<tr>
<td>Admixture</td>
<td>1.1</td>
<td>Linux</td>
<td>The ADMIXTOOLS package implements 5 methods described in Patterson et al (2012)Ancient Admixture in Human History.</td>
<td>11/6/2013</td>
<td>1/22/2015</td>
<td>[details]</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>2/19/2014</td>
<td>2/19/2014</td>
<td>[details]</td>
</tr>
<tr>
<td>abalone</td>
<td></td>
<td>Linux</td>
<td>Nanopore base caller.</td>
<td>6/2/2017</td>
<td></td>
<td>[details]</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>11/8/2012</td>
<td>11/8/2012</td>
<td>[details]</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>4/2/2014</td>
<td></td>
<td>[details]</td>
</tr>
<tr>
<td>ALLMAPS</td>
<td>20150719</td>
<td>Linux</td>
<td>ALLMAPS is capable of computing a scaffold ordering that minimizes the colinearity to a collection of maps, including genetic, physical or comparative maps into the final chromosome build.</td>
<td>7/10/2015</td>
<td></td>
<td>[details]</td>
</tr>
<tr>
<td>ALLPATHS-LG</td>
<td>49144</td>
<td>Linux</td>
<td>Illumina short reads assembly tool.</td>
<td>12/14/2011</td>
<td>12/3/2014</td>
<td>[details]</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>1/12/2015</td>
<td>1/14/2015</td>
<td>[details]</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>7/28/2017</td>
<td>7/27/2017</td>
<td>[details]</td>
</tr>
<tr>
<td>Analysis</td>
<td>0.8.4</td>
<td>Linux</td>
<td>C++ software for evolutionary genetic analysis.</td>
<td>9/2/2014</td>
<td></td>
<td>[details]</td>
</tr>
<tr>
<td>ANGEL</td>
<td>0.0.31</td>
<td>Linux</td>
<td>ANGEL is a software for analyzing metagenomic sequencing data. The software analyzes a number of different lang tools from paired reads to inferred gene orthologs.</td>
<td>7/23/2014</td>
<td>1/14/2014</td>
<td>[details]</td>
</tr>
</tbody>
</table>
BioHPC Lab Software

There are 302 software titles installed in BioHPC Lab. 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](https://biocl.ncsu.edu/labs/userguide/labs/software).

Please read details and instructions before running any program, it may contain important information on how to properly use the software in BioHPC Lab.

454 gsAssembler, grgMapper, a5, ABrujin, ABySS, AdapterRemoval, Admidtools, Admixtre, albacore, Alder, AlleleSeq, ALLMAPS, ALLPATHS-LG, AMOS, AMPHORA, anaylzes, ANGSD, Apropos, apollo, Atlas-Link, ATLAS_GapFill, ATLAS, Augustus, bamtools, Basset, BayeScan, BBSmap, BCTools, bcfetch, Beagle, Beagle4, Bead1, bead2, beadstamps, BEDtools, bc, bgc, biobambam, Bioconductor, BioPerl, BioPython, Birdsuite, Bismark, blast, BLAST, blast2dgo, BLAT, btmaps, Boost, Bowman, Bowtie2, breseq, BSSeeker2, BUSCO, BWA, canu, CAPS, CBSU RNAseq, cd-hit, CEIGMA, CellRanger, Circos, Circoscape, CLUMPY, Clustal Omega, CLUSTALW, Cluster, clustalw, CNAster, cortex_v, CrossMap, CRT, cuda, Cufflinks, cutadapt, dad, dar1, 1.8.1, MFEI, dDecom, DESENTE, diamond, Discovar, Dr2loop, duality, druk, drukmap, Dust, DropGC, Dropseq, DropseqPipe, dsk, ea-utils, ecopcr, EDGE, EIGENSOFT, EMBoss, entropy, emrino, exabases, exonerate, eXpress, Falcon, Falcon_unzip, FastGBS, fasta, fastclaster, FastML, fast_species_detector, FastQC, fastStructure, FastTree, FASTX, finestructure, flash, Flexible Adapter Remover, FMAP, freebayes, FunGene Pipeline, GATK, GBSR, GCA, GEM library, GEMMA, gened, GeneMark, GeneMark, Genome STRIP, GenomeMapper, GenomeStudio (illumina), GenomicConsensus, gern, germine, GMAP/GSNAK, GNU Compilers, GNU parallel, Grinder, GROMACS, Gubbins, HapCompass, HAPCUT, HAPCUT2, hapflk, HaplotypeMerger, Haplomerger2, HapSeq2, HIC-Pro, HISAT2, HMMER, Homer, HOTSPOT, HTSeq, HuMan2, HuPy, iAssiyem, IBDLD, IBDX-UD, iGV, lma2, lma2p, IMAGE, impute2, impute_v2, Impute2, Impute3, IntervalMap, InterProScan, iRep, java, javadoc, jellyfish, JoinMap, jula, jupiter, kallisto, Kent source utilities, kmer, LACHESIS, kMLink, LDAK, iem, LINS, Links, LocustSearch, longranger, LUCY, LUCY2, LUMPY, MACS, MACS simulator, MACS2, MAFFT, MAGIC-BLAST, MAKER, MAQ, MASH, MarSAQR, Mauve, maq, megaphl, MEME Suite, MERLIN, MetaBAT, metaCRISPR, MetaMOS, MetaPathways, MetaPhAn, MetaVelvet, MetaVelvet.SL, MigraTe, mma, mmRDeep2, MISO, mqc, MrBayes, MrFast, MrISD, MSMC, MSMS-CA Genome assembler, msdsats, MSTmap, msymy, MultiQc, MUMmer, muscle, mutech, nctp, Nemo, NetBeans, NEURON, nextGenMap, NSAbcDave, ngsDist, ngsF, ngsTools, NGSUtils, Novoalign, NovoalignCS, Oases, OBITools, Orthomcl, PATAlign, PAML, pandas, pandaseq, PanPase, PASTEL, phsalign, pbtools, PBnJelly, PBSuite, PeakSplitter, PEAR, PennCNV, philtools, PhageFinder, PHAST, PHYLIP, PhyloCSF, phyloPython, PhyML, Picard, Pindel, philPipes, PIQ, Platypus, plink, Plot, popbpth, prism, prodigal, progressiveCactus, prokka, pyRAD, PySNPTools, PyVCF, QHMA, QUIMO, QIME2, QIIME, Quale, QuantumSNP2, QUAST, QUMA, R, RACA, RADAR, RAPTR-SV, RawML, Ray, Rconformer, REAPR, RepeatMasker, RepeatModeler, RF-Mix, RNAMMER, maQAST, Roary, RSEM, RseqC, RStudio, rare, SangerGW, samblaster, Samtools, Satsuma, scikit-learn, scythe, Sentience, SeqPrep, sgraph, SHARP, shoe, SHOR, Enmap, shortBRED, SHRMP, sickle, SignalP, smuOP, snavel, smcmap, SMRT Analysis, snakemake, snap, SNAPP, SNAPP.py, SOAP, SOAPdenovo, SOAPdenovo-Trans, SOAPPenzo, SomaticSniper, SPAdes, SRA Toolkit, srtz, stacks, stampy, STAR, statmodtools, Strelka, StringTie, STRUCTURE, superov, SURPI, suita, SVDetect, svtools, SweepFinder, sweepsems, taffix, Tandem Repeats Finder, TRF, TASSEL 3, TASSEL 4, tool, Tophat, tradRNA, Trans-Proteomic Pipeline (TPP), TransComb, TransDecoder, transrate, TRAP, treeCl, treeMix, trimomatic, Trinity, Trimmomatic, TRRNAseq-SE, UCSC Kent utilities, UMItools, usearch, Variant Effect Predictor, VarScan, vcf2gff4, vcf2cooker, vcf2gff, vcf2tools, VcfTool, VEGPA, ViennaRNA, VIP, VirusFinder 2, ViZin, vsearch, WASH, wgs-assembler (Celera), WgsS2 (Genewise), Xander_assembler, yah
Software

Details for **bamtools** (hide)

- **Name:** bamtools
- **Version:** 2.3.0
- **OS:** Linux
- **About:** BAM file processing and filtering.
- **Added:** 12/13/2011 2:56:40 PM
- **Updated:** 3/24/2014 3:26:48 PM
- **Link:** [http://seqanswers.com/wiki/BamTools](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:

```bash
bamtools [options]
```

There are 1 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>
</table>
| 2.3.0 (latest) | full path: /programs/bamtools-2.3.0/bin/bamtools [options]  
add to PATH: export PATH=/programs/bamtools-2.3.0/bin:$PATH |
| 2.2.3 | full path: /programs/bamtools-2.2.3/bin/bamtools [options]  
add to PATH: export PATH=/programs/bamtools-2.2.3/bin:$PATH |
## Details for Trinity (trinity)

**Name:** Trinity

**Version:** 2.0.4

**OS:** Linux

**About:** De novo transcriptome assembly.

**Added:** 12/15/2011 3:22:56 PM

**Updated:** 3/9/2015 1:18:17 PM

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

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

/programs/trinityrnaseq-2.0.4/Trinity [options]

You can also add the program to your PATH

```
export PATH=/programs/trinityrnaseq-2.0.4:$PATH
```

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

It is recommended to run on medium memory nodes (e.g., chemn01). A large memory machines (e.g., checm05) may be needed to process larger datasets.

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

/programs/trinityrnaseq-2.0.4/Trinity [options] > & trinity.log &

If you need to run TransDecoder that is part of Trinity release, you will need to download the Plasm-A.hmm.gz file to your working directory, and add path to bin/ before you run the tool (You will need to run `export ...` command every time before you run TransDecoder).

wget ftp://ftp.sanger.ac.uk/pub/databases/PlasmA/current_release/Plasm-A.hmm.gz

export PATH=/programs/hummer/binaries:$PATH

If you prefer to run any of the earlier versions of the program (2013-10-05 or earlier), you will need to first modify the environment to use an earlier version of jar by executing the following commands:

```
export JAVA_HOME=/usr/lib/jvm/jce-1.0.0-openssl3.aff_04
export PATH=$JAVA_HOME/bin:$PATH
```

There are 8 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>
</table>
| 2.0.4   | full path: /programs/trinityrnaseq-2.0.4/Trinity [options]  
add to PATH: export PATH=/programs/trinityrnaseq-2.0.4:$PATH |
| 20140413p1 | full path: /programs/trinityrnaseq_r20140413p1/Trinity [options]  
add to PATH: export PATH=/programs/trinityrnaseq_r20140413p1:$PATH |
| 20131110 | full path: /programs/trinityrnaseq_r20131110/Trinity [options]  
add to PATH: export PATH=/programs/trinityrnaseq_r20131110:$PATH |
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.
Software

Want to install software yourself?

Many software titles can be installed in user space, especially Python or Perl.

You can download and use Docker image.

You can install software inside Docker container.

=> Topic for another workshop, next in Spring 2019
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!*