Introduction to BioHPC Lab

BioHPC Lab Workshop

Jaroslaw Pillardy

Bioinformatics Facility
Institute of Biotechnology
Cornell University

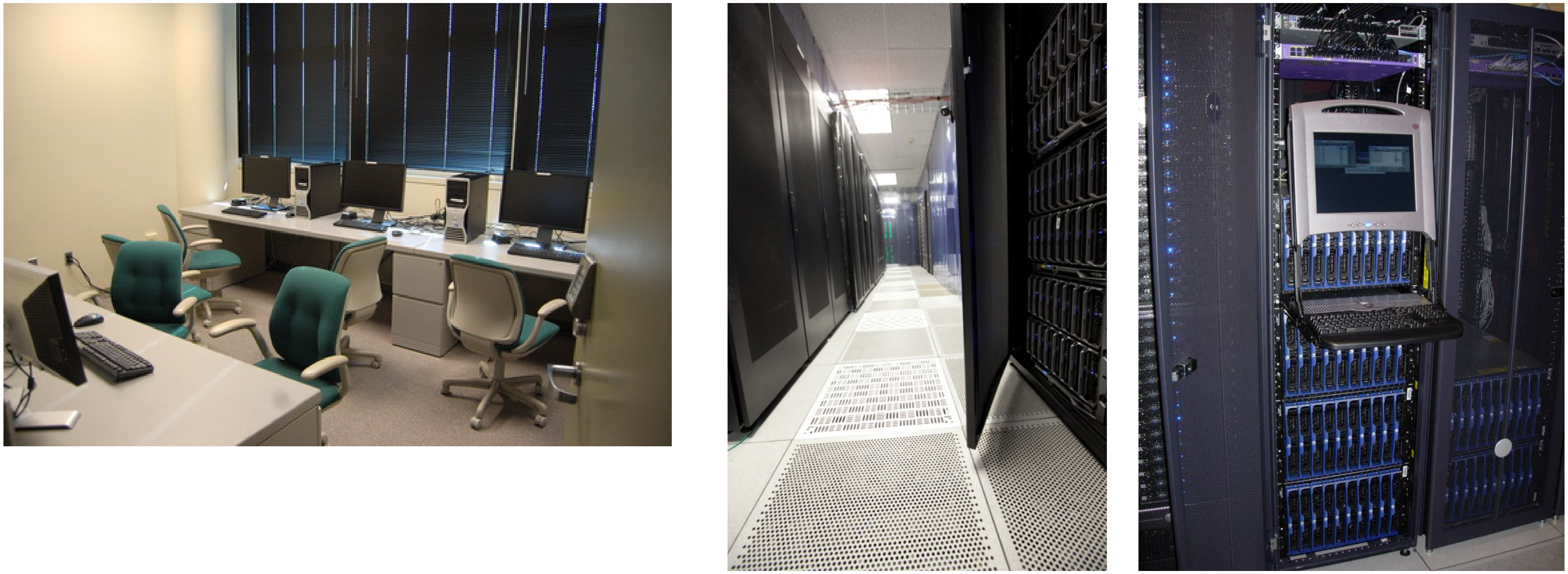
http://cbsu.tc.cornell.edu/lab/lab.aspx

http://cbsu.tc.cornell.edu/lab/doc/Introduction_to_BioHPC_Lab_v2.pdf

BioHPC Laboratory

BioHPC Laboratory 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 182 titles)
- Located in Rhodes Hall and accessible remotely and on-site



BioHPC Lab: computing

Two ways of using BioHPC Lab computing resources:

- Buy hours and run computations on BioHPC Lab 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 ask us to buy it for you) and host it in BioHPC Lab

  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 Lab: storage

- 512 TB of networked storage available in one volume
- Very robust and scalable architecture: cluster storage based on RAID6 servers connected by Gluster, new servers can be added to expand storage.
- Very affordable! $87.13 per TB per year is lower than even Amazon archival storage (Amazon Glacier: $120 per TB per year + data transfer fees)
- Users get free storage allocations with the accounts

BioHPC Lab: software

- BioHPC Lab 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.
- 182 titles as of 2/27/2015
- 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 Lab 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" remote workstations:

32	8-core, 16GB RAM, 1TB HDD (cbsum1c1b0NN, cbsum1c2b0NN)

"medium memory" remote workstations

1	16-core, 64GB RAM, 1TB HDD
16	12-core, 128GB RAM, 4TB HDD, 1TB SSD (cbsummNN)

"large memory" remote workstations

1	48-core, 512GB RAM, 12TB HDD
5	64-core, 512GB RAM, 12TB HDD
3	64-core, 512GB RAM, 9TB HDD, 1TB SSD (cbsulmNN)

networked storage: total 520TB available in one volume

2	40TB HDD server
5	88TB HDD server

login machine (cbsulogin, cbsulogin2)

2	12-core, 64GB RAM, 1.5TB HDD

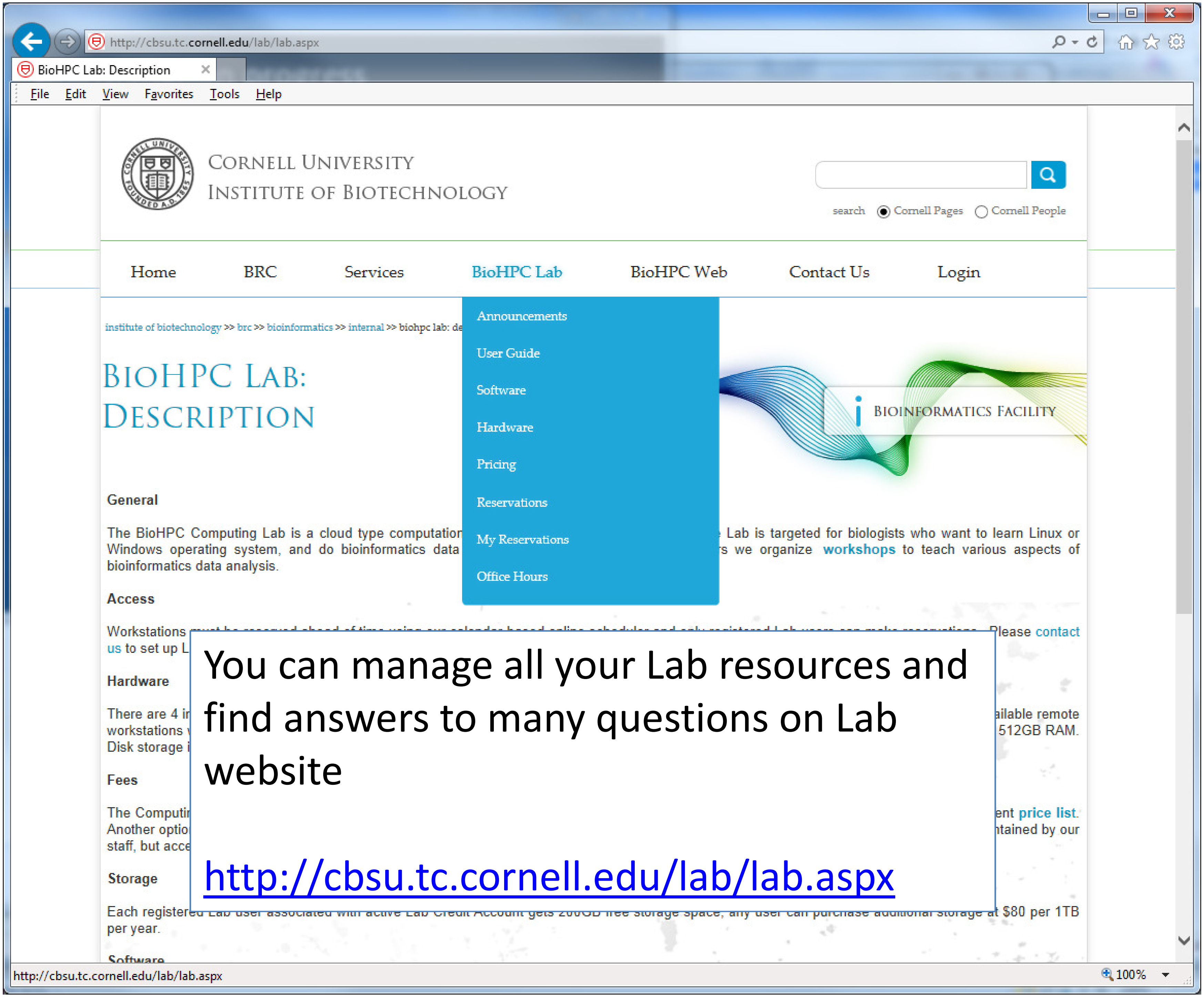
Setting up an account

- In order to get an account send an e-mail to cbsu@cornell.edu with your name, Cornell Netid and affiliation

- 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.



BIOHPC LAB: DESCRIPTION

General

The BioHPC Computing Lab is a cloud type computation … Lab is targeted for biologists who want to learn Linux or Windows operating system, and do bioinformatics data … s we organize workshops to teach various aspects of bioinformatics data analysis.

Access

Workstations … Please contact us to set up L…

Hardware

There are 4 i… ailable remote workstations … 512GB RAM. Disk storage i…

Fees

The Computi… ent price list. Another optio… ntained by our staff, but acce…

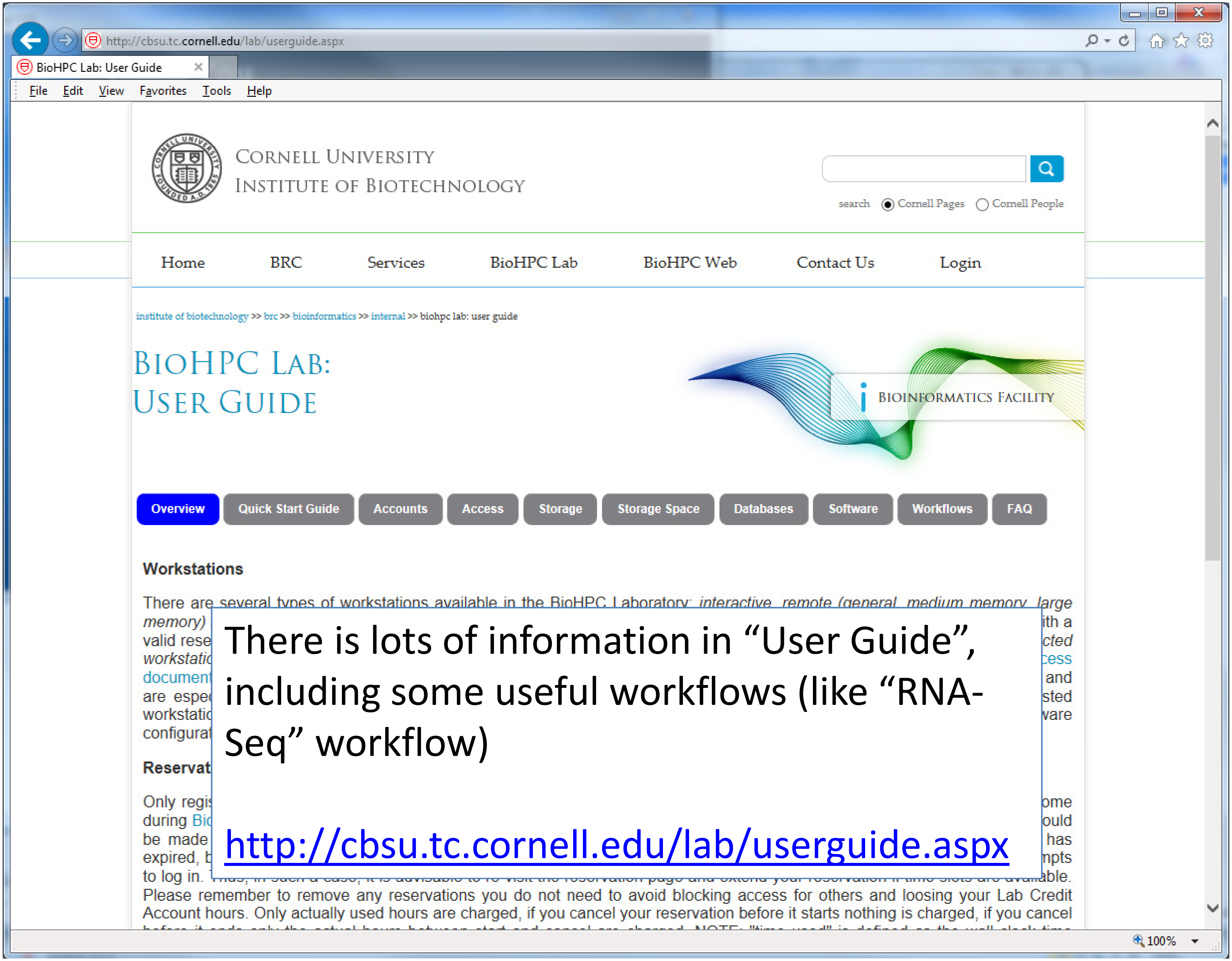
Storage

Each registered Lab user associated with active Lab Credit Account gets 200GB free storage space, any user can purchase additional storage at $80 per 1TB per year.

Software

You can manage all your Lab resources and find answers to many questions on Lab website

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



BioHPC Lab: User Guide

Workstations

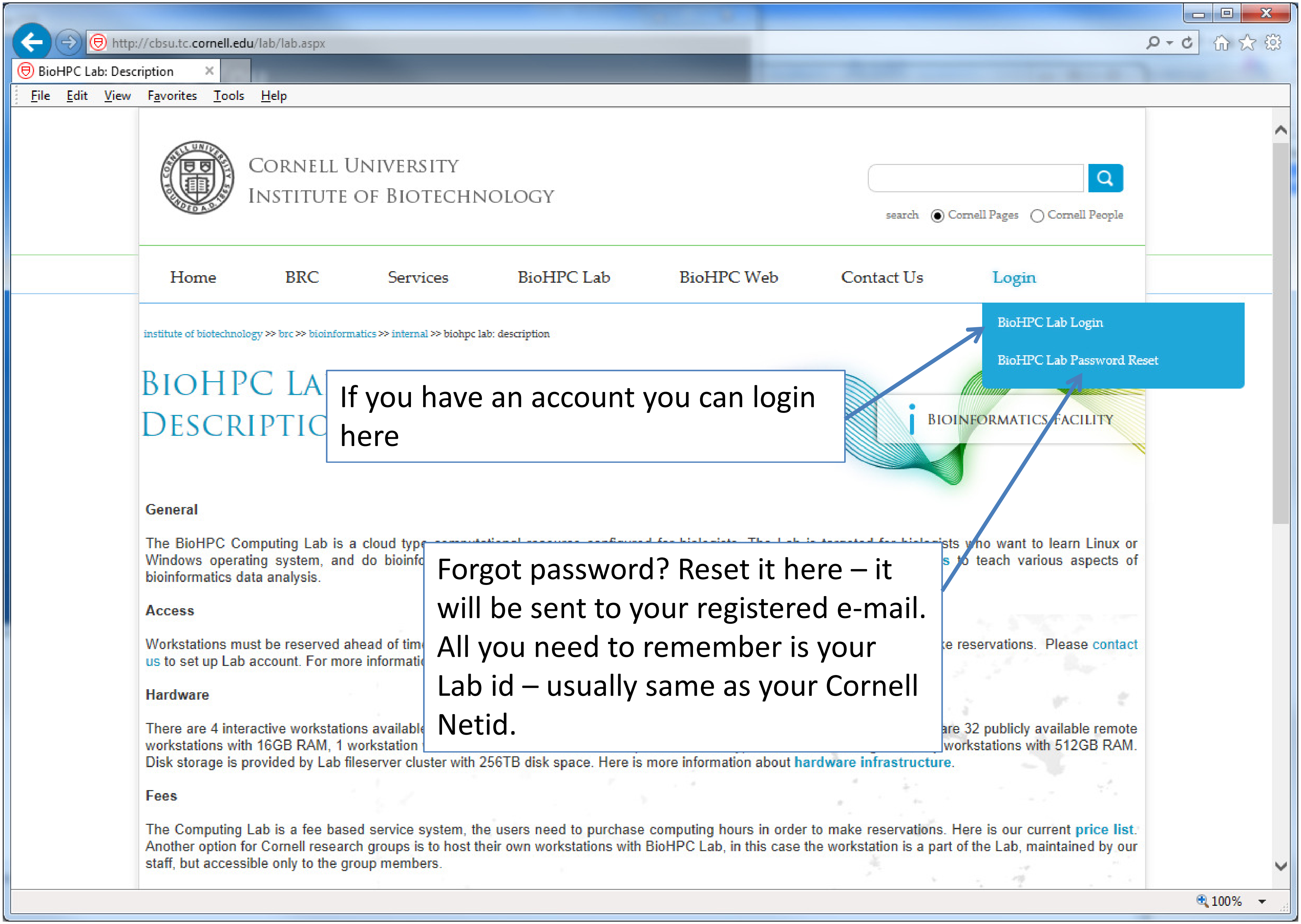
There are several types of workstations available in the BioHPC Laboratory: *interactive*, *remote (general, medium memory, large memory)*

Reservat

Please remember to remove any reservations you do not need to avoid blocking access for others and loosing your Lab Credit Account hours. Only actually used hours are charged, if you cancel your reservation before it starts nothing is charged, if you cancel

There is lots of information in "User Guide", including some useful workflows (like "RNA-Seq" workflow)

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



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: description

BIOHPC LA

DESCRIPTIO

If you have an account you can login here

General

The BioHPC Computing Lab is a cloud type … sts who want to learn Linux or Windows operating system, and do bioinfo… s to teach various aspects of bioinformatics data analysis.

Access

Workstations must be reserved ahead of tim… e reservations. Please contact us to set up Lab account. For more informati…

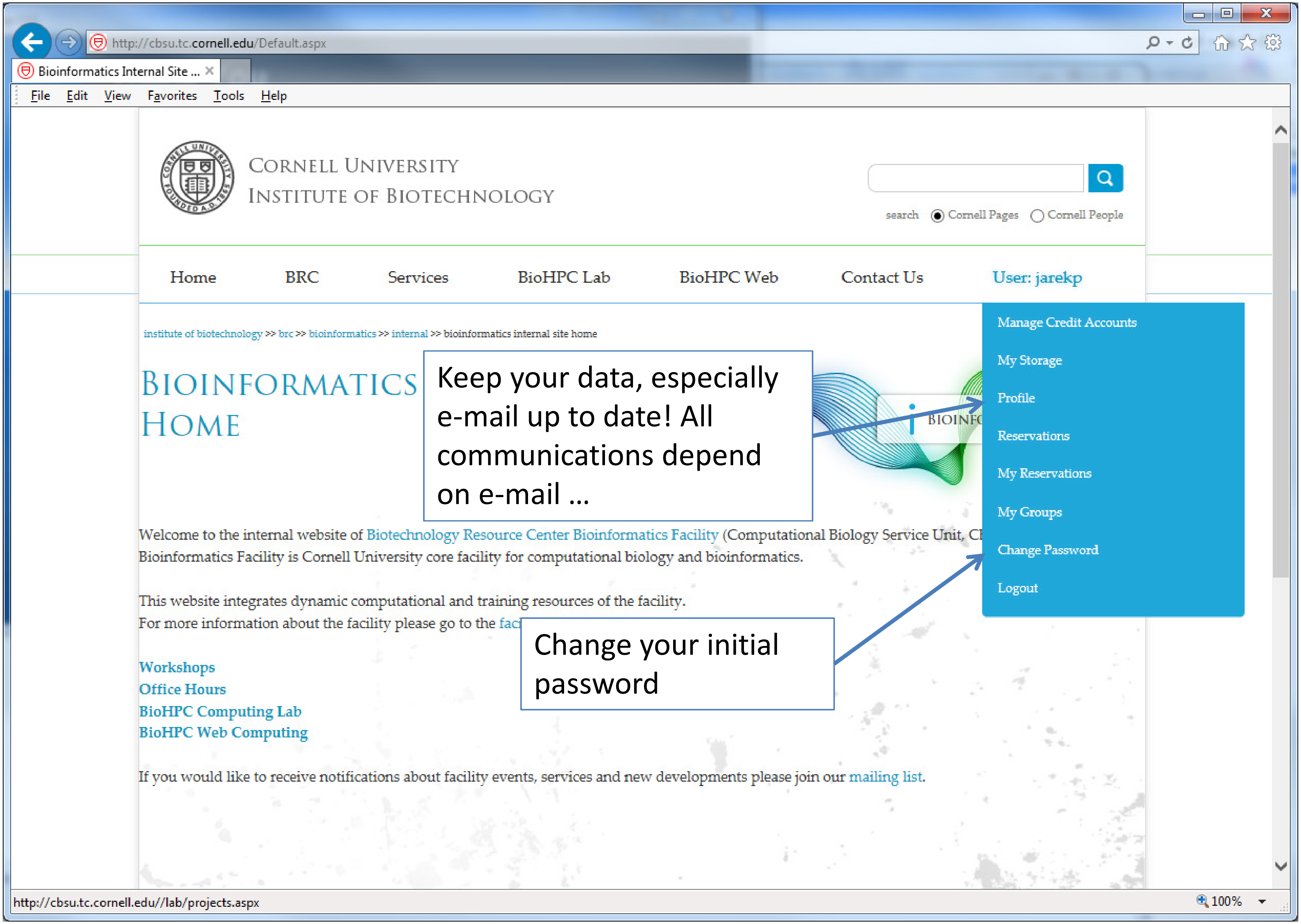
Hardware

There are 4 interactive workstations available … are 32 publicly available remote workstations with 16GB RAM, 1 workstation … workstations with 512GB RAM. Disk storage is provided by Lab fileserver cluster with 256TB disk space. Here is more information about hardware infrastructure.

Fees

The Computing Lab is a fee based service system, the users need to purchase computing hours in order to make reservations. Here is our current price list. Another option for Cornell research groups is to host their own workstations with BioHPC Lab, in this case the workstation is a part of the Lab, maintained by our staff, but accessible only to the group members.

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.



Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (Computational Biology Service Unit, CI
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 faci

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

Keep your data, especially e-mail up to date! All communications depend on e-mail …

Change your initial password

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!*

Lab Credit Accounts

In order to reserve workstations you need to have hours available in your Lab Credit Account.

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

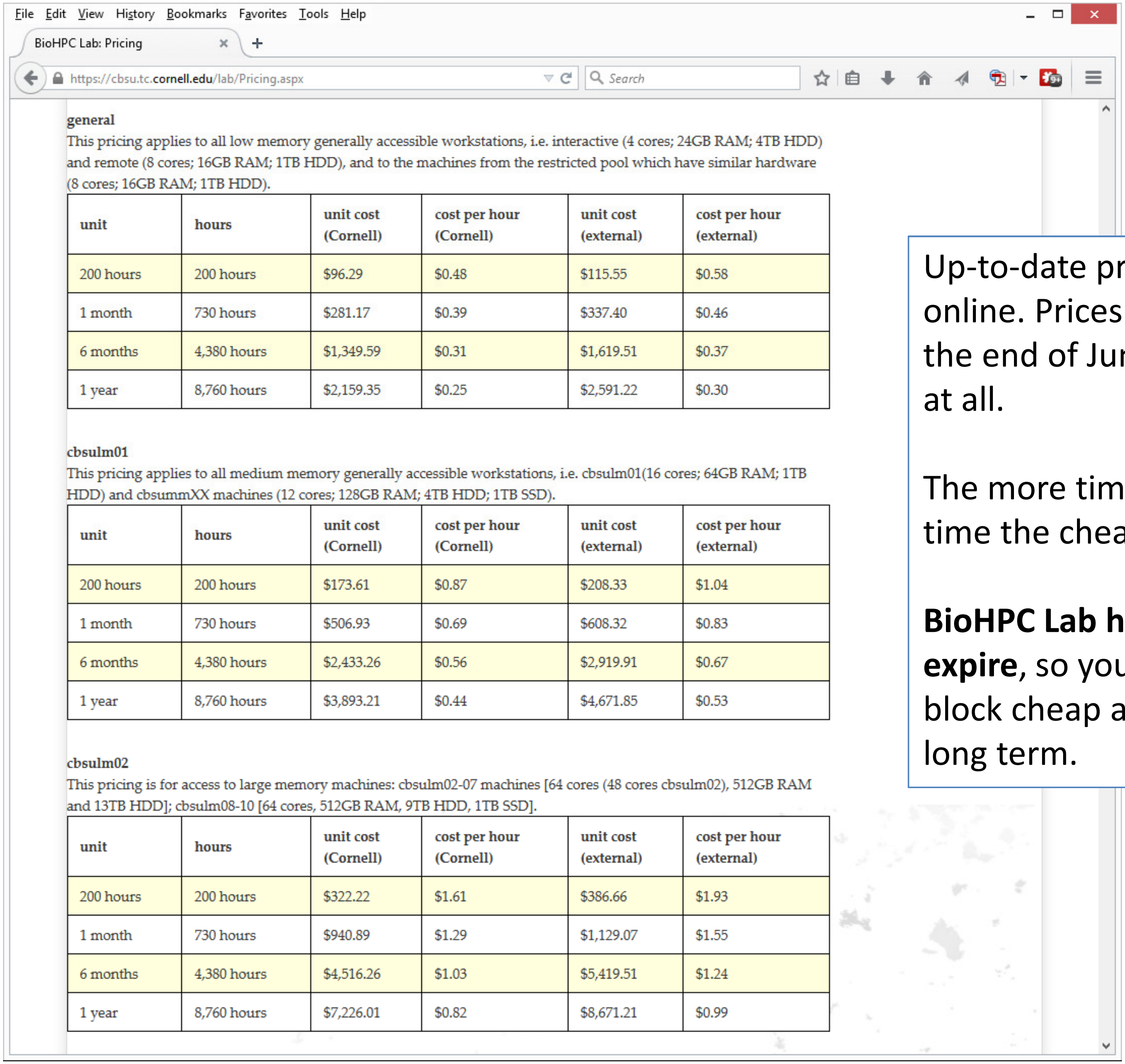
There are **3 types of workstations** linked to **3 types of hours**:

- **general**
  cbsum1c1*, cbsum1c2*, cbsuwrkst*

- medium memory (**cbsulm01**)
  cbsulm01, cbsumm*

- large memory (**cbsulm02**)
  cbsulm*

http://cbsu.tc.cornell.edu/Lab/Pricing.aspx



general

This pricing applies to all low memory generally accessible workstations, i.e. interactive (4 cores; 24GB RAM; 4TB HDD) and remote (8 cores; 16GB RAM; 1TB HDD), and to the machines from the restricted pool which have similar hardware (8 cores; 16GB RAM; 1TB HDD).

unit	hours	unit cost (Cornell)	cost per hour (Cornell)	unit cost (external)	cost per hour (external)
200 hours	200 hours	$96.29	$0.48	$115.55	$0.58
1 month	730 hours	$281.17	$0.39	$337.40	$0.46
6 months	4,380 hours	$1,349.59	$0.31	$1,619.51	$0.37
1 year	8,760 hours	$2,159.35	$0.25	$2,591.22	$0.30

cbsulm01

This pricing applies to all medium memory generally accessible workstations, i.e. cbsulm01(16 cores; 64GB RAM; 1TB HDD) and cbsummXX machines (12 cores; 128GB RAM; 4TB HDD; 1TB SSD).

unit	hours	unit cost (Cornell)	cost per hour (Cornell)	unit cost (external)	cost per hour (external)
200 hours	200 hours	$173.61	$0.87	$208.33	$1.04
1 month	730 hours	$506.93	$0.69	$608.32	$0.83
6 months	4,380 hours	$2,433.26	$0.56	$2,919.91	$0.67
1 year	8,760 hours	$3,893.21	$0.44	$4,671.85	$0.53

cbsulm02

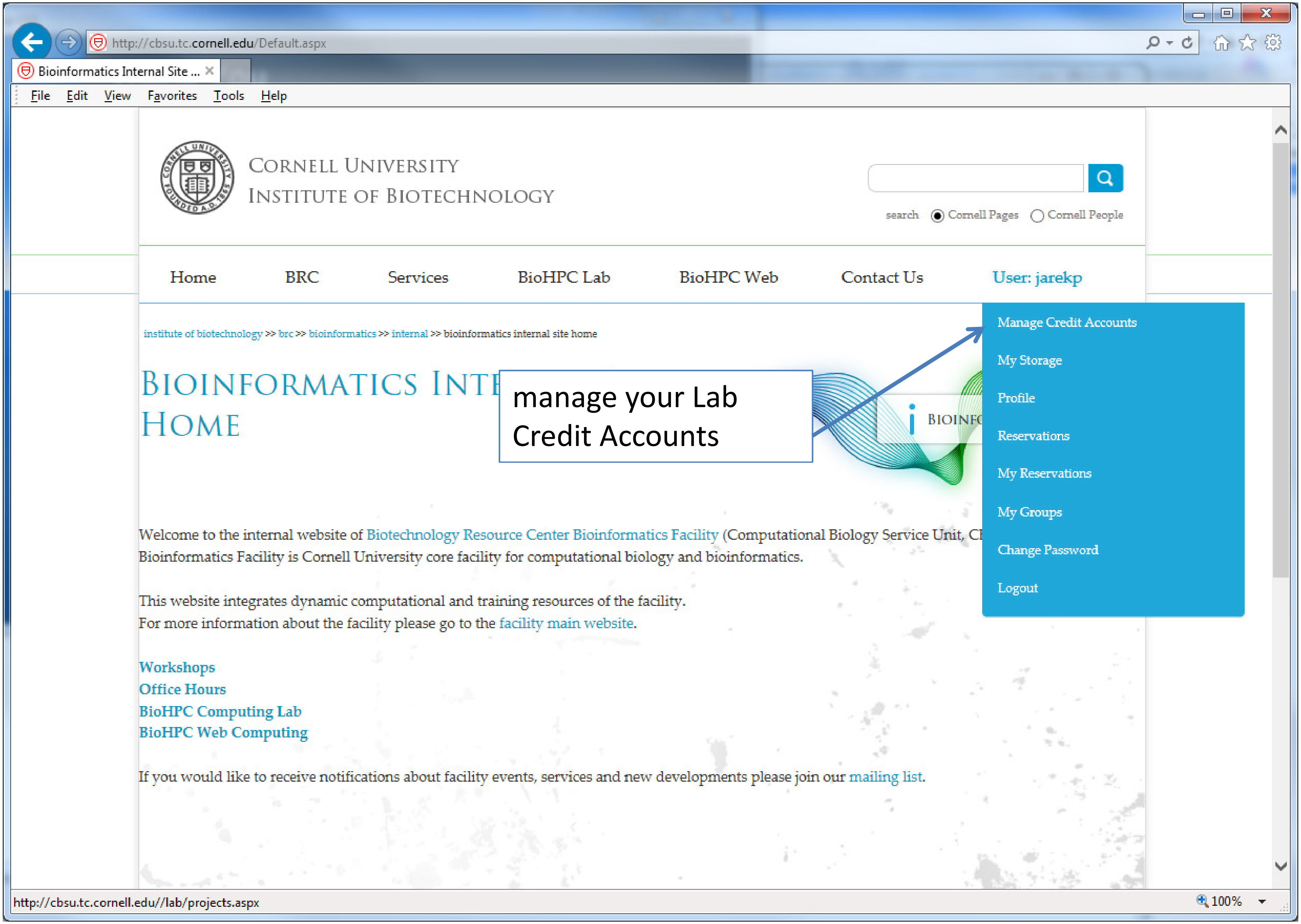
This pricing is for access to large memory machines: cbsulm02-07 machines [64 cores (48 cores cbsulm02), 512GB RAM and 13TB HDD]; cbsulm08-10 [64 cores, 512GB RAM, 9TB HDD, 1TB SSD].

unit	hours	unit cost (Cornell)	cost per hour (Cornell)	unit cost (external)	cost per hour (external)
200 hours	200 hours	$322.22	$1.61	$386.66	$1.93
1 month	730 hours	$940.89	$1.29	$1,129.07	$1.55
6 months	4,380 hours	$4,516.26	$1.03	$5,419.51	$1.24
1 year	8,760 hours	$7,226.01	$0.82	$8,671.21	$0.99

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 Lab hours NEVER expire, so you can buy a large block cheap and use them in a long term.



manage your Lab Credit Accounts

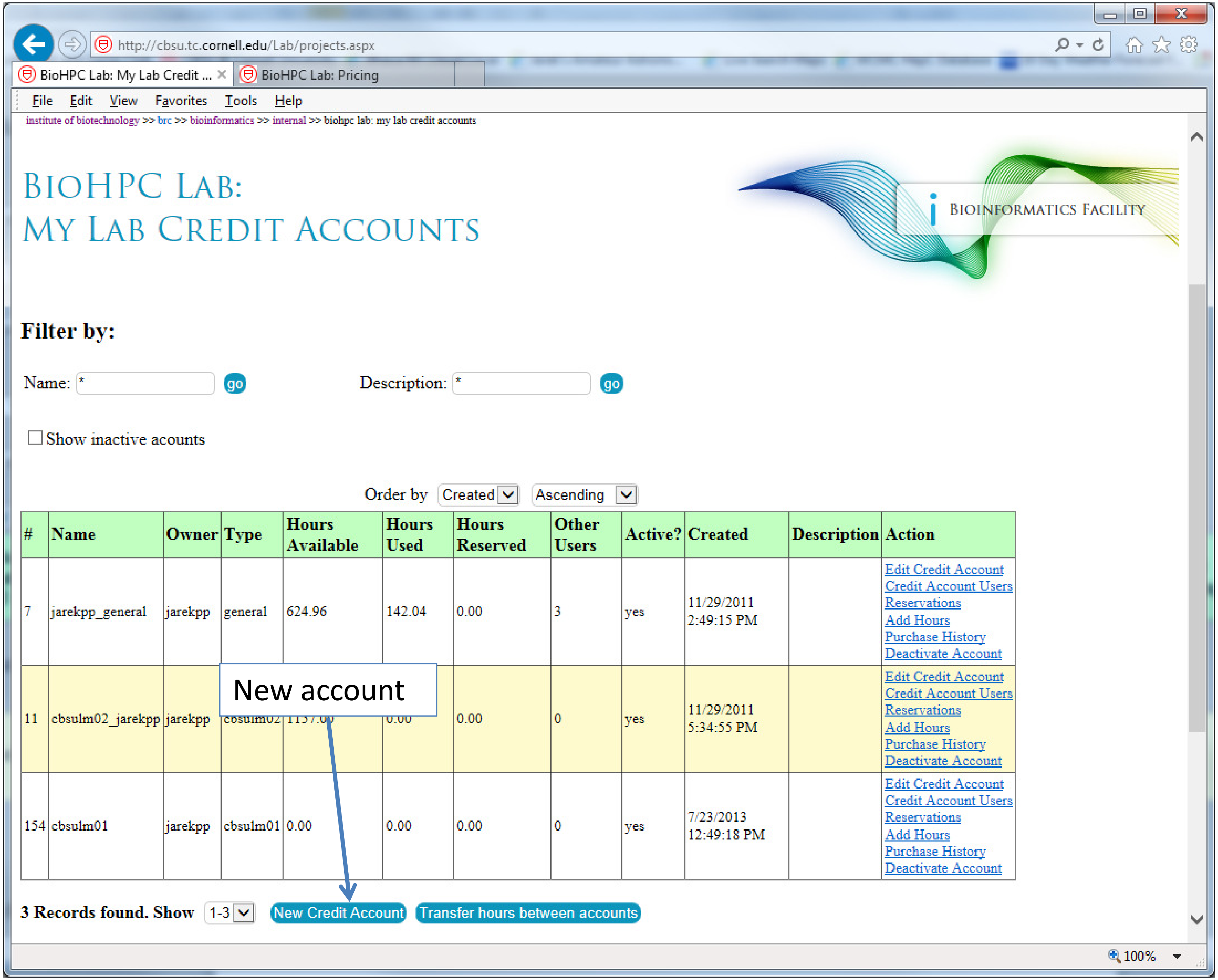
BIOINFORMATICS INTE
HOME

Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (Computational Biology Service Unit, C
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 Web Computing

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



BIOHPC LAB:
MY LAB CREDIT ACCOUNTS

Filter by:

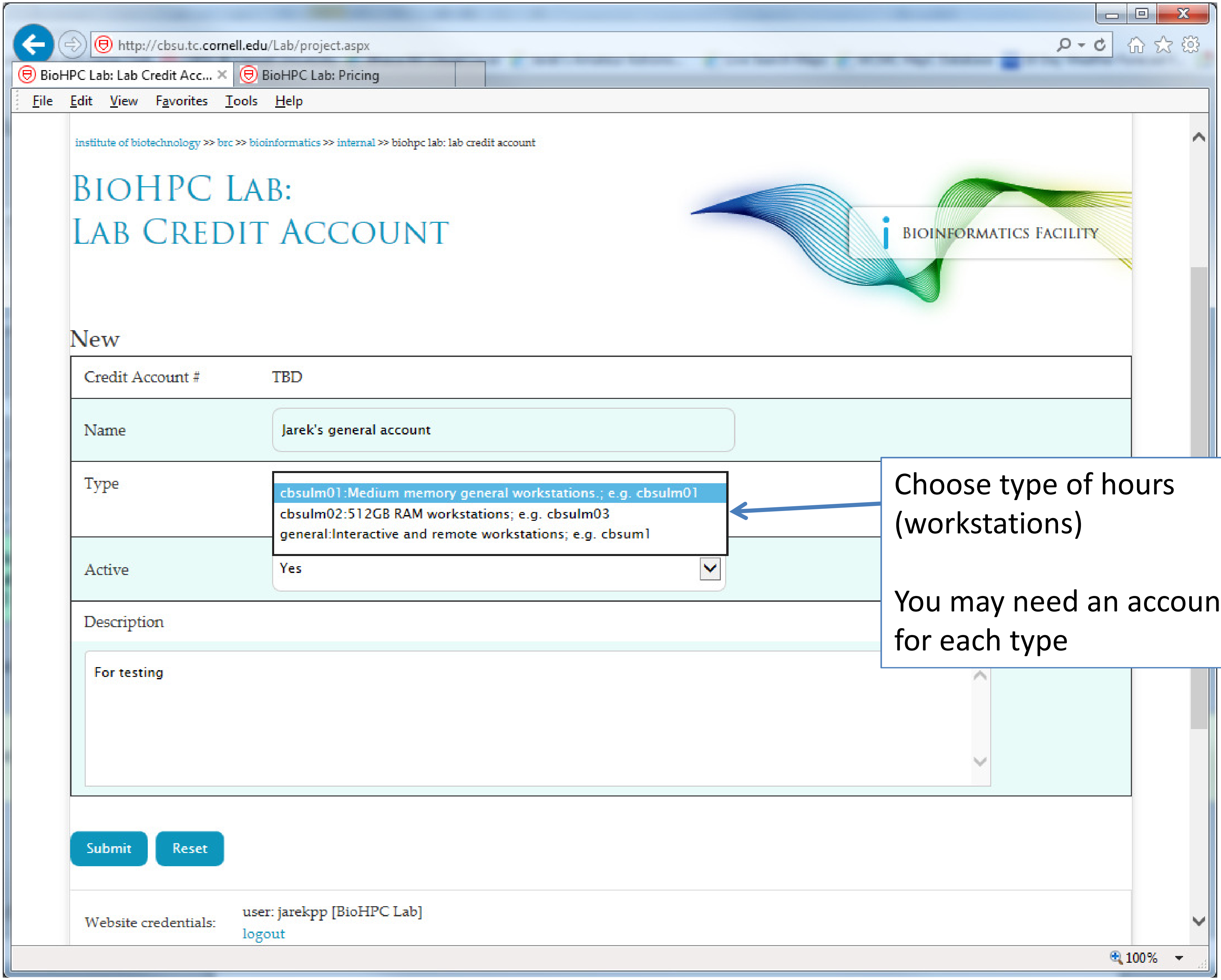
Name: * go Description: * go

☐ Show inactive acounts

Order by Created Ascending

#	Name	Owner	Type	Hours Available	Hours Used	Hours Reserved	Other Users	Active?	Created	Description	Action
7	jarekpp_general	jarekpp	general	624.96	142.04	0.00	3	yes	11/29/2011 2:49:15 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
11	cbsulm02_jarekpp	jarekpp	cbsulm02	1137.0	0.00	0.00	0	yes	11/29/2011 5:34:55 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
154	cbsulm01	jarekpp	cbsulm01	0.00	0.00	0.00	0	yes	7/23/2013 12:49:18 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account

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



BIOHPC LAB:
LAB CREDIT ACCOUNT

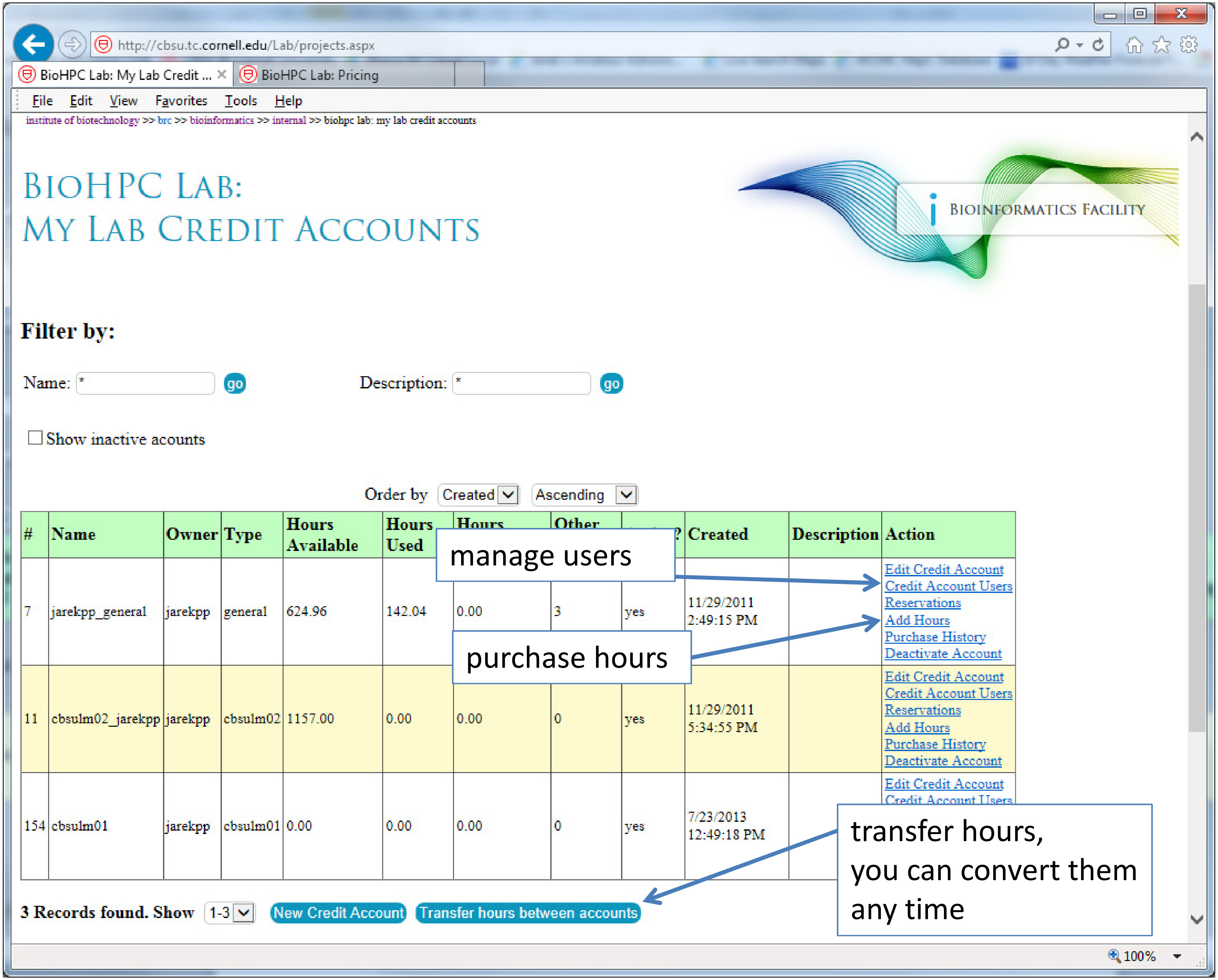
New

Credit Account #	TBD
Name	Jarek's general account
Type	cbsulm01:Medium memory general workstations.; e.g. cbsulm01 cbsulm02:512GB RAM workstations; e.g. cbsulm03 general:Interactive and remote workstations; e.g. cbsum1
Active	Yes
Description	For testing

Website credentials: user: jarekpp [BioHPC Lab]
logout

Choose type of hours (workstations)

You may need an account for each type



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: my lab credit accounts

BIOHPC LAB:
MY LAB CREDIT ACCOUNTS

Filter by:

Name: * go

Description: * go

☐ Show inactive acounts

Order by Created Ascending

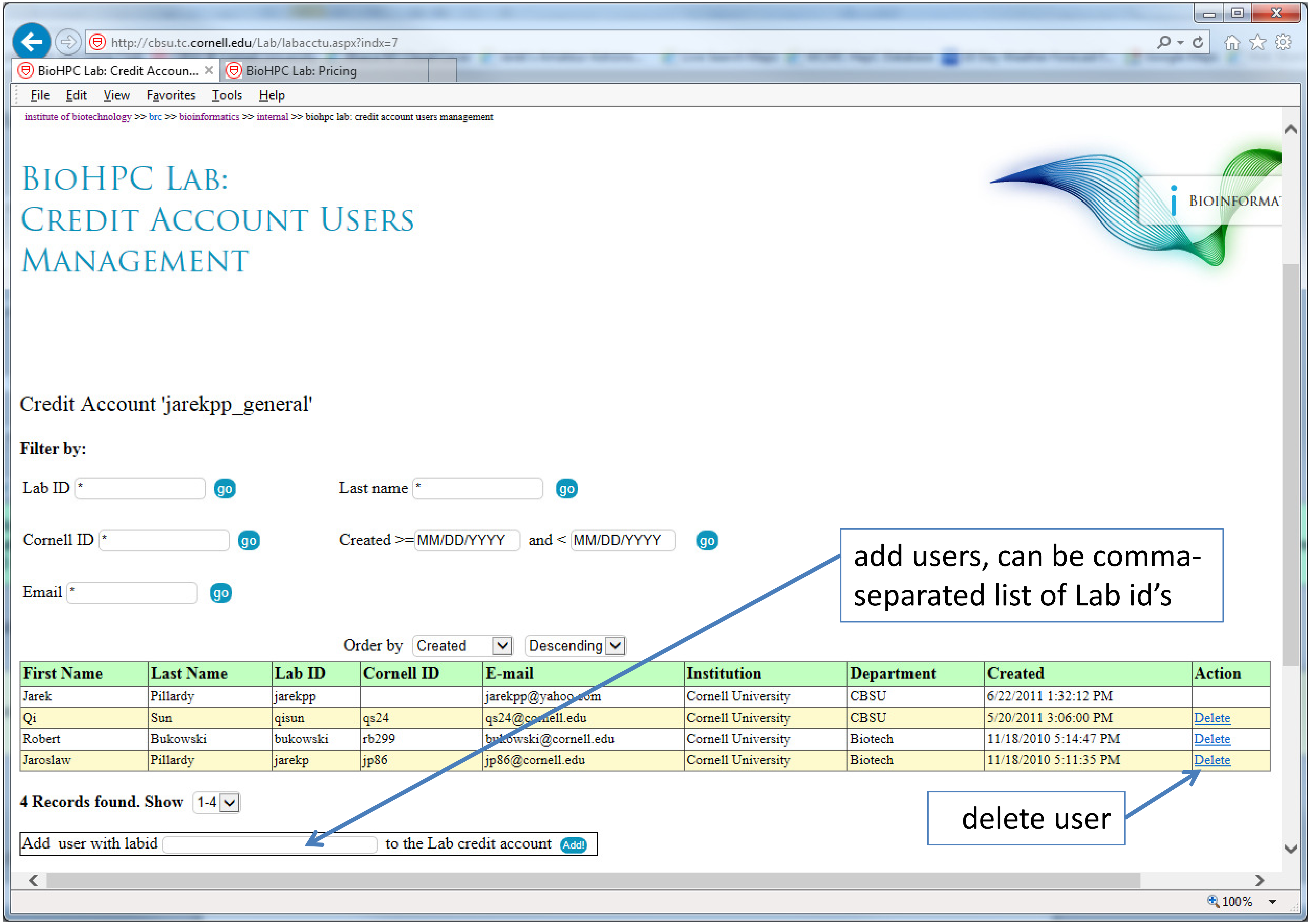
#	Name	Owner	Type	Hours Available	Hours Used	Hours	Other	?	Created	Description	Action
7	jarekpp_general	jarekpp	general	624.96	142.04	0.00	3	yes	11/29/2011 2:49:15 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
11	cbsulm02_jarekpp	jarekpp	cbsulm02	1157.00	0.00	0.00	0	yes	11/29/2011 5:34:55 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
154	cbsulm01	jarekpp	cbsulm01	0.00	0.00	0.00	0	yes	7/23/2013 12:49:18 PM		Edit Credit Account Credit Account Users

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

manage users

purchase hours

transfer hours,
you can convert them
any time



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: credit account users management

BIOHPC LAB: CREDIT ACCOUNT USERS MANAGEMENT

Credit Account 'jarekpp_general'

Filter by:

Lab ID * go
Last name * go
Cornell ID * go
Created >= MM/DD/YYYY and < MM/DD/YYYY go
Email * go

Order by Created Descending

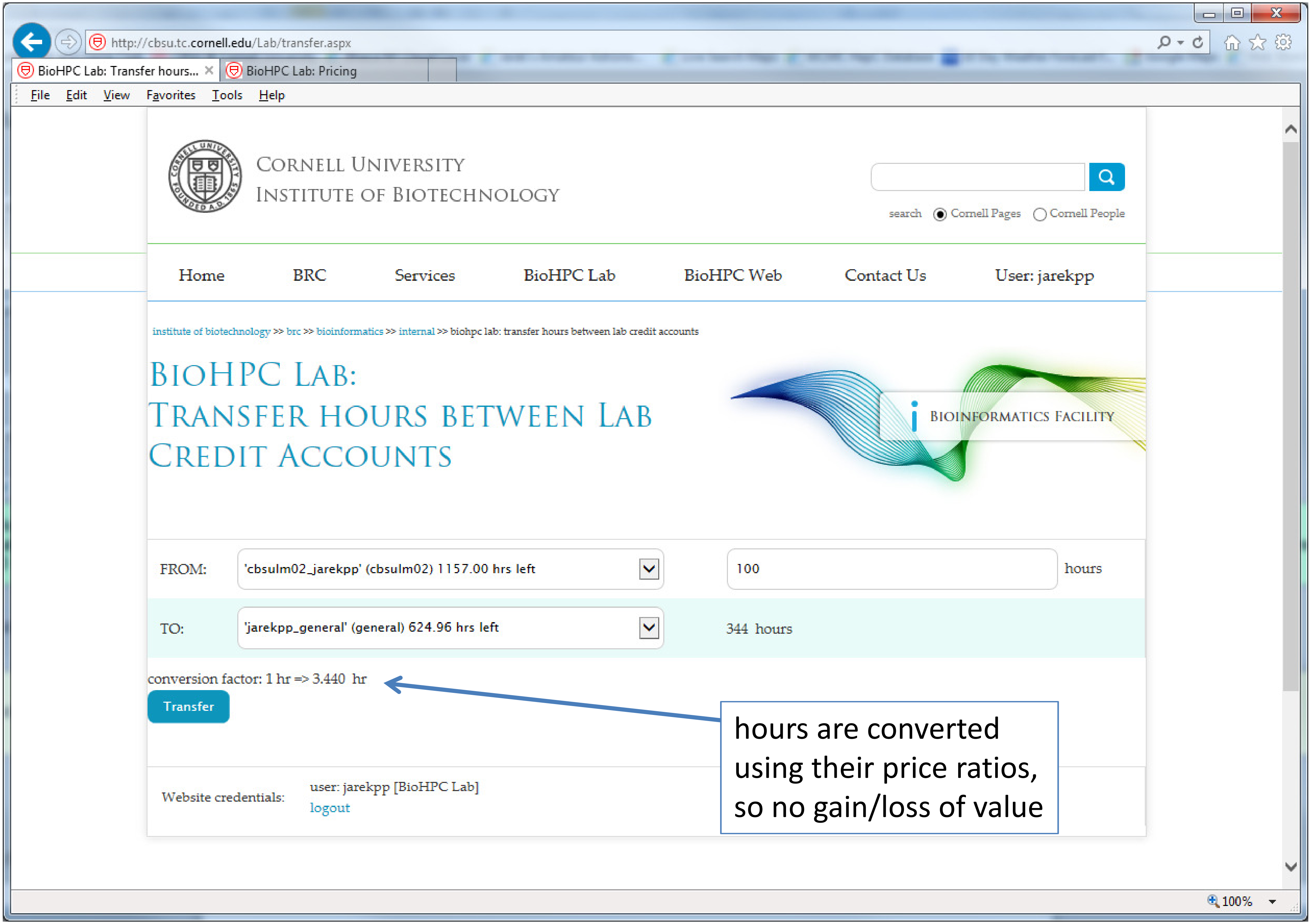
First Name	Last Name	Lab ID	Cornell ID	E-mail	Institution	Department	Created	Action
Jarek	Pillardy	jarekpp		jarekpp@yahoo.com	Cornell University	CBSU	6/22/2011 1:32:12 PM	
Qi	Sun	qisun	qs24	qs24@cornell.edu	Cornell University	CBSU	5/20/2011 3:06:00 PM	Delete
Robert	Bukowski	bukowski	rb299	bukowski@cornell.edu	Cornell University	Biotech	11/18/2010 5:14:47 PM	Delete
Jaroslaw	Pillardy	jarekp	jp86	jp86@cornell.edu	Cornell University	Biotech	11/18/2010 5:11:35 PM	Delete

4 Records found. Show 1-4

Add user with labid ______ to the Lab credit account Add!

add users, can be comma-separated list of Lab id's

delete user



CORNELL UNIVERSITY
INSTITUTE OF BIOTECHNOLOGY

institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: transfer hours between lab credit accounts

BIOHPC LAB:
TRANSFER HOURS BETWEEN LAB CREDIT ACCOUNTS

BIOINFORMATICS FACILITY

FROM:	'cbsulm02_jarekpp' (cbsulm02) 1157.00 hrs left	100 hours
TO:	'jarekpp_general' (general) 624.96 hrs left	344 hours

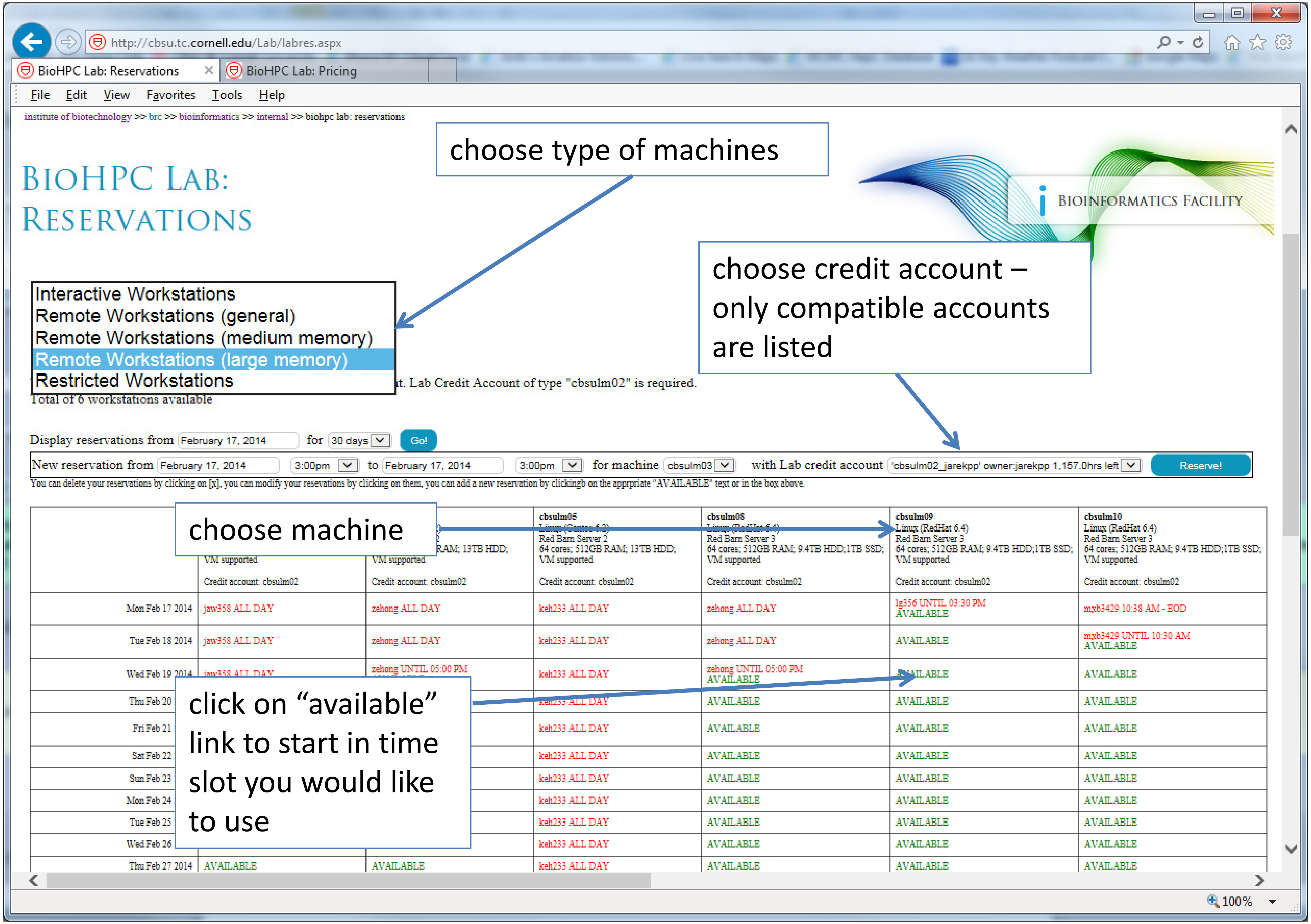
conversion factor: 1 hr => 3.440 hr

Website credentials: user: jarekpp [BioHPC Lab]
logout

hours are converted using their price ratios, so no gain/loss of value

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!*



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: reservations

BIOHPC LAB: RESERVATIONS

choose type of machines

- Interactive Workstations
- Remote Workstations (general)
- Remote Workstations (medium memory)
- Remote Workstations (large memory)
- Restricted Workstations

t. Lab Credit Account of type "cbsulm02" is required.

Total of 6 workstations available

choose credit account – only compatible accounts are listed

Display reservations from February 17, 2014 for 30 days

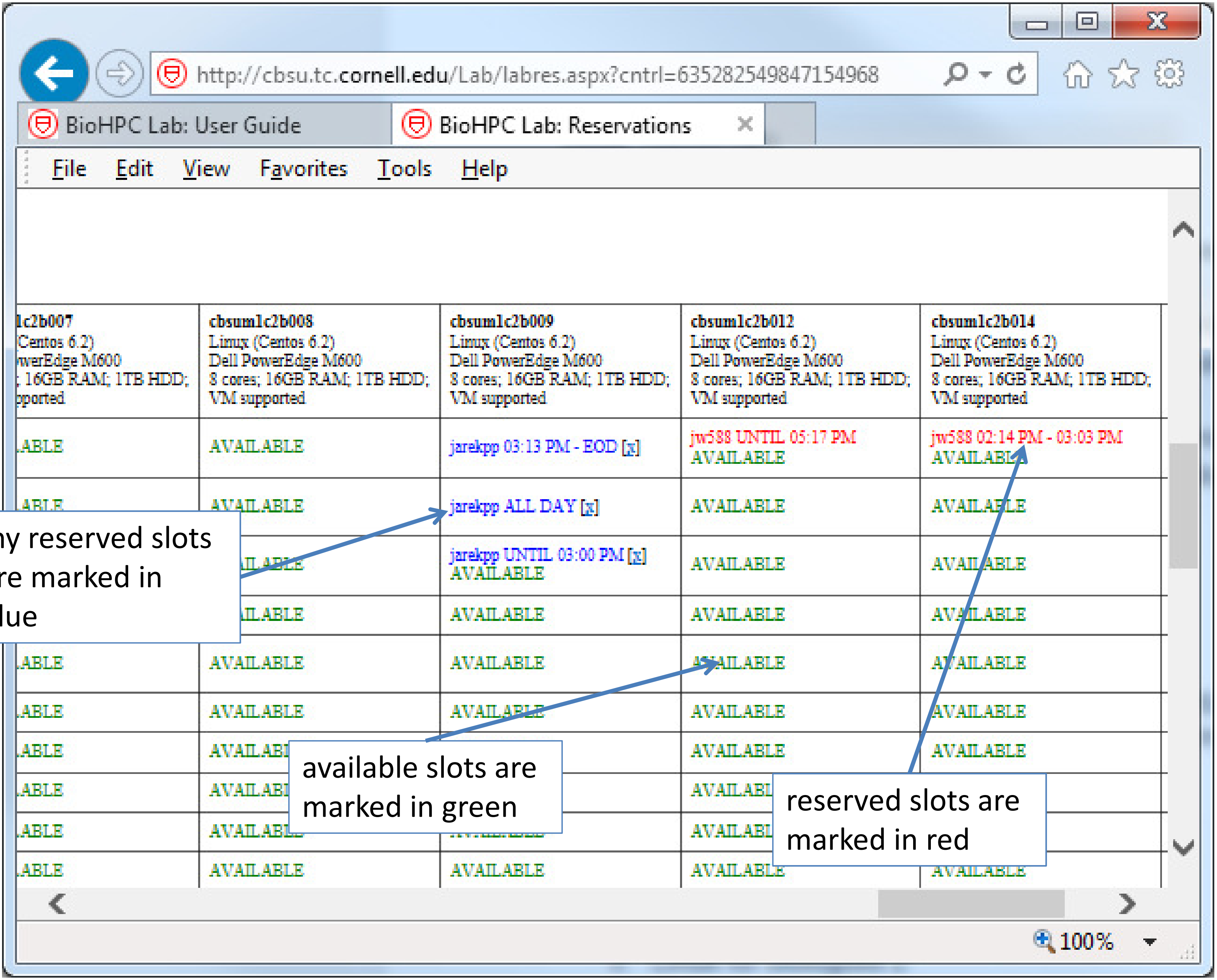
New reservation from February 17, 2014 3:00pm to February 17, 2014 3:00pm for machine cbsulm03 with Lab credit account 'cbsulm02_jarekpp' owner:jarekpp 1,157.0hrs left

You can delete your reservations by clicking on [x], you can modify your resevations by clicking on them, you can add a new reservation by clickingb on the apprpriate "AVAILABLE" text or in the box above.

choose machine

	VM supported Credit account: cbsulm02	RAM; 13TB HDD; VM supported Credit account: cbsulm02	**cbsulm05** Linux (Centos 6.2) Red Barn Server 2 64 cores; 512GB RAM; 13TB HDD; VM supported Credit account: cbsulm02	**cbsulm08** Linux (RedHat 6.4) Red Barn Server 3 64 cores; 512GB RAM; 9.4TB HDD;1TB SSD; VM supported Credit account: cbsulm02	**cbsulm09** Linux (RedHat 6.4) Red Barn Server 3 64 cores; 512GB RAM; 9.4TB HDD;1TB SSD; VM supported Credit account: cbsulm02	**cbsulm10** Linux (RedHat 6.4) Red Barn Server 3 64 cores; 512GB RAM; 9.4TB HDD;1TB SSD; VM supported Credit account: cbsulm02
Mon Feb 17 2014	jaw358 ALL DAY	zehong ALL DAY	keh233 ALL DAY	zehong ALL DAY	lg356 UNTIL 03:30 PM AVAILABLE	mxb3429 10:38 AM - EOD
Tue Feb 18 2014	jaw358 ALL DAY	zehong ALL DAY	keh233 ALL DAY	zehong ALL DAY	AVAILABLE	mxb3429 UNTIL 10:30 AM AVAILABLE
Wed Feb 19 2014	jaw358 ALL DAY	zehong UNTIL 05:00 PM	keh233 ALL DAY	zehong UNTIL 05:00 PM AVAILABLE	AVAILABLE	AVAILABLE
Thu Feb 20			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Fri Feb 21			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Sat Feb 22			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Sun Feb 23			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Mon Feb 24			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Tue Feb 25			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Wed Feb 26			keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE
Thu Feb 27 2014	AVAILABLE	AVAILABLE	keh233 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE

click on “available” link to start in time slot you would like to use

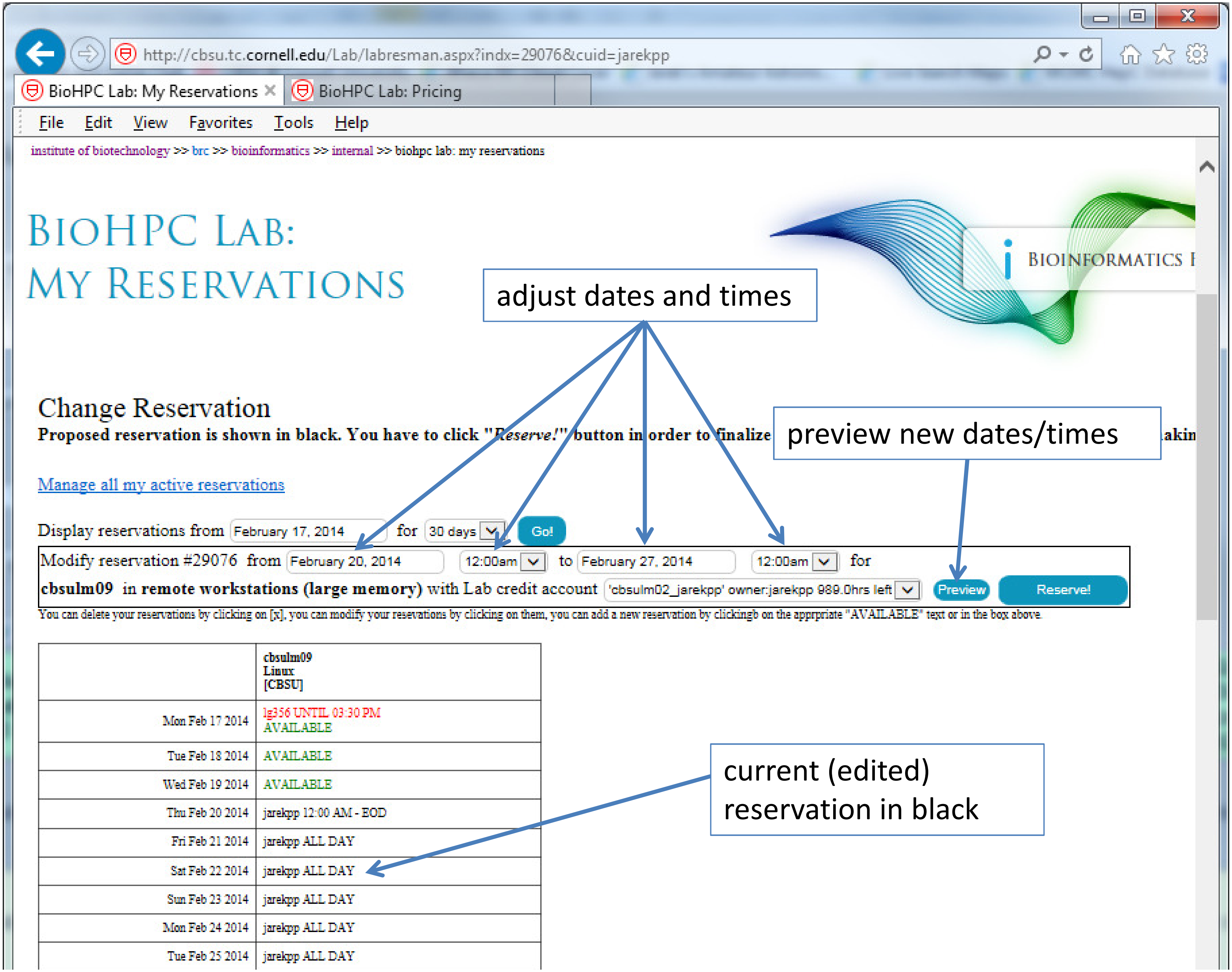


lc2b007 Centos 6.2) werEdge M600 ; 16GB RAM; 1TB HDD; pported	cbsumlc2b008 Linux (Centos 6.2) Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported	cbsumlc2b009 Linux (Centos 6.2) Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported	cbsumlc2b012 Linux (Centos 6.2) Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported	cbsumlc2b014 Linux (Centos 6.2) Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported
ABLE	AVAILABLE	jarekpp 03:13 PM - EOD [x]	jw588 UNTIL 05:17 PM AVAILABLE	jw588 02:14 PM - 03:03 PM AVAILABLE
ABLE	AVAILABLE	jarekpp ALL DAY [x]	AVAILABLE	AVAILABLE
	ILABLE	jarekpp UNTIL 03:00 PM [x] AVAILABLE	AVAILABLE	AVAILABLE
	ILABLE	AVAILABLE	AVAILABLE	AVAILABLE
ABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
ABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
ABLE	AVAILABL		AVAILABLE	AVAILABLE
ABLE	AVAILABL		AVAILABL	
ABLE	AVAILABL		AVAILABL	
ABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE

my reserved slots are marked in blue

available slots are marked in green

reserved slots are marked in red



BIOHPC LAB:
MY RESERVATIONS

adjust dates and times

Change Reservation

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

preview new dates/times

Manage all my active reservations

Display reservations from February 17, 2014 for 30 days Go!

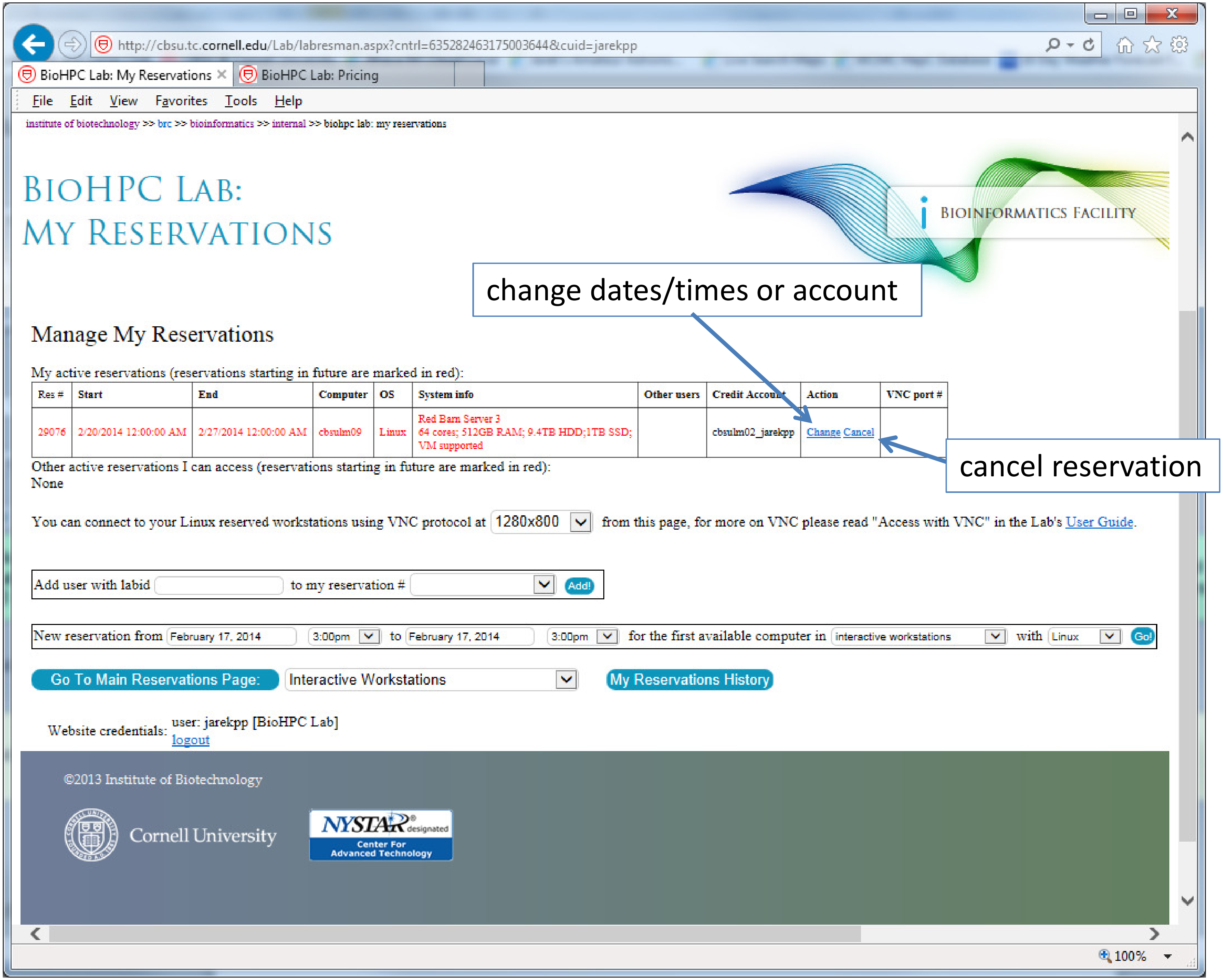
Modify reservation #29076 from February 20, 2014 12:00am to February 27, 2014 12:00am for

cbsulm09 **in remote workstations (large memory)** with Lab credit account 'cbsulm02_jarekpp' owner:jarekpp 989.0hrs left Preview Reserve!

You can delete your reservations by clicking on [x], you can modify your resevations by clicking on them, you can add a new reservation by clickingb on the apprpriate "AVAILABLE" text or in the box above.

	cbsulm09 **Linux** **[CBSU]**
Mon Feb 17 2014	lg356 UNTIL 03:30 PM AVAILABLE
Tue Feb 18 2014	AVAILABLE
Wed Feb 19 2014	AVAILABLE
Thu Feb 20 2014	jarekpp 12:00 AM - EOD
Fri Feb 21 2014	jarekpp ALL DAY
Sat Feb 22 2014	jarekpp ALL DAY
Sun Feb 23 2014	jarekpp ALL DAY
Mon Feb 24 2014	jarekpp ALL DAY
Tue Feb 25 2014	jarekpp ALL DAY

current (edited)
reservation in black



BIOHPC LAB:
MY RESERVATIONS

Manage My Reservations

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

Res #	Start	End	Computer	OS	System info	Other users	Credit Account	Action	VNC port #
29076	2/20/2014 12:00:00 AM	2/27/2014 12:00:00 AM	cbsulm09	Linux	Red Barn Server 3 64 cores; 512GB RAM; 9.4TB HDD;1TB SSD; VM supported		cbsulm02_jarekpp	Change Cancel	

Other active reservations I can access (reservations starting in future are marked in red):
None

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.

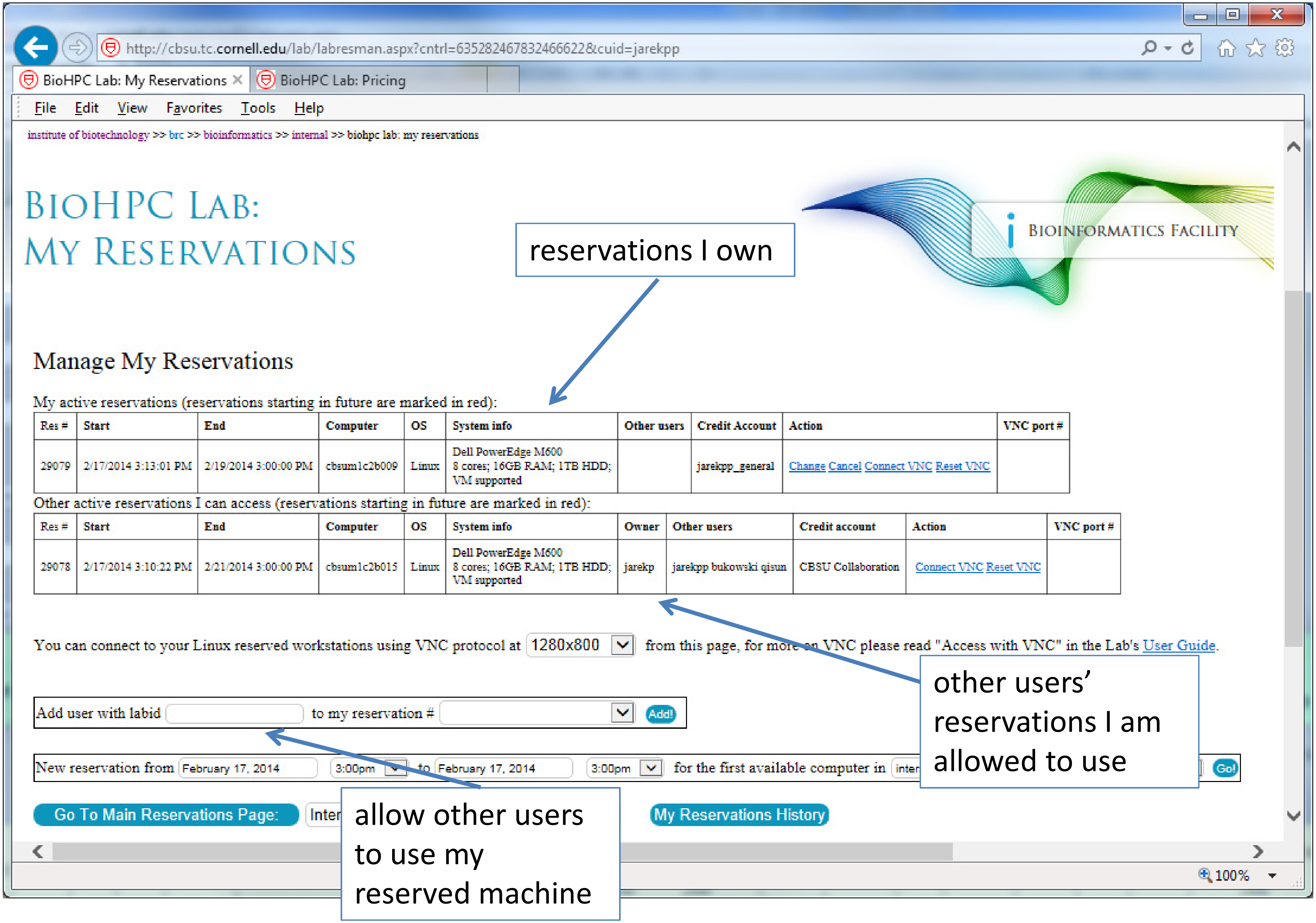
Add user with labid to my reservation # Add!

New reservation from February 17, 2014 3:00pm to February 17, 2014 3:00pm for the first available computer in interactive workstations with Linux Go!

Website credentials: user: jarekpp [BioHPC Lab]
logout

change dates/times or account

cancel reservation



BIOHPC LAB:
MY RESERVATIONS

reservations I own

Manage My Reservations

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

Res #	Start	End	Computer	OS	System info	Other users	Credit Account	Action	VNC port #
29079	2/17/2014 3:13:01 PM	2/19/2014 3:00:00 PM	cbsumlc2b009	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported		jarekpp_general	Change Cancel Connect VNC Reset VNC	

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

Res #	Start	End	Computer	OS	System info	Owner	Other users	Credit account	Action	VNC port #
29078	2/17/2014 3:10:22 PM	2/21/2014 3:00:00 PM	cbsumlc2b015	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported	jarekp	jarekpp bukowski qisun	CBSU Collaboration	Connect VNC Reset VNC	

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.

other users' reservations I am allowed to use

Add user with labid ______ to my reservation # ______ Add!

New reservation from February 17, 2014 3:00pm to February 17, 2014 3:00pm for the first available computer in inter Go!

allow other users to use my reserved machine

Reservations

You can only make reservation if you have enough hours of appropriate type.

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*

```
#!/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
- 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 (520TB), 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

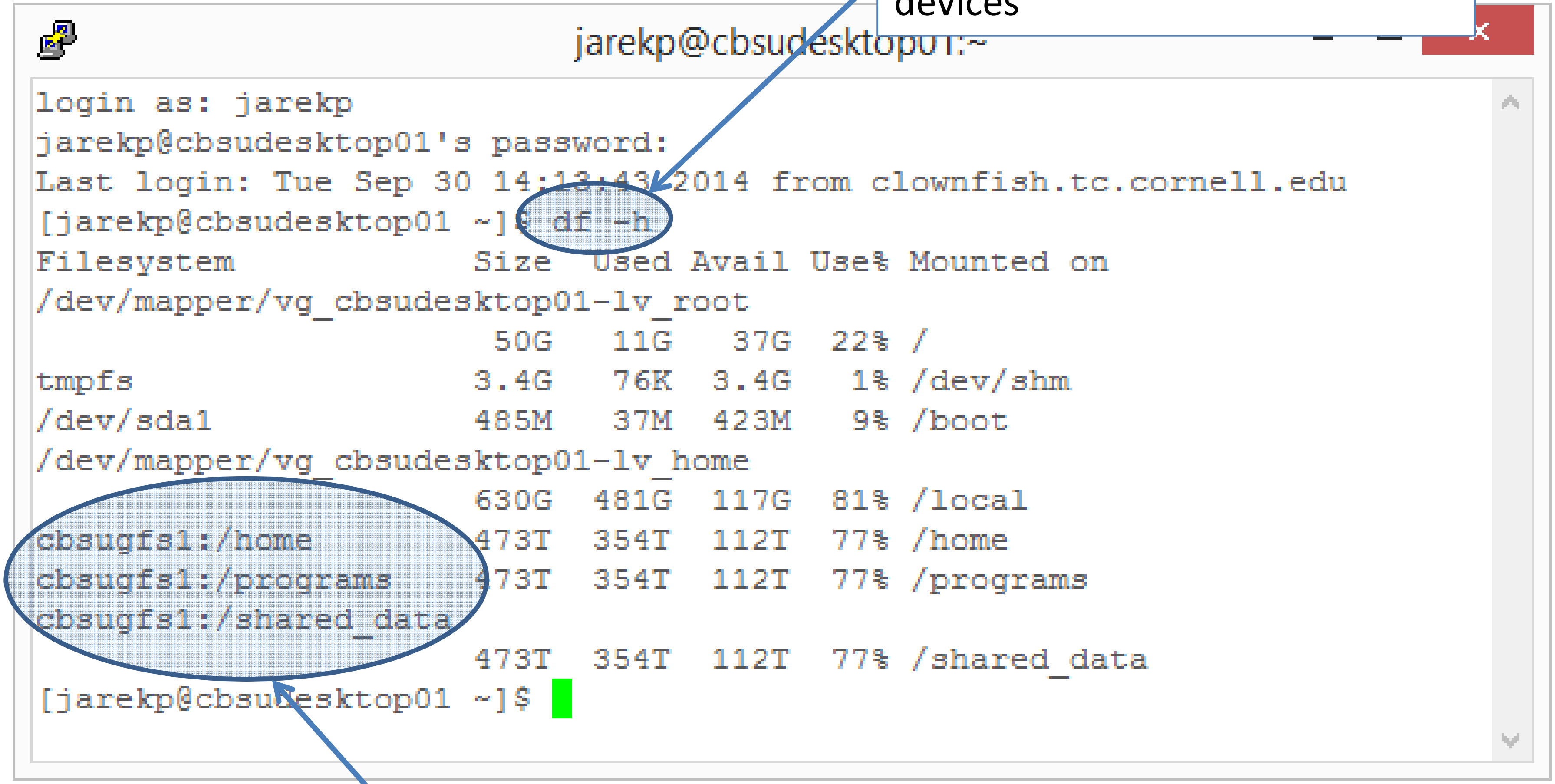
Linux directory structure is *continuous*, i.e. regardless of the physical location of storage it all seems to be part of one directory tree starting from root (/).

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

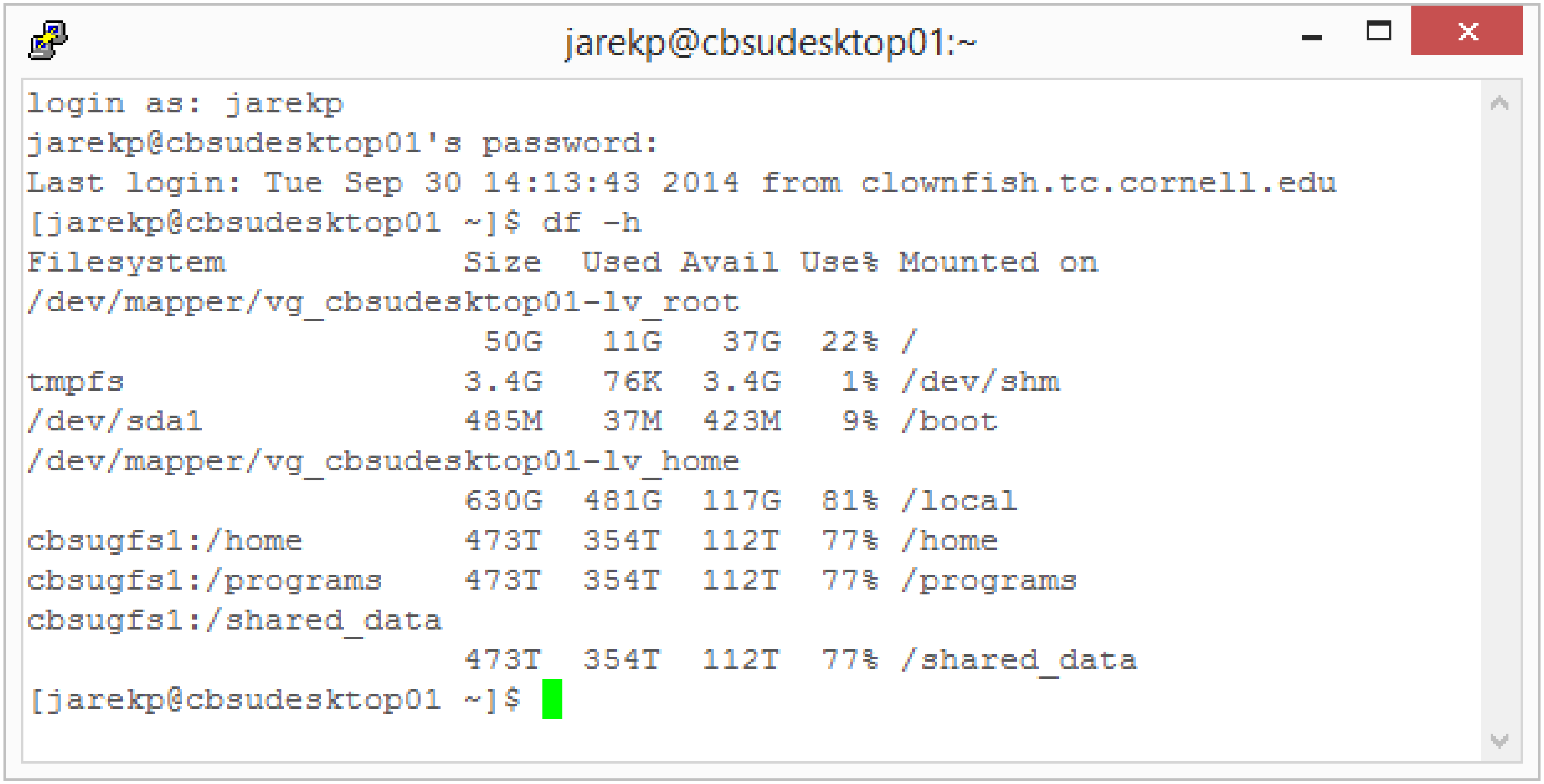
If in doubt “df -h” lists all storage devices



```
login as: jarekp
jarekp@cbsudesktop01's password:
Last login: Tue Sep 30 14:13:43 2014 from clownfish.tc.cornell.edu
[jarekp@cbsudesktop01 ~]$ df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/vg_cbsudesktop01-lv_root
                       50G   11G   37G  22% /
tmpfs                 3.4G   76K  3.4G   1% /dev/shm
/dev/sda1             485M   37M  423M   9% /boot
/dev/mapper/vg_cbsudesktop01-lv_home
                      630G  481G  117G  81% /local
cbsugfs1:/home        473T  354T  112T  77% /home
cbsugfs1:/programs    473T  354T  112T  77% /programs
cbsugfs1:/shared_data
                      473T  354T  112T  77% /shared_data
[jarekp@cbsudesktop01 ~]$
```

These are network devices – starting with “computername:”

Storage



```
login as: jarekp
jarekp@cbsudesktop01's password:
Last login: Tue Sep 30 14:13:43 2014 from clownfish.tc.cornell.edu
[jarekp@cbsudesktop01 ~]$ df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/vg_cbsudesktop01-lv_root
                       50G   11G   37G  22% /
tmpfs                 3.4G   76K  3.4G   1% /dev/shm
/dev/sda1             485M   37M  423M   9% /boot
/dev/mapper/vg_cbsudesktop01-lv_home
                      630G  481G  117G  81% /local
cbsugfs1:/home        473T  354T  112T  77% /home
cbsugfs1:/programs    473T  354T  112T  77% /programs
cbsugfs1:/shared_data
                      473T  354T  112T  77% /shared_data
[jarekp@cbsudesktop01 ~]$ 
```

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

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.

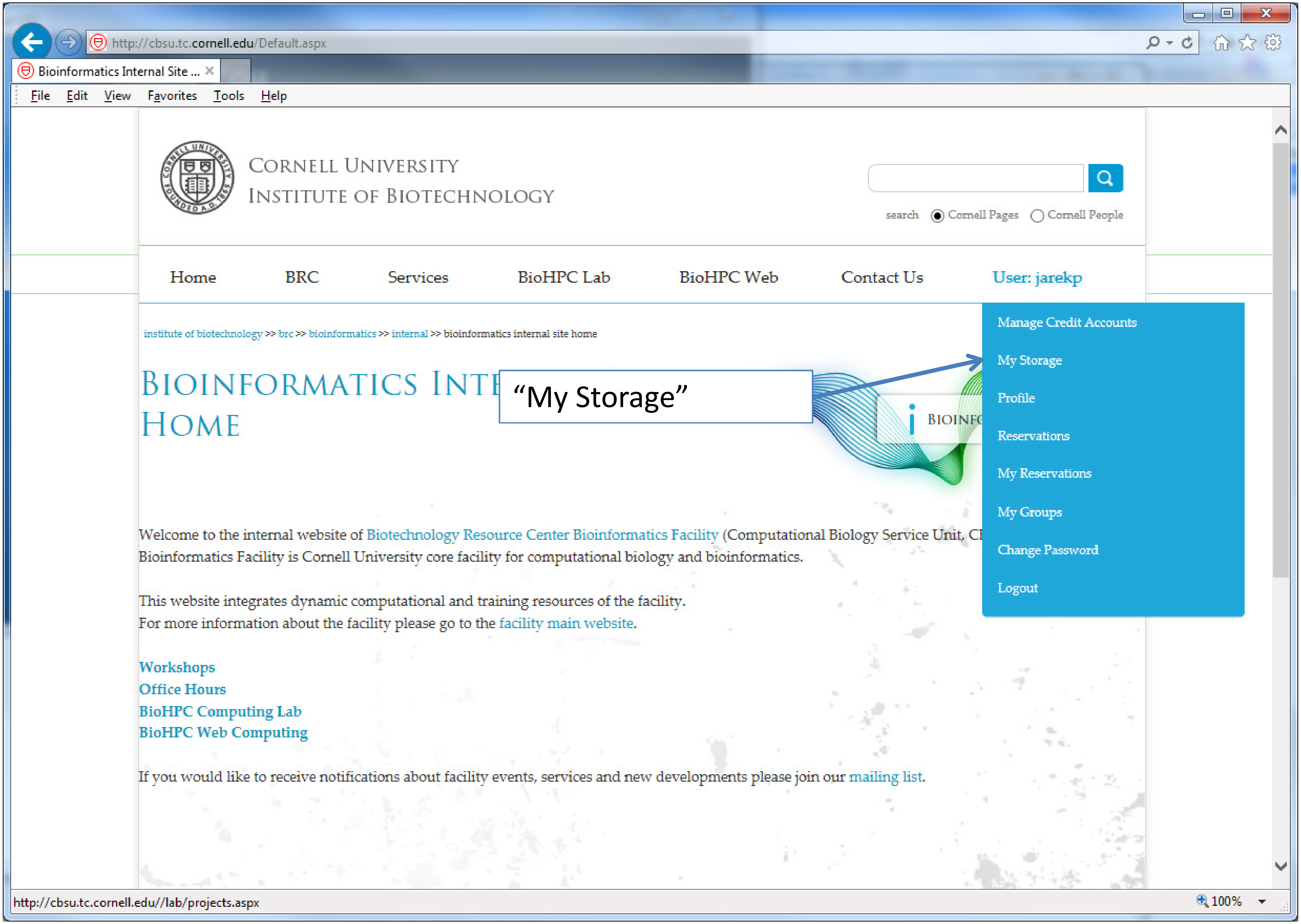
All home directories on BioHPC Lab are networked.

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 at 5 am.



CORNELL UNIVERSITY
INSTITUTE OF BIOTECHNOLOGY

institute of biotechnology >> brc >> bioinformatics >> internal >> bioinformatics internal site home

BIOINFORMATICS INTE
HOME

Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (Computational Biology Service Unit, CB
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 Web Computing

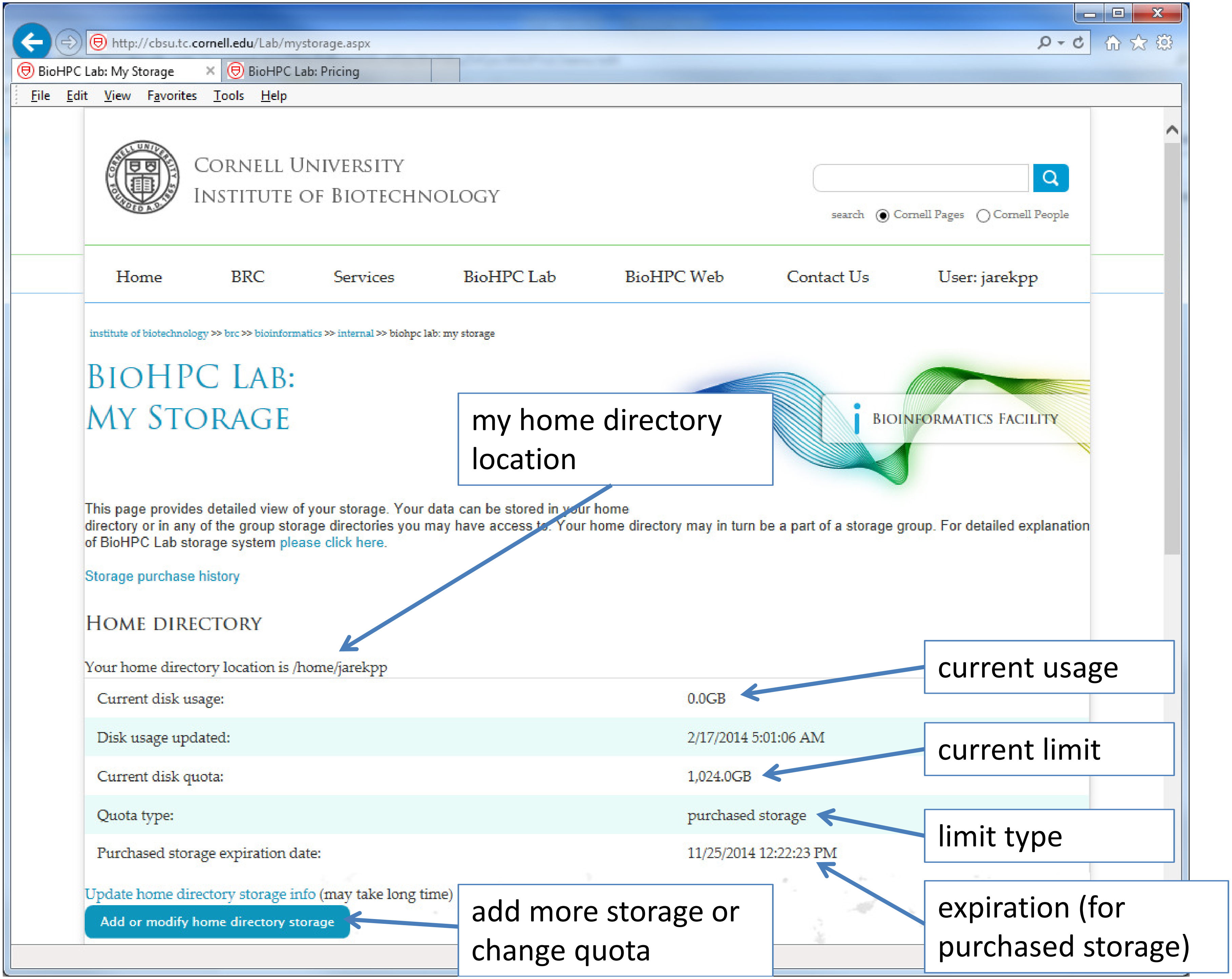
If you would like to receive notifications about facility events, services and new developments please join our mailing list.

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 LAB: MY STORAGE

This page provides 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 Lab storage system please click here.

Storage purchase history

HOME DIRECTORY

Your home directory location is /home/jarekpp

Current disk usage:	0.0GB
Disk usage updated:	2/17/2014 5:01:06 AM
Current disk quota:	1,024.0GB
Quota type:	purchased storage
Purchased storage expiration date:	11/25/2014 12:22:23 PM

Update home directory storage info (may take long time)

Add or modify home directory storage

my home directory location

current usage

current limit

limit type

expiration (for purchased storage)

add more storage or change quota

Storage

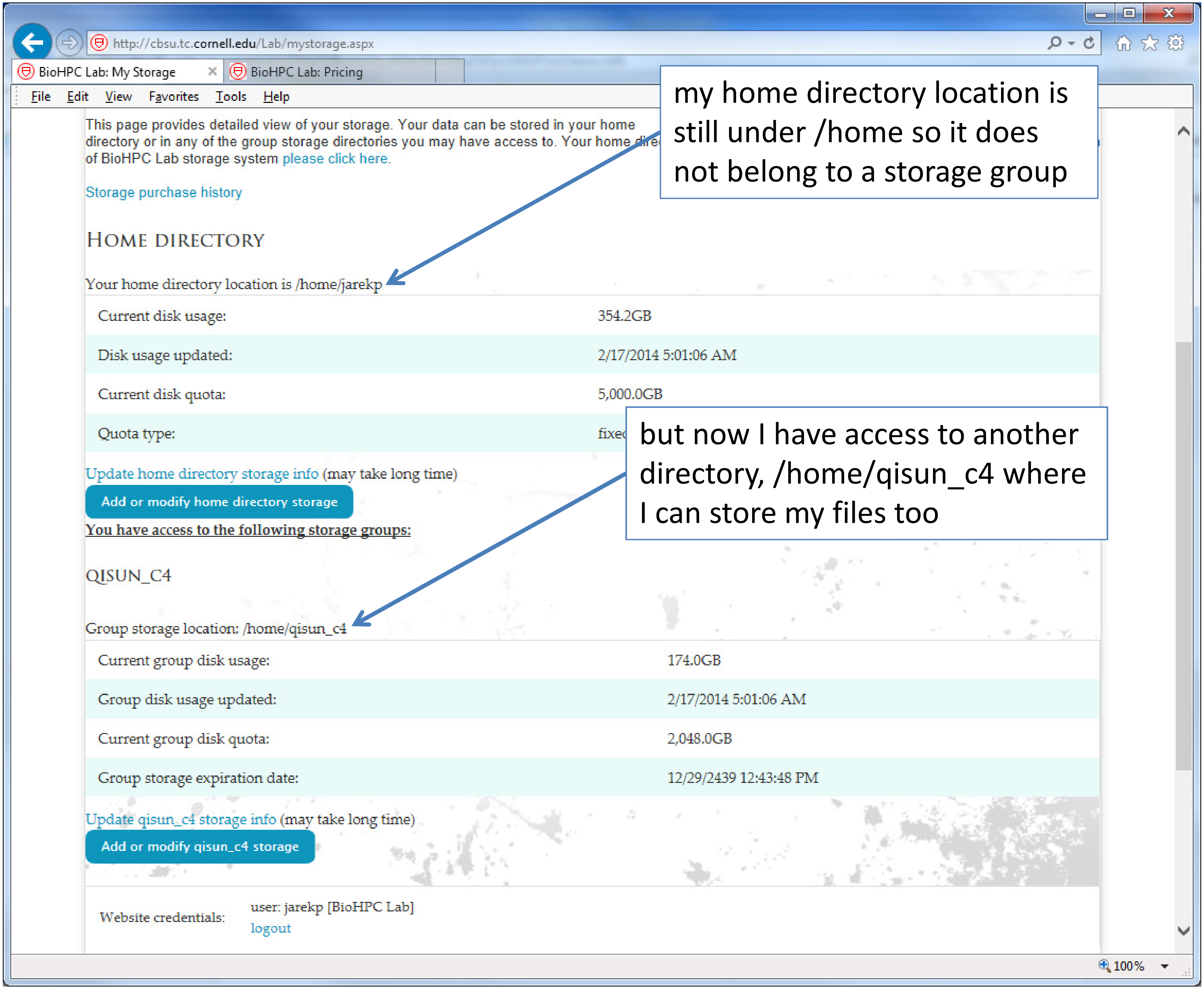
Extra storage can be purchased for $87.13 per TB per year.

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”



This page provides 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 dire
of BioHPC Lab storage system please click here.

Storage purchase history

HOME DIRECTORY

Your home directory location is /home/jarekp

Current disk usage:	354.2GB
Disk usage updated:	2/17/2014 5:01:06 AM
Current disk quota:	5,000.0GB
Quota type:	fixe

Update home directory storage info (may take long time)

You have access to the following storage groups:

QISUN_C4

Group storage location: /home/qisun_c4

Current group disk usage:	174.0GB
Group disk usage updated:	2/17/2014 5:01:06 AM
Current group disk quota:	2,048.0GB
Group storage expiration date:	12/29/2439 12:43:48 PM

Update qisun_c4 storage info (may take long time)

Website credentials: user: jarekp [BioHPC Lab]
logout

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

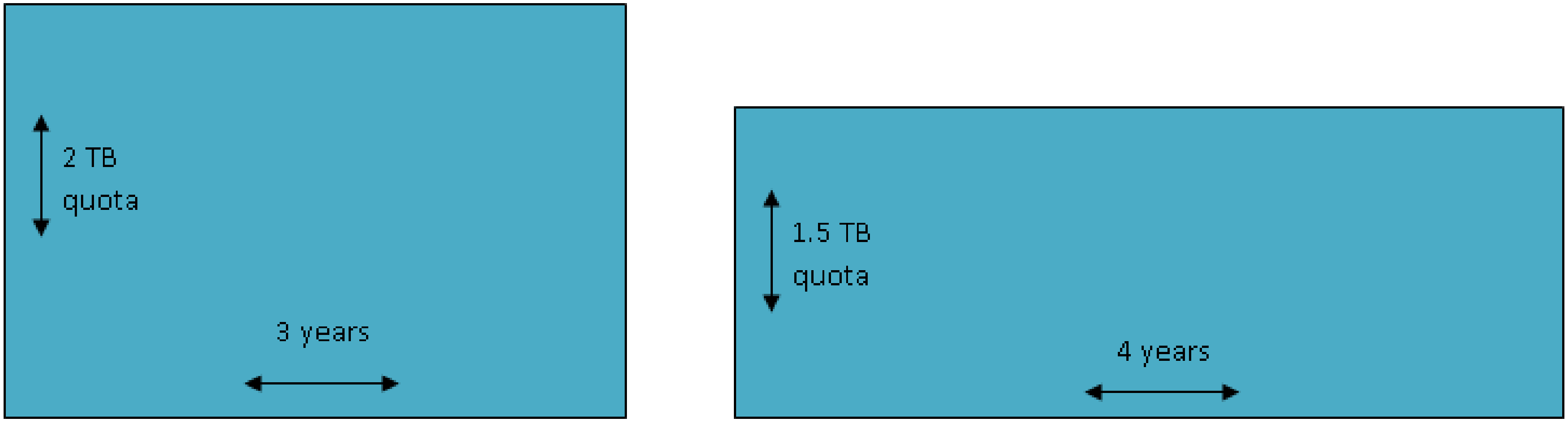
Storage

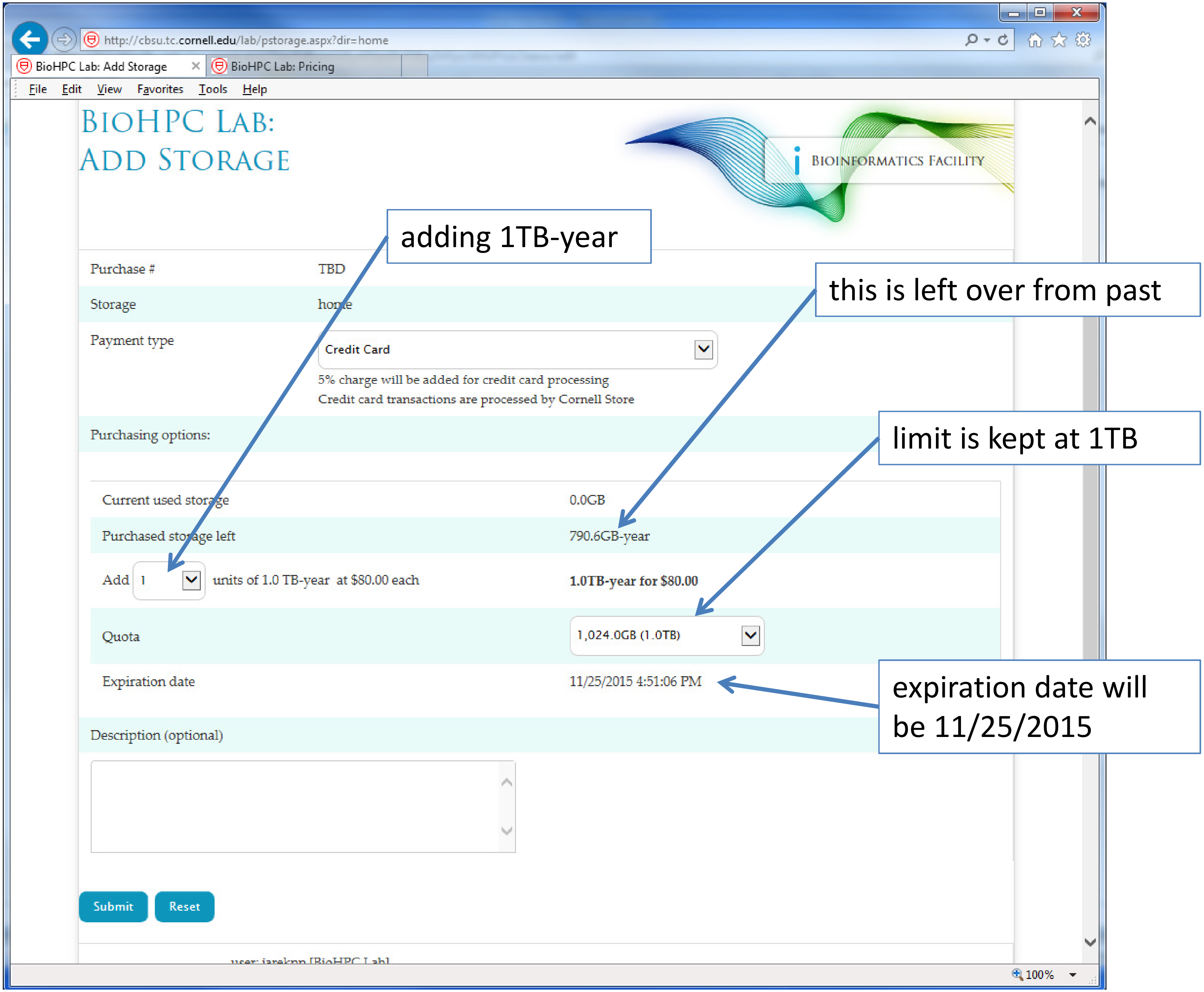
- The storage can be only purchased in 1 TB-year chunks, it needs to be done up front, and you can set your quota to an appropriate size, which in turn will decide the expiration date.
- You can buy as many of the 1TB-year chunks as you want and then set the quota at the level you want, the expiration date will be computed as the result.

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.





BIOHPC LAB:
ADD STORAGE

Purchase #	TBD
Storage	home
Payment type	Credit Card 5% charge will be added for credit card processing Credit card transactions are processed by Cornell Store
Purchasing options:	

Current used storage	0.0GB
Purchased storage left	790.6GB-year
Add 1 units of 1.0 TB-year at $80.00 each	**1.0TB-year for $80.00**
Quota	1,024.0GB (1.0TB)
Expiration date	11/25/2015 4:51:06 PM

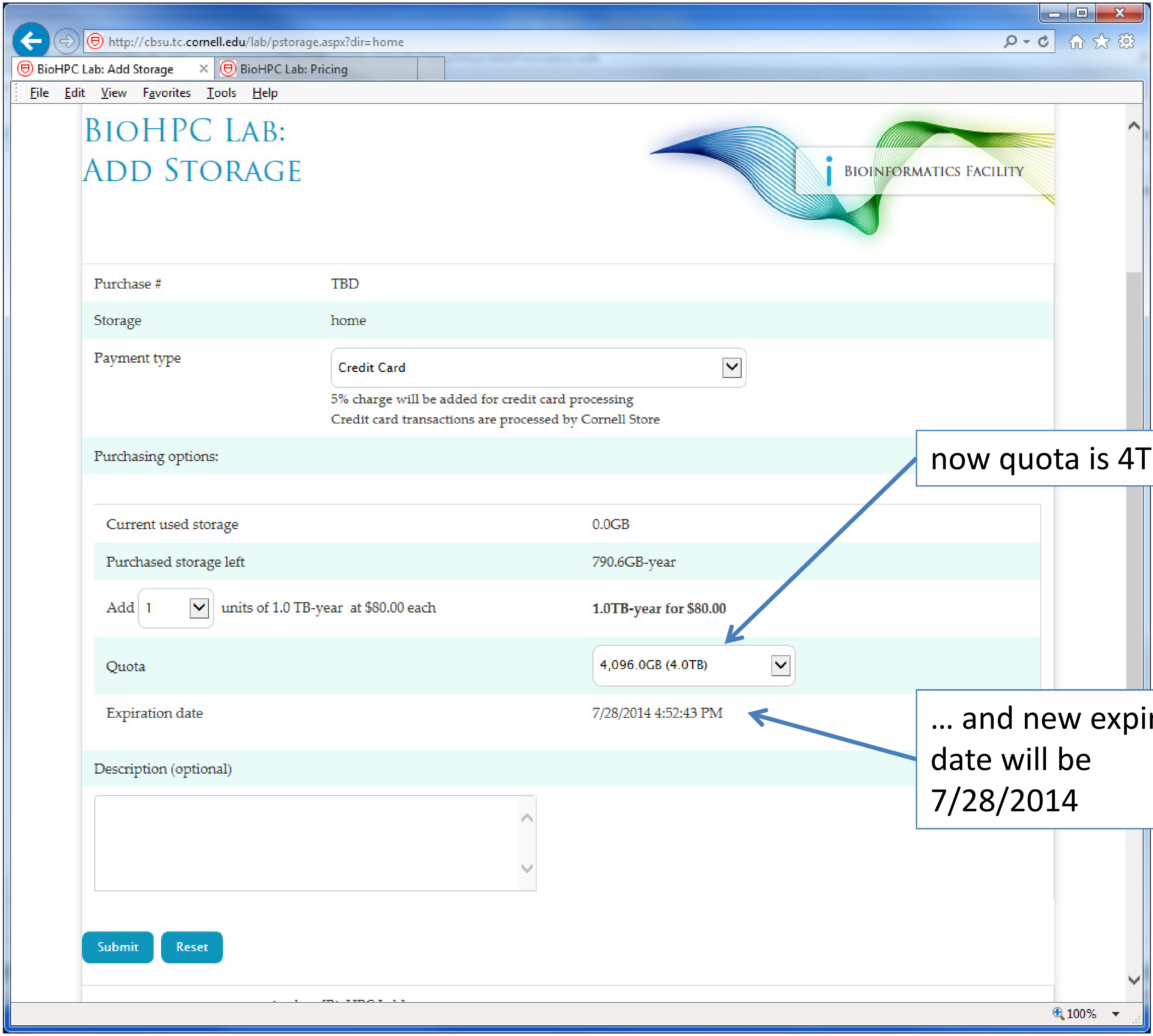
Description (optional)

adding 1TB-year

this is left over from past

limit is kept at 1TB

expiration date will be 11/25/2015



BIOHPC LAB:
ADD STORAGE

Purchase #	TBD
Storage	home
Payment type	Credit Card 5% charge will be added for credit card processing Credit card transactions are processed by Cornell Store

Purchasing options:

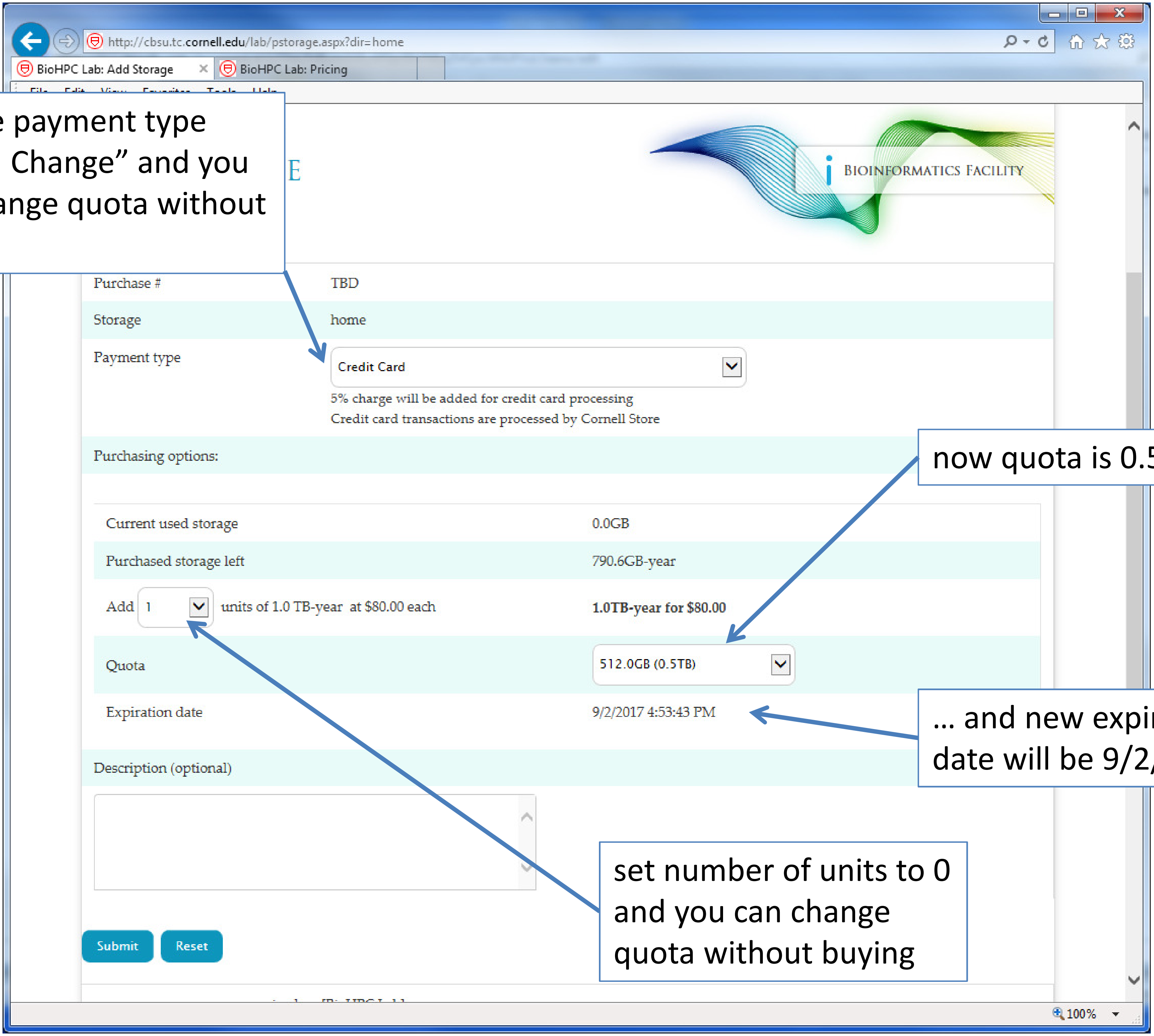
Current used storage	0.0GB
Purchased storage left	790.6GB-year
Add 1 units of 1.0 TB-year at $80.00 each	**1.0TB-year for $80.00**
Quota	4,096.0GB (4.0TB)
Expiration date	7/28/2014 4:52:43 PM

Description (optional)

now quota is 4TB

… and new expiration date will be 7/28/2014

Choose payment type "Quota Change" and you can change quota without buying



Purchase #	TBD
Storage	home
Payment type	Credit Card 5% charge will be added for credit card processing Credit card transactions are processed by Cornell Store

Purchasing options:

Current used storage	0.0GB
Purchased storage left	790.6GB-year
Add 1 units of 1.0 TB-year at $80.00 each	**1.0TB-year for $80.00**
Quota	512.0GB (0.5TB)
Expiration date	9/2/2017 4:53:43 PM

Description (optional)

now quota is 0.5TB

... and new expiration date will be 9/2/2017

set number of units to 0 and you can change quota without buying

Storage

Similarly as with Lab Credit Accounts computing hours you are charged for *reservation* of storage, i.e. your TB-year storage purchased is used and subtracted based on your *quota*, NOT the amount of actually stored data.

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
- ✖ 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 Lab 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 <u>account</u>, 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 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 terminal window. Type

    ssh jarekp@cbsuwrkstX.tc.cornell.edu

    (replace the "**cbsuwrkstX"** with the workstation that you just reserved, and "**jarekp**" 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 the 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)

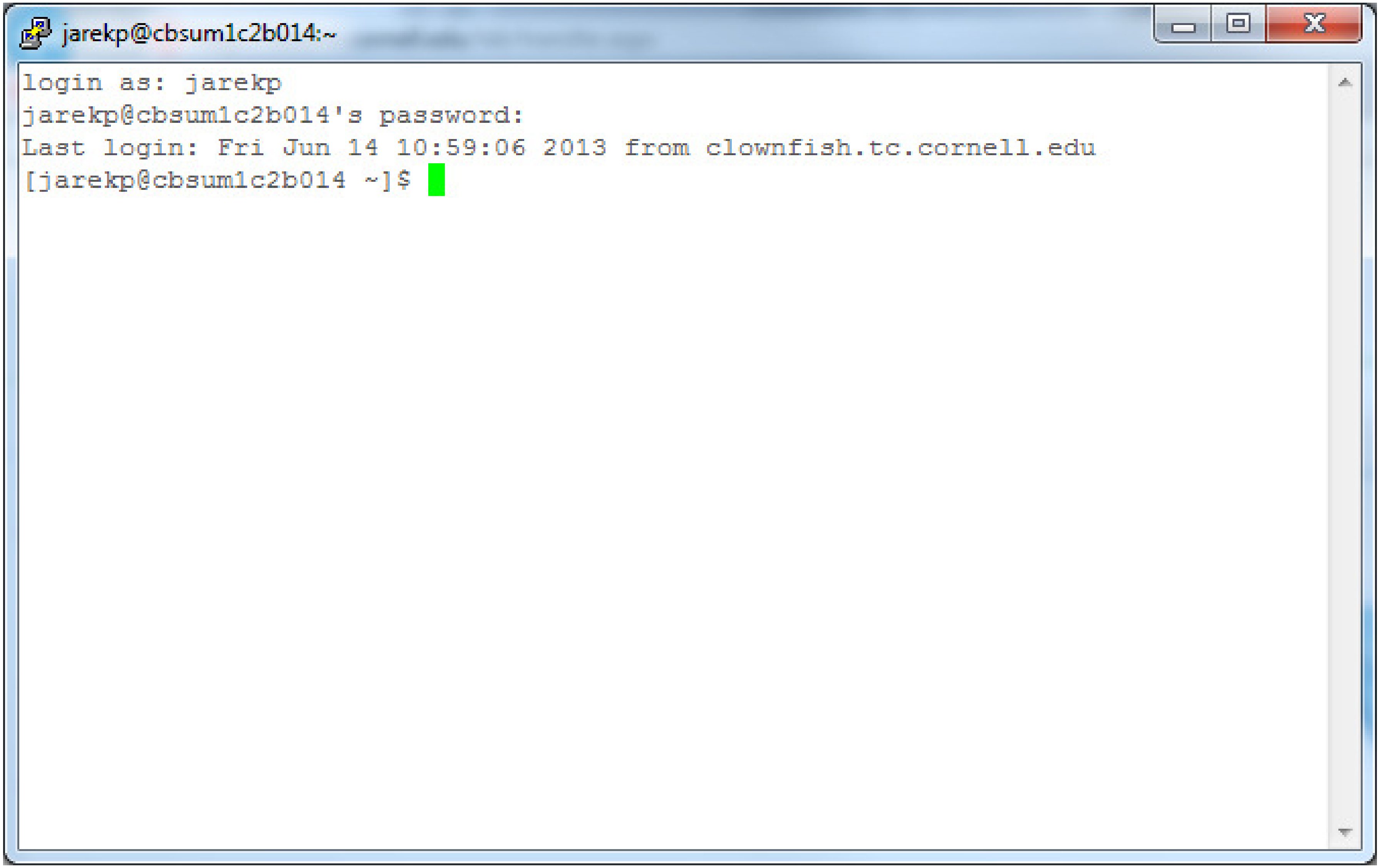
Once logged in to cbsulogin, ssh further to your reserved machine

ssh jarekp@cbsuwrkst3.tc.cornell.edu

Backup login machine is cbsulogin2.tc.cornell.edu

https://cbsu.tc.cornell.edu/lab/doc/BioHPCLabexternal.pdf

Terminal window



```
login as: jarekp
jarekp@cbsum1c2b014's password:
Last login: Fri Jun 14 10:59:06 2013 from clownfish.tc.cornell.edu
[jarekp@cbsum1c2b014 ~]$ 
```

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. We will be using the shell called **bash** (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

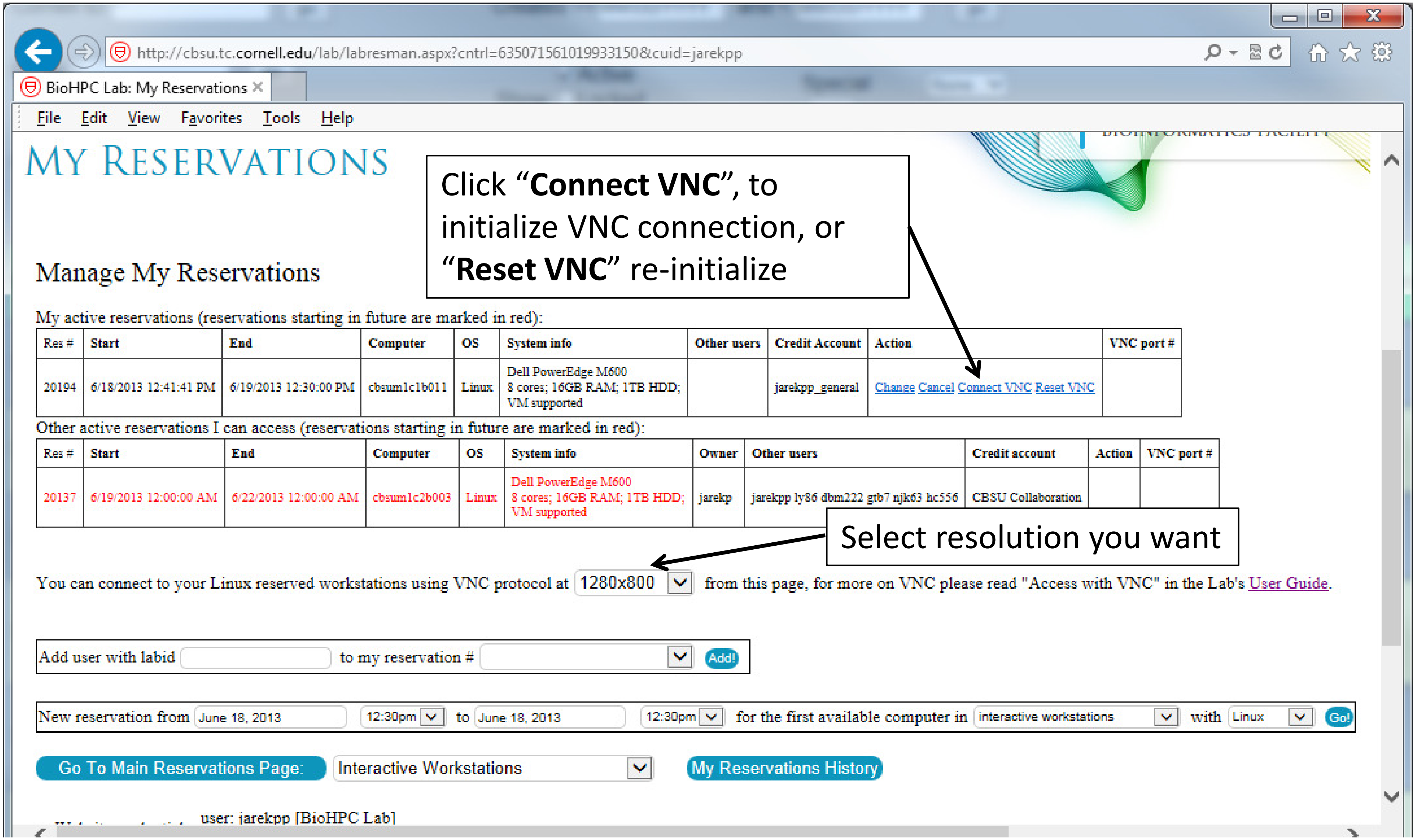
- Find your assigned machine on the list on workshop page
  https://cbsu.tc.cornell.edu/ww/machines.aspx?i=67
- Windows: open PuTTY program, enter your workstation name and connect. Provide your user name and password when prompted.
- Linux or Mac: Open terminal window and type ssh command “ssh labid@workstation.tc.cornell.edu”. Provide your user name and password when prompted.

Connecting to BioHPC Lab

GUI (graphical) connection: VNC

Logging in to a Linux workstation via web browser VNC client (GUI)

In web browser, navigate to 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

Manage My Reservations

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

Res #	Start	End	Computer	OS	System info	Other users	Credit Account	Action	VNC port #
20194	6/18/2013 12:41:41 PM	6/19/2013 12:30:00 PM	cbsumlc1b011	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported		jarekpp_general	Change Cancel Connect VNC Reset VNC	

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

Res #	Start	End	Computer	OS	System info	Owner	Other users	Credit account	Action	VNC port #
20137	6/19/2013 12:00:00 AM	6/22/2013 12:00:00 AM	cbsumlc2b003	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported	jarekp	jarekpp ly86 dbm222 gtb7 njk63 hc556	CBSU Collaboration		

Select resolution you want

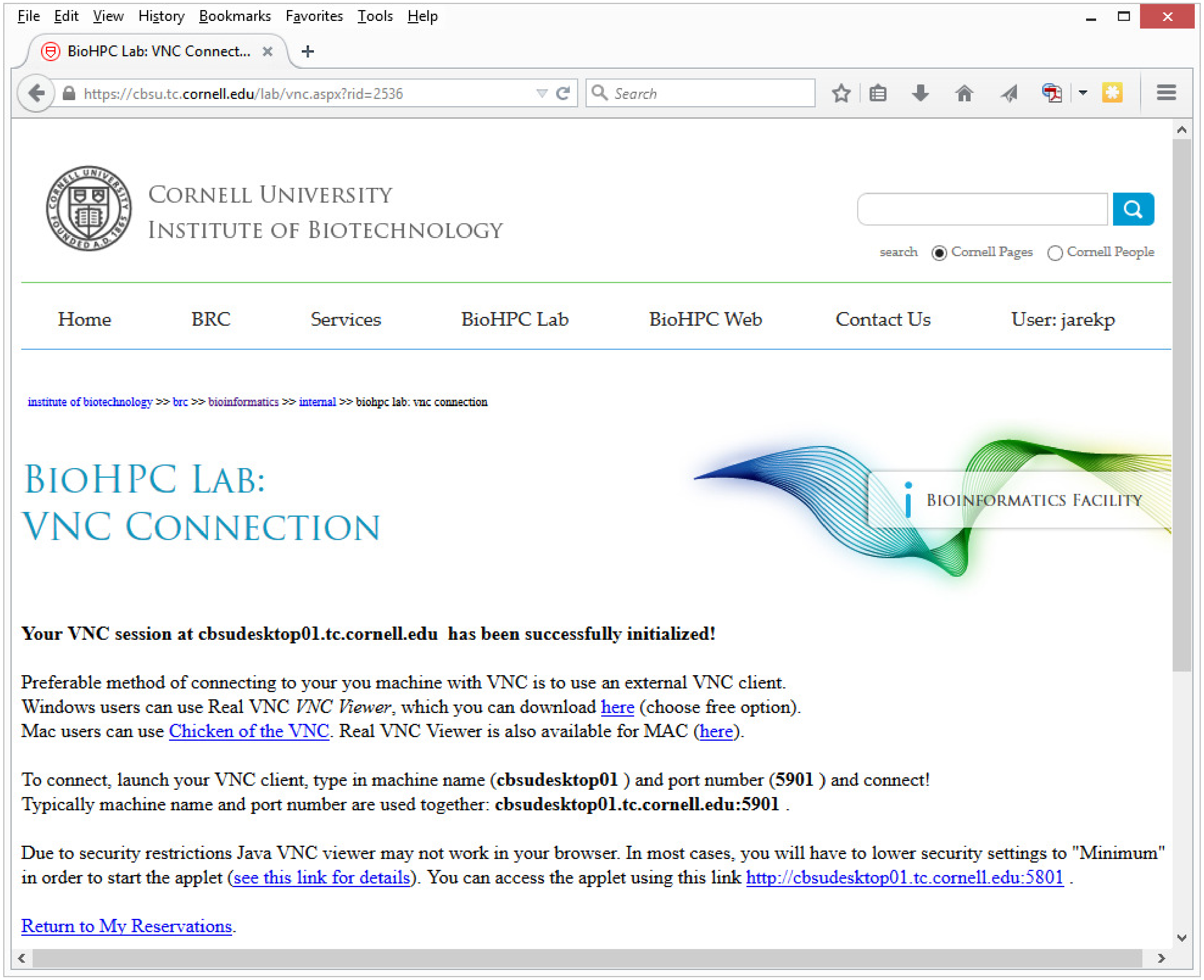
You can connect to your Linux reserved workstations using VNC protocol at 1280x800 from this page, for more on VNC please read "Access with VNC" in the Lab's User Guide.

Add user with labid ______ to my reservation # ______ Add!

New reservation from June 18, 2013 12:30pm to June 18, 2013 12:30pm for the first available computer in interactive workstations with Linux Go!

user: jarekpp [BioHPC Lab]

Logging in to a Linux workstation
(GUI)



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc lab: vnc connection

BIOHPC LAB:
VNC CONNECTION

Your VNC session at cbsudesktop01.tc.cornell.edu has been successfully initialized!

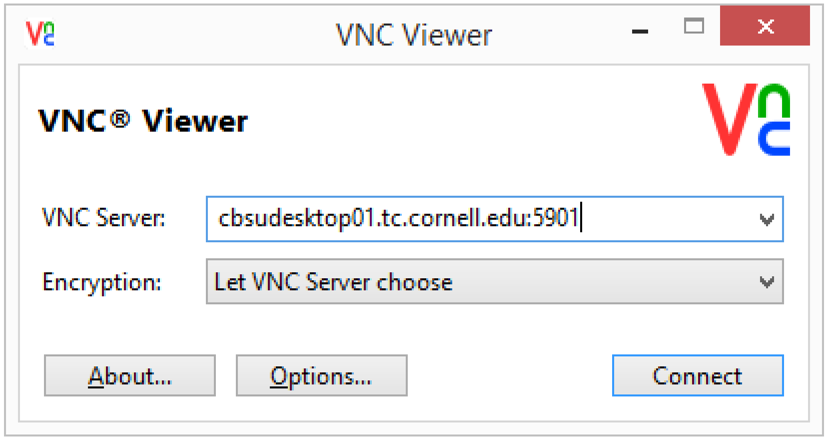
Preferable method of connecting to your you machine with VNC is to use an external VNC client.
Windows users can use Real VNC *VNC Viewer*, which you can download here (choose free option).
Mac users can use Chicken of the VNC. Real VNC Viewer is also available for MAC (here).

To connect, launch your VNC client, type in machine name (**cbsudesktop01**) and port number (**5901**) and connect!
Typically machine name and port number are used together: **cbsudesktop01.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://cbsudesktop01.tc.cornell.edu:5801 .

Return to My Reservations.

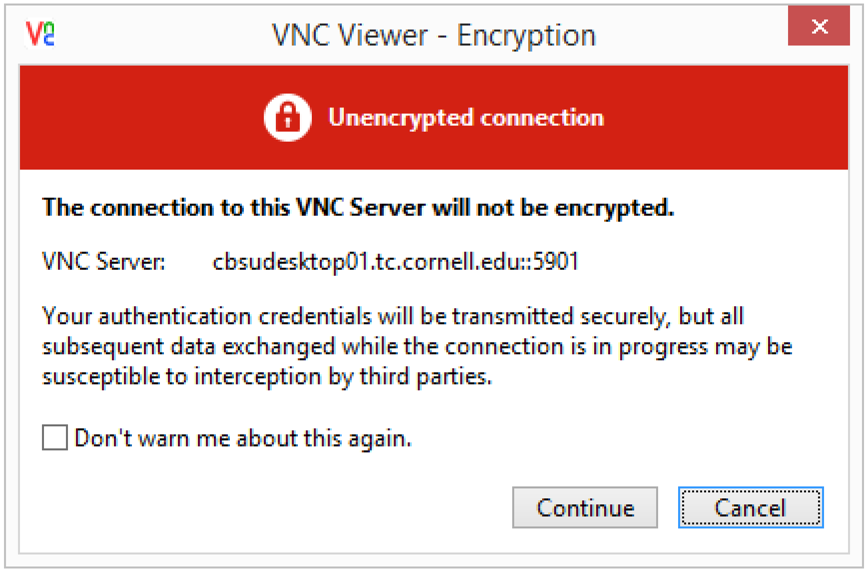
Logging in to a Linux workstation
(GUI)



VNC® Viewer

VNC Server: cbsudesktop01.tc.cornell.edu:5901

Encryption: Let VNC Server choose



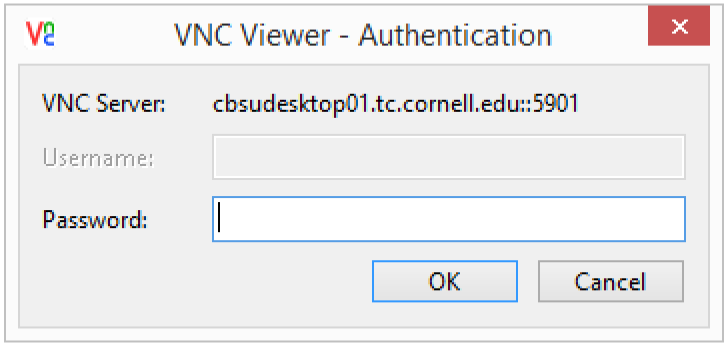
Unencrypted connection

The connection to this VNC Server will not be encrypted.

VNC Server: cbsudesktop01.tc.cornell.edu::5901

Your authentication credentials will be transmitted securely, but all subsequent data exchanged while the connection is in progress may be susceptible to interception by third parties.

☐ Don't warn me about this again.



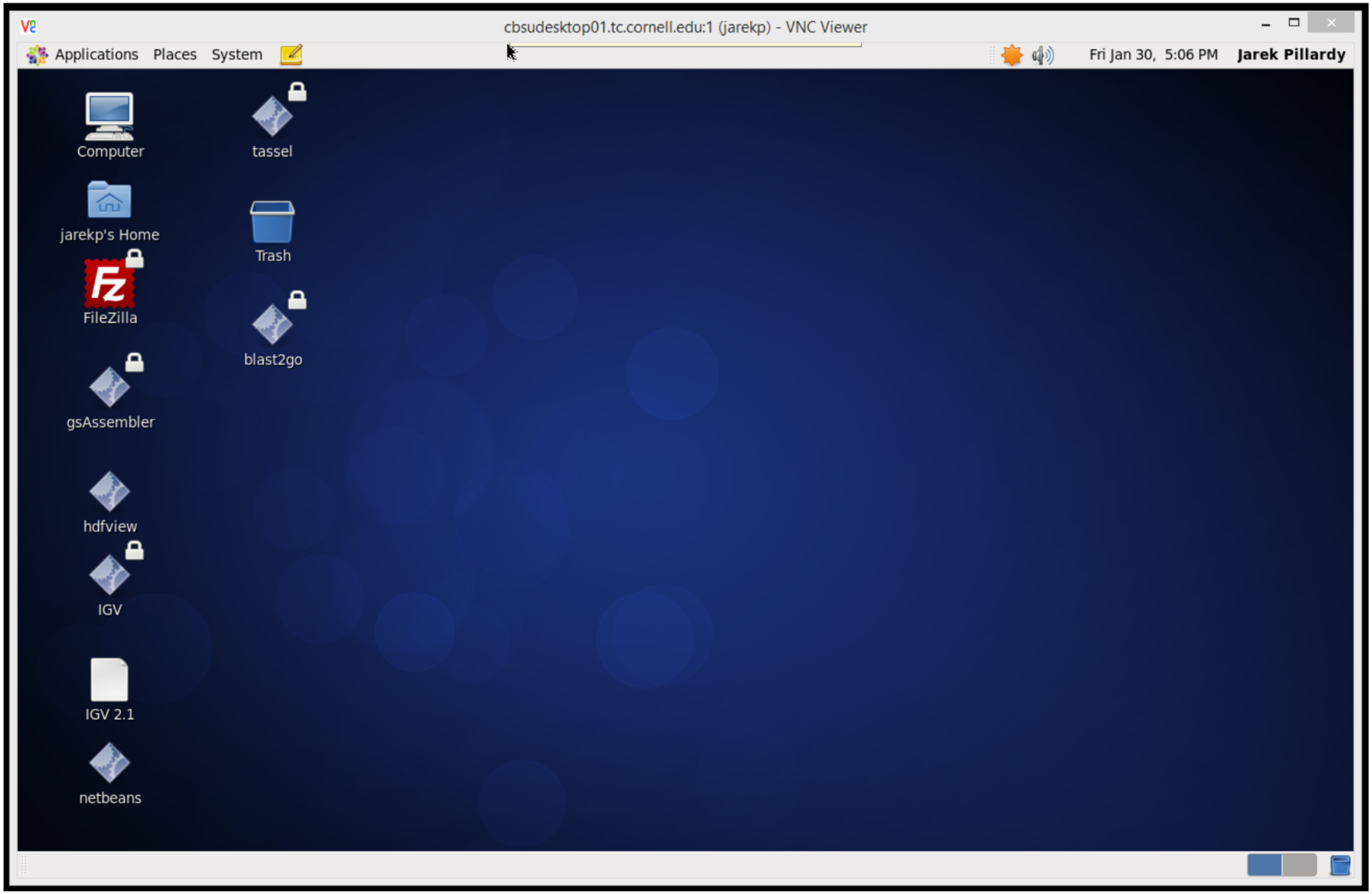
VNC Server: cbsudesktop01.tc.cornell.edu::5901

Username:

Password:

Logging in to a Linux workstation via web browser VNC client
(GUI)

Right-click anywhere within blue 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 Lab: 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 Lab

GUI (graphical) connection: X-Windows

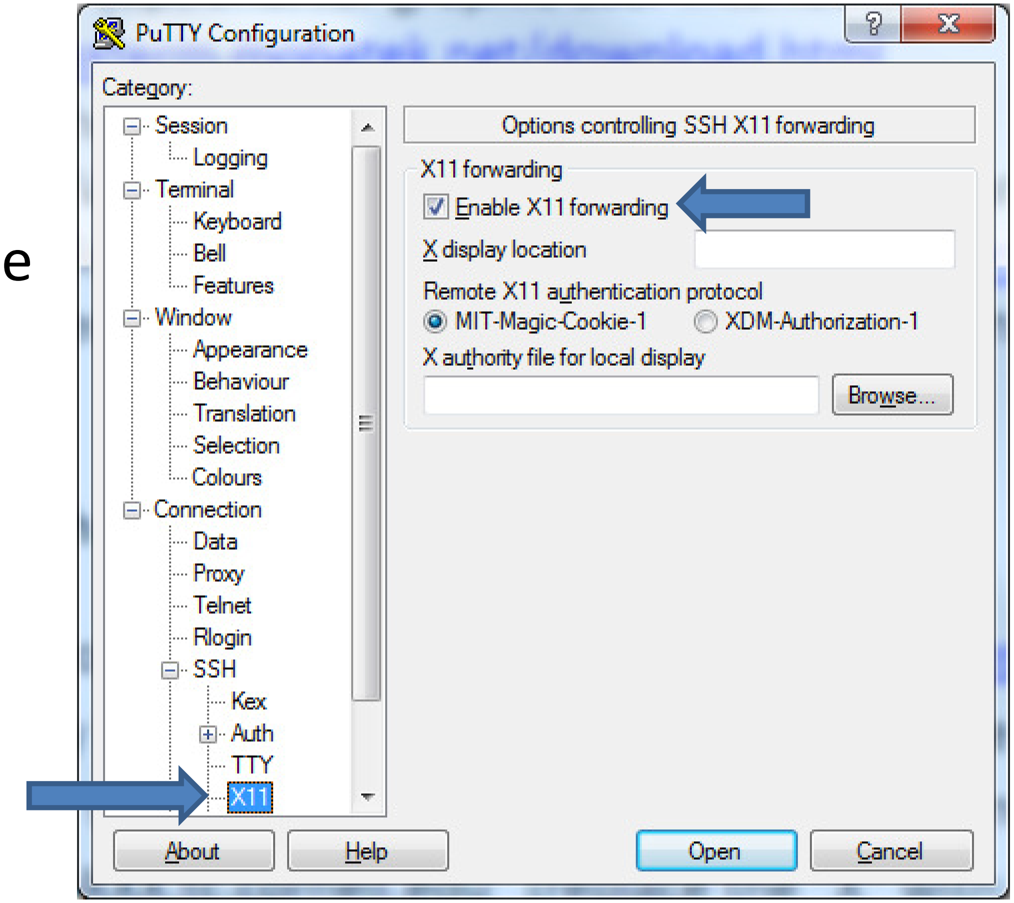
Not persistent – programs will get killed when client disconnects.

Connecting to BioHPC Lab: X-Windows Windows

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

- Start MobaXterm

- Connect to BioHPC Lab machine using PuTTY. Make sure X11 forwarding is enabled. X11 is a synonym for X-Windows



Options controlling SSH X11 forwarding

X11 forwarding

☑ Enable X11 forwarding

X display location

Remote X11 authentication protocol

● MIT-Magic-Cookie-1 ○ XDM-Authorization-1

X authority file for local display

Connecting to BioHPC Lab: 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 Lab machine, but it will display graphics on your local computer.

Connecting to BioHPC Lab: X-Windows
Linux or Mac

- Connect to BioHPC Lab machine using ssh with X11 forwarding :
  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 Lab machine, but it will display graphics on your local computer.

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
- ✖ 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 (secure file transfer protocol)

  Transfer can be done to and from Lab machines.

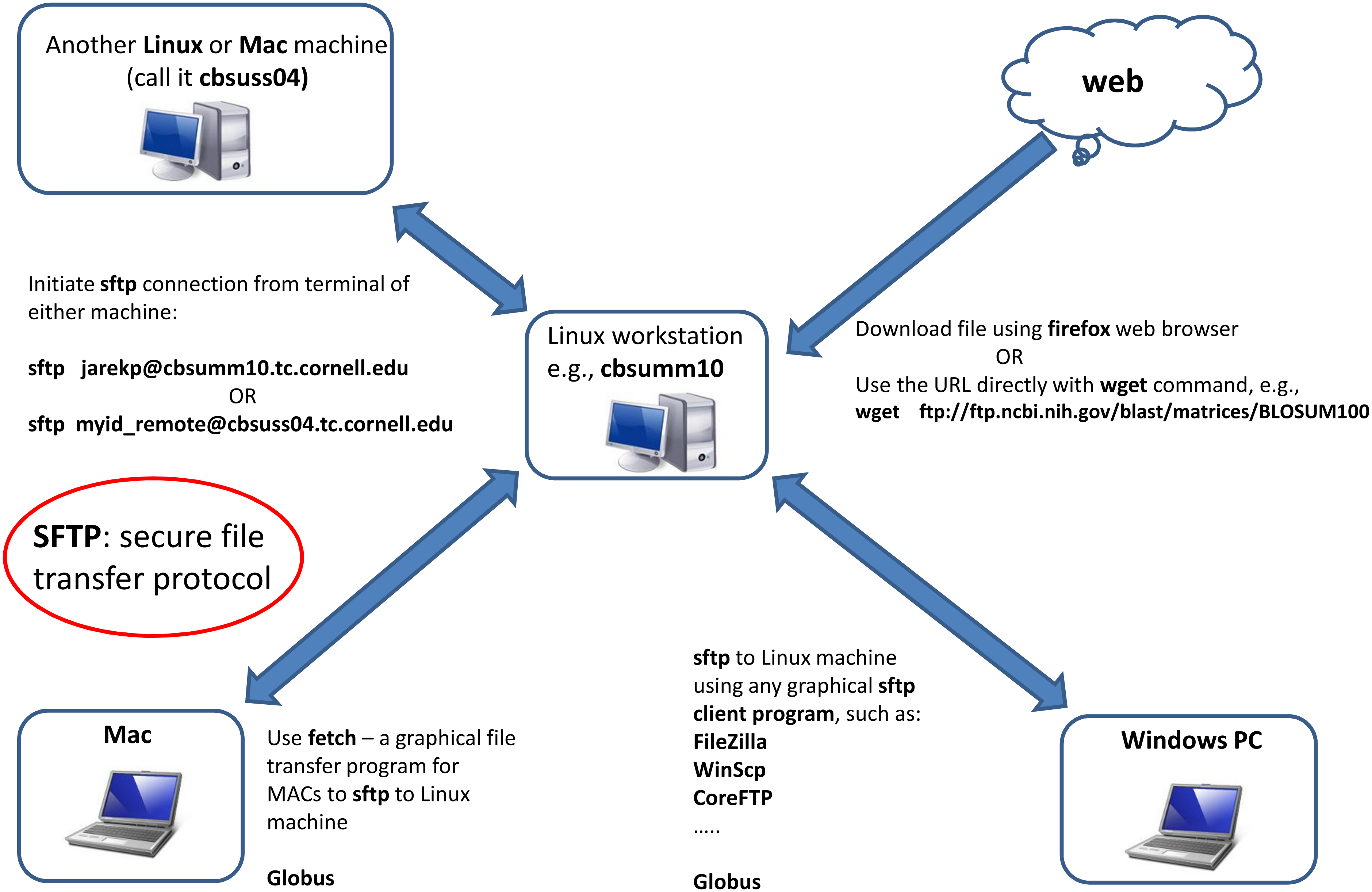
- Globus

  Transfer can be done to and from Lab machines.

- wget

  Transfer can be done to Lab machines only.

File Transfer: summary



Another **Linux** or **Mac** machine
(call it **cbsuss04)**

web

Initiate **sftp** connection from terminal of either machine:

```
sftp   jarekp@cbsumm10.tc.cornell.edu
                    OR
sftp  myid_remote@cbsuss04.tc.cornell.edu
```

Linux workstation
e.g., **cbsumm10**

Download file using **firefox** web browser

OR

Use the URL directly with **wget** command, e.g.,

```
wget    ftp://ftp.ncbi.nih.gov/blast/matrices/BLOSUM100
```

SFTP: secure file transfer protocol

Mac

Use **fetch** – a graphical file transfer program for MACs to **sftp** to Linux machine

Globus

sftp to Linux machine using any graphical **sftp client program**, such as:

FileZilla
WinScp
CoreFTP
…..

Globus

Windows PC

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:**

```
wget    ftp://ftp.ncbi.nih.gov/blast/matrices/BLOSUM100
```

(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

```
wget    ftp://ftp.ncbi.nih.gov/blast/matrices/BLOSUM100
```

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
- **CoreFTP LE**: http://www.coreftp.com/
- **FileZilla** (client): http://filezilla-project.org/

Recommended!

- … 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)

- http://www2.cit.cornell.edu/services/systems_support/filefetch.html#fetchinst
- graphical user interface
- Drag-and-drop functionality

File transfer: sftp

between a lab workstation and another Linux machine

Suppose we want to transfer a file from **cbsuss04.tc.cornell.edu** (another Linux machine; substitute “your” Linux machine here) and **cbsuwrkst2** lab workstation.

Option 1: when logged in to **cbsuwrkst2**, sftp to **cbsuss04** by running the following commands:

cd /workdir/jarekp	*(this is where we want the file to be on cbsuwrkst2)*
sftp jarekp@cbsuss04.tc.cornell.edu	*(instead of “jarekp”, use your own user name on cbsuss04; you will be asked for password)*
cd /data/jarekp/blat	*(on cbsuss04, go to the directory where the file is)*
get e_coli_1000_1.fq	*(transfer, or “get” the file from cbsuss04)*
quit	*(exit sftp client and disconnect from cbsuss04 – we are back on cbsuwrkst2)*

Option 2: when logged in to **cbsuss04**, sftp to **cbsuwrkst2** by running the following commands:

cd /data/jarekp/blat	*(this is where the file is on cbsuss04)*
sftp jarekp@cbsuwrkst2.tc.cornell.edu	*(instead of “jarekp”, use your own user name on cbsuss04; you will be asked your lab password)*
cd /workdir/jarekp	*(on cbsuwrkst2, go to the directory where the file is supposed to be stored)*
put e_coli_1000_1.fq	*(transfer, or “put” the file on cbsuwrkst2)*
quit	*(exit sftp client and disconnect from cbsuwrkst2– we are back on cbsuss04)*

Option 3: **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 Lab Globus endpoint server is **biohpc#cbsulogin**.

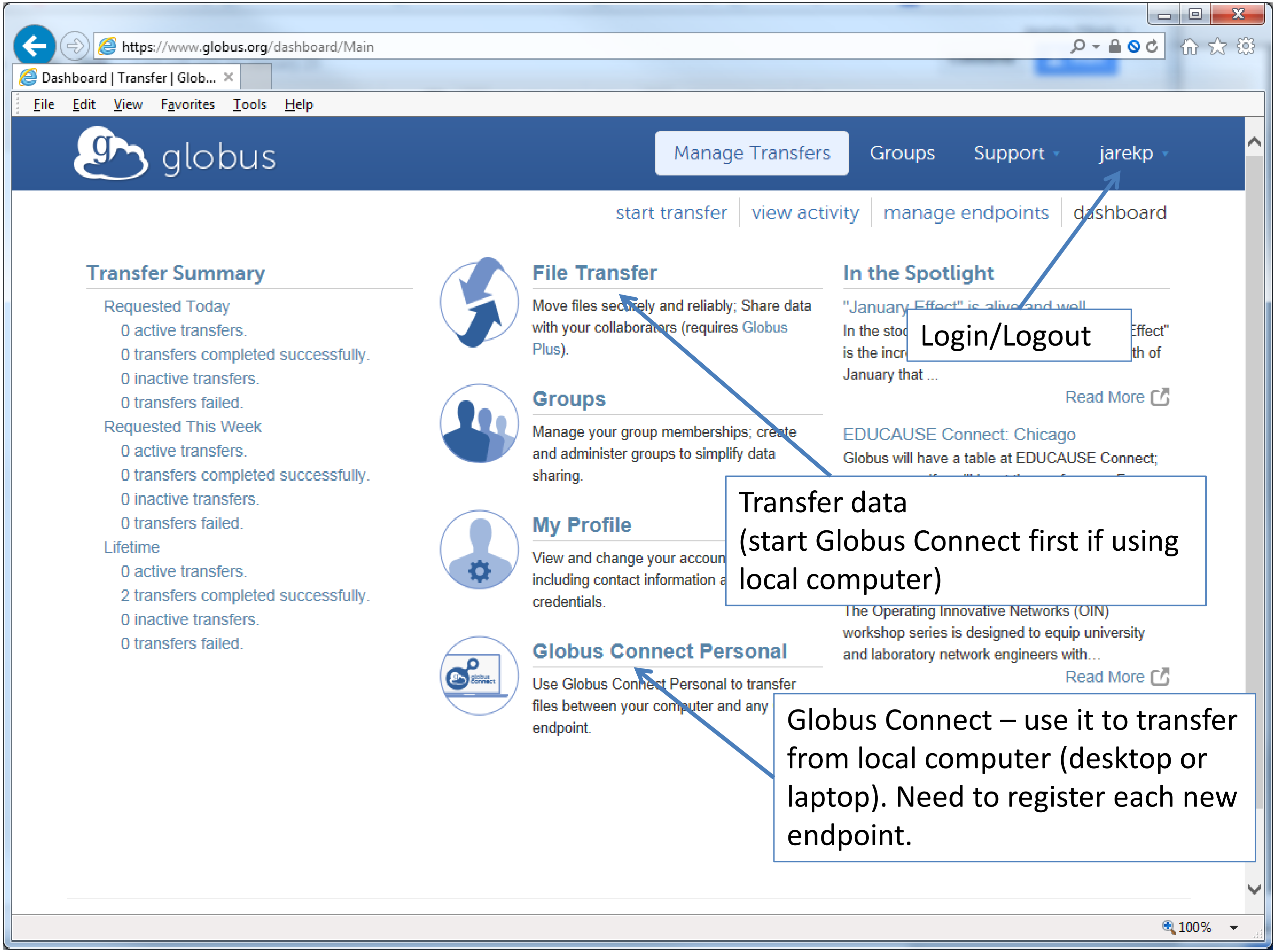
You can transfer data to another server or to your computer (using Globus Connect client).

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 Lab 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 Summary

Requested Today
- 0 active transfers.
- 0 transfers completed successfully.
- 0 inactive transfers.
- 0 transfers failed.

Requested This Week
- 0 active transfers.
- 0 transfers completed successfully.
- 0 inactive transfers.
- 0 transfers failed.

Lifetime
- 0 active transfers.
- 2 transfers completed successfully.
- 0 inactive transfers.
- 0 transfers failed.

File Transfer

Move files securely and reliably; Share data with your collaborators (requires Globus Plus).

Groups

Manage your group memberships; create and administer groups to simplify data sharing.

My Profile

View and change your accoun including contact information a credentials.

Globus Connect Personal

Use Globus Connect Personal to transfer files between your computer and any endpoint.

In the Spotlight

"January Effect" is alive and well

In the stoc Effect" is the incr th of January that ...

Read More

EDUCAUSE Connect: Chicago

Globus will have a table at EDUCAUSE Connect;

The Operating Innovative Networks (OIN) workshop series is designed to equip university and laboratory network engineers with...

Read More

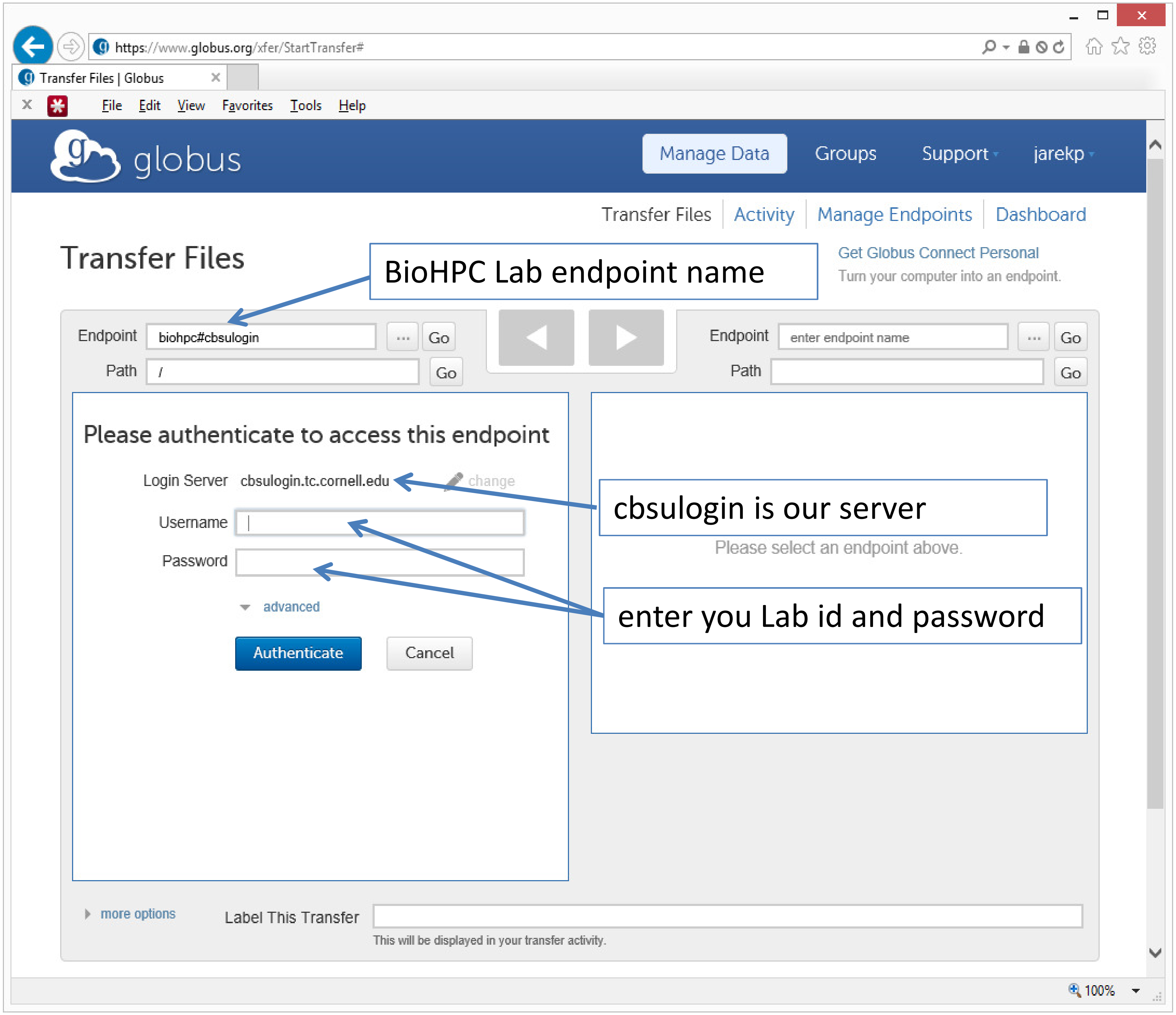
Login/Logout

Transfer data
(start Globus Connect first if using local computer)

Globus Connect – use it to transfer from local computer (desktop or laptop). Need to register each new 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 Lab 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

BioHPC Lab endpoint name

Endpoint: biohpc#cbsulogin

Path: /

Please authenticate to access this endpoint

Login Server: cbsulogin.tc.cornell.edu

Username:

Password:

cbsulogin is our server

Please select an endpoint above.

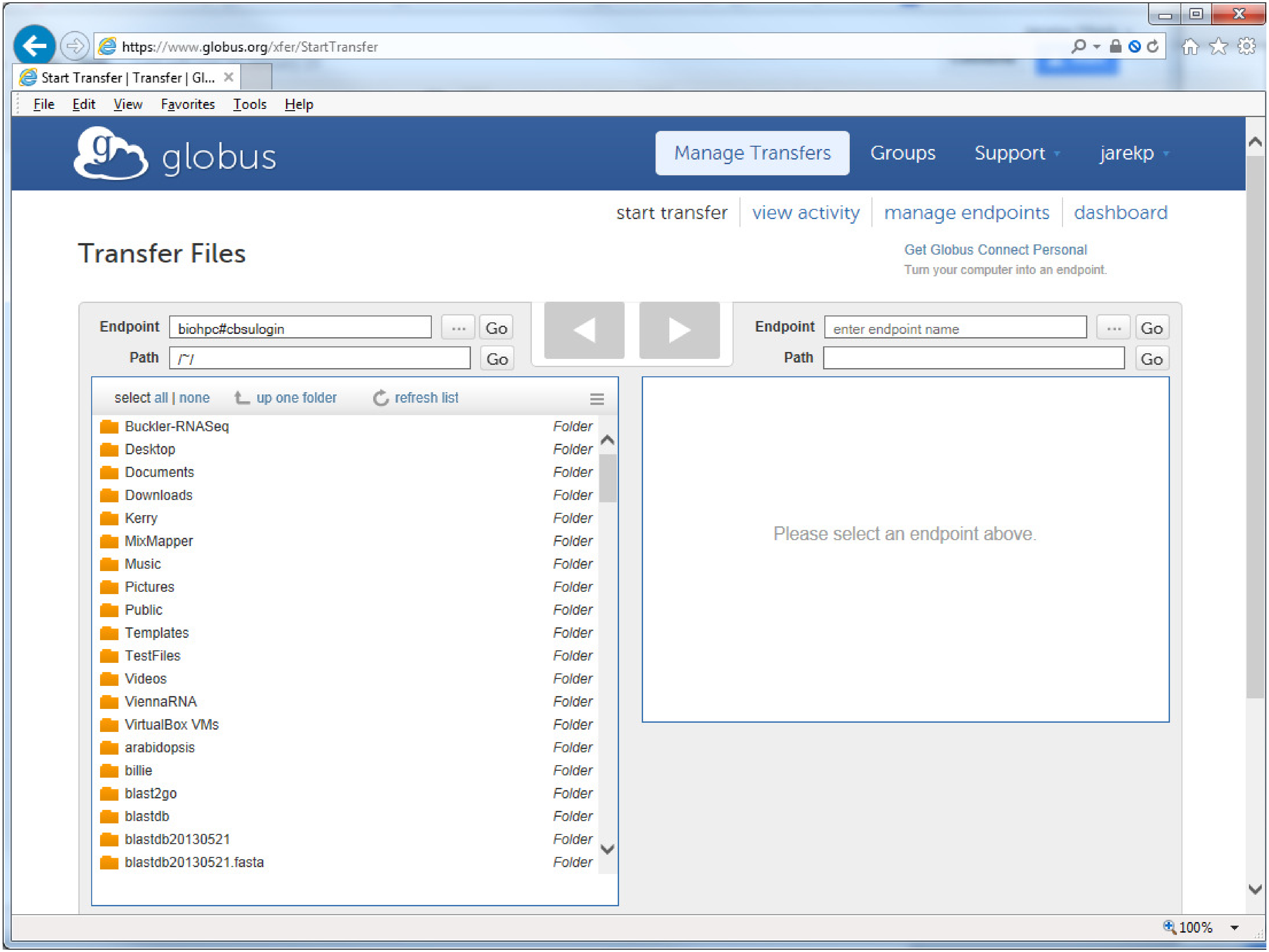
enter you Lab id and password

Label This Transfer

This will be displayed in your transfer activity.

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 Lab 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

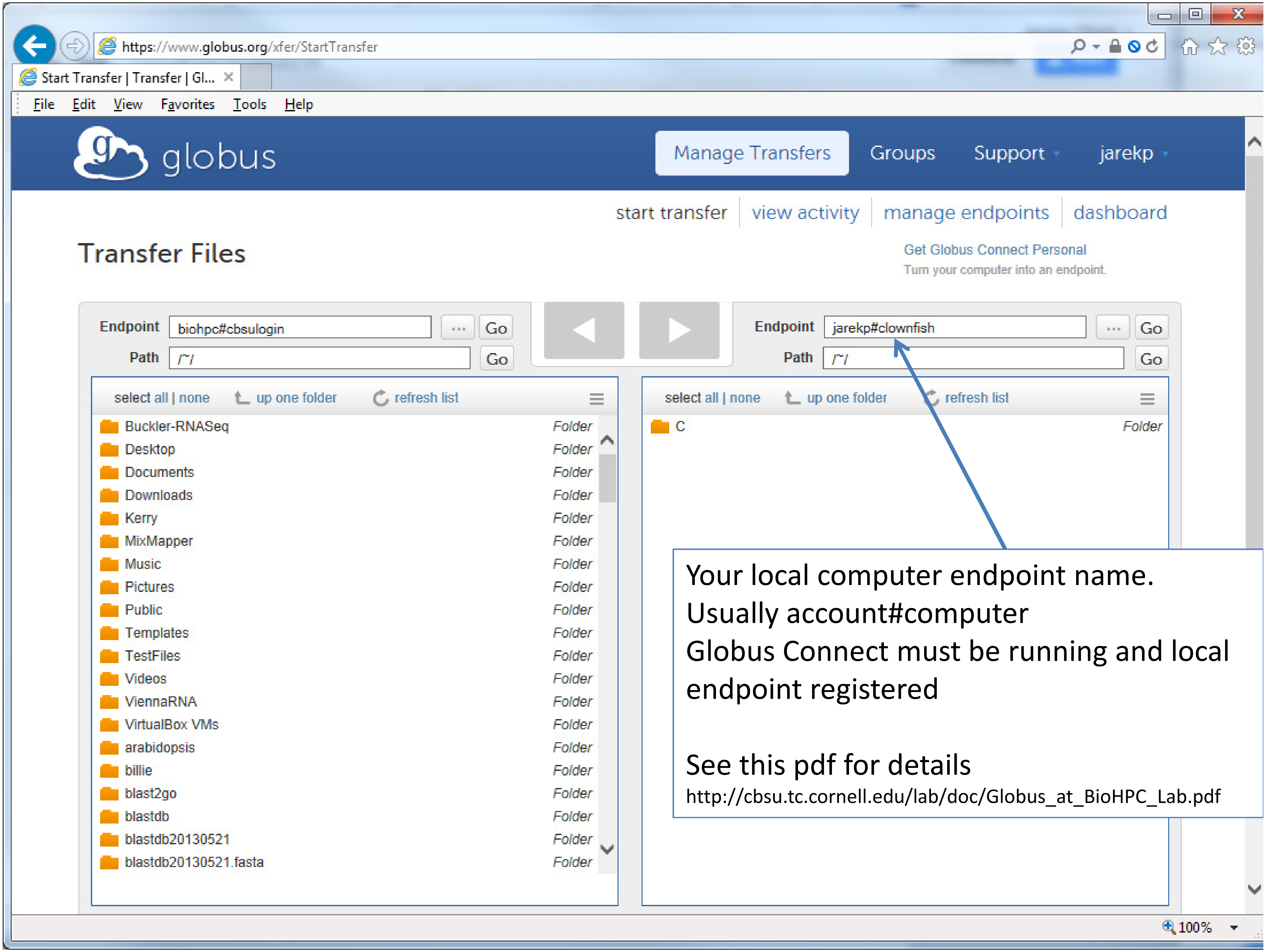
Get Globus Connect Personal
Turn your computer into an endpoint.

Buckler-RNASeq	Folder
Desktop	Folder
Documents	Folder
Downloads	Folder
Kerry	Folder
MixMapper	Folder
Music	Folder
Pictures	Folder
Public	Folder
Templates	Folder
TestFiles	Folder
Videos	Folder
ViennaRNA	Folder
VirtualBox VMs	Folder
arabidopsis	Folder
billie	Folder
blast2go	Folder
blastdb	Folder
blastdb20130521	Folder
blastdb20130521.fasta	Folder

Please select an endpoint above.

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 Lab 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).



Endpoint biohpc#cbsulogin
Path /~/

Endpoint jarekp#clownfish
Path /~/

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

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!*

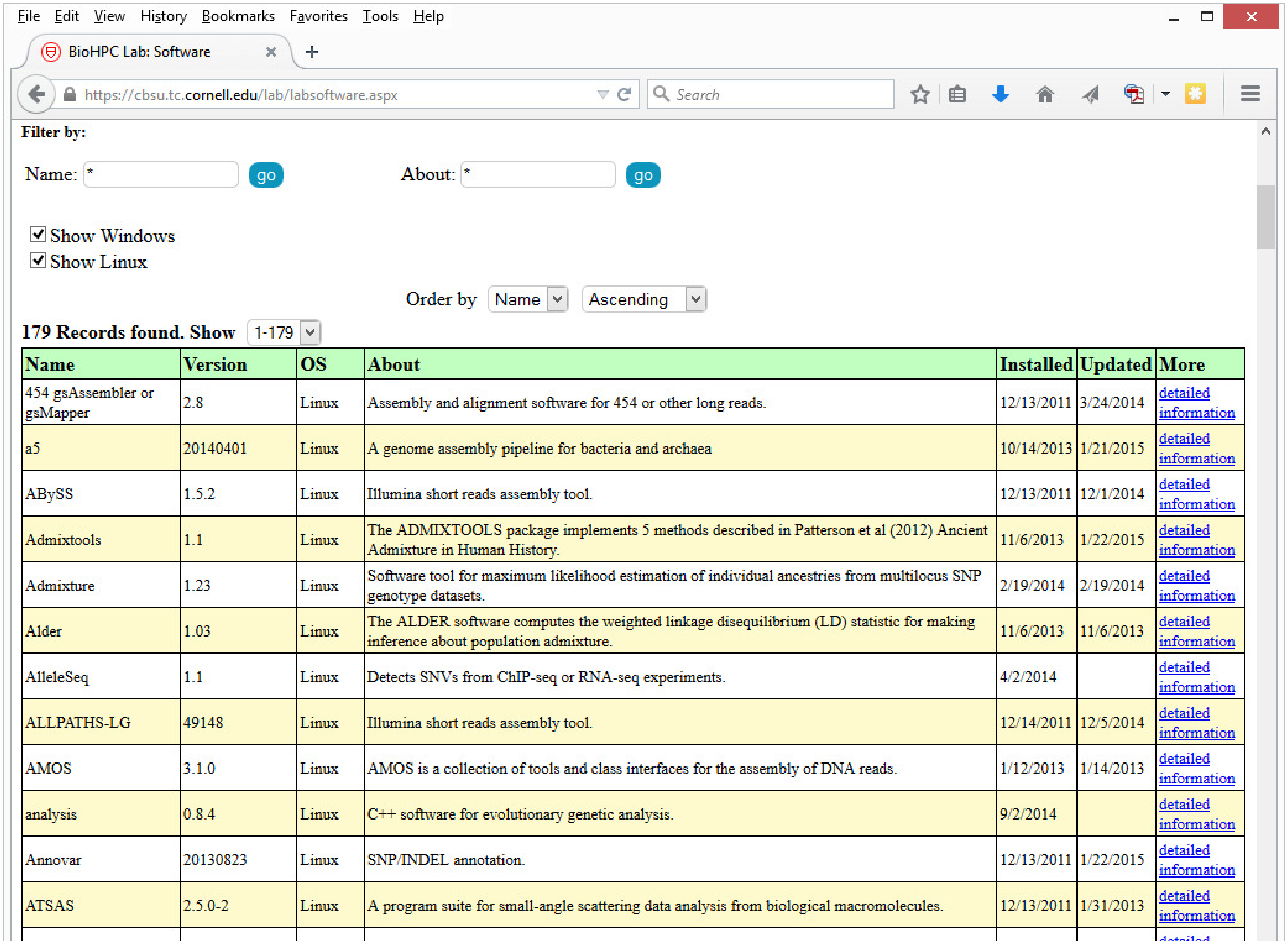
Software

https://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.

Software



Filter by:

Name: *

About: *

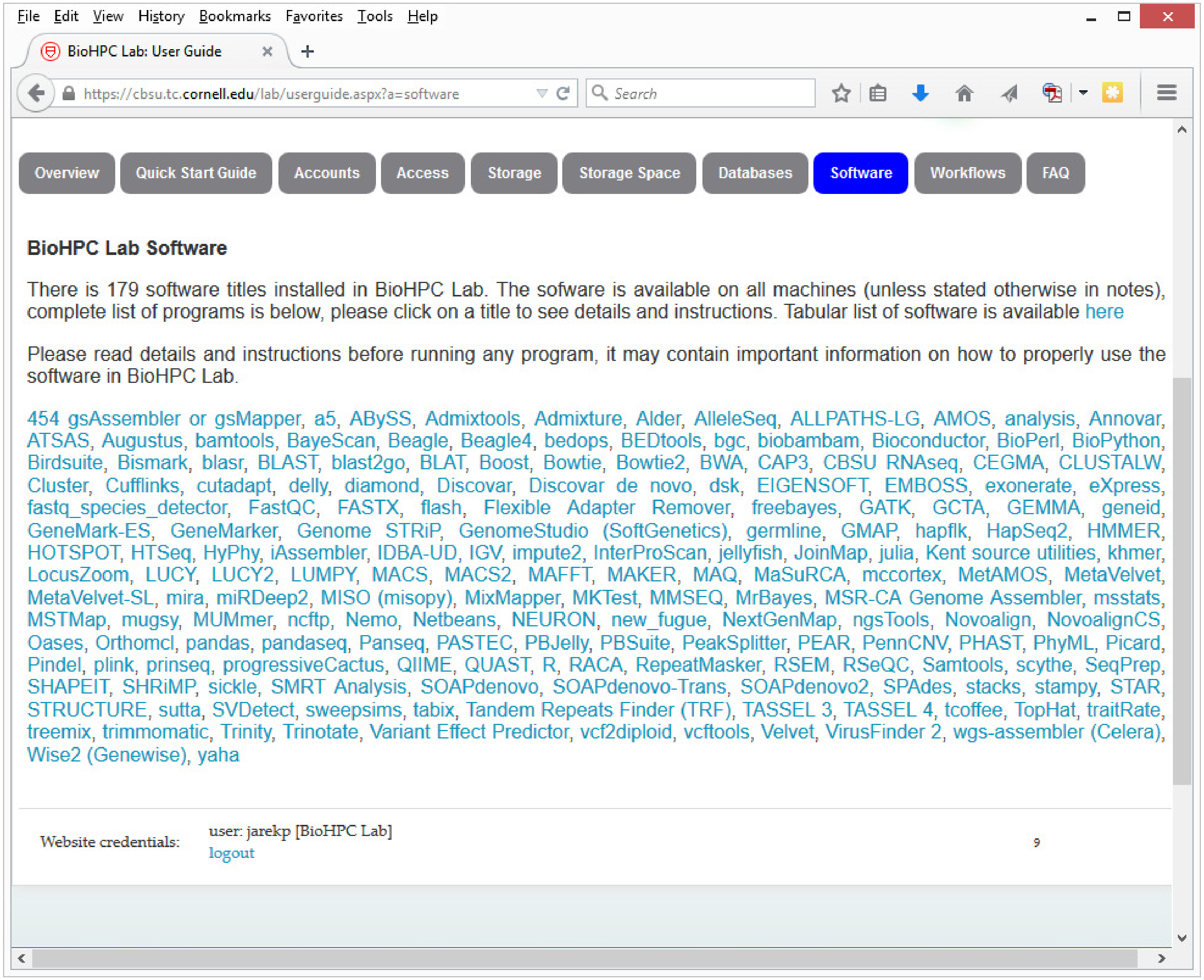
☑ Show Windows
☑ Show Linux

Order by Name Ascending

179 Records found. Show 1-179

Name	Version	OS	About	Installed	Updated	More
454 gsAssembler or gsMapper	2.8	Linux	Assembly and alignment software for 454 or other long reads.	12/13/2011	3/24/2014	detailed information
a5	20140401	Linux	A genome assembly pipeline for bacteria and archaea	10/14/2013	1/21/2015	detailed information
ABySS	1.5.2	Linux	Illumina short reads assembly tool.	12/13/2011	12/1/2014	detailed information
Admixtools	1.1	Linux	The ADMIXTOOLS package implements 5 methods described in Patterson et al (2012) Ancient Admixture in Human History.	11/6/2013	1/22/2015	detailed information
Admixture	1.23	Linux	Software tool for maximum likelihood estimation of individual ancestries from multilocus SNP genotype datasets.	2/19/2014	2/19/2014	detailed information
Alder	1.03	Linux	The ALDER software computes the weighted linkage disequilibrium (LD) statistic for making inference about population admixture.	11/6/2013	11/6/2013	detailed information
AlleleSeq	1.1	Linux	Detects SNVs from ChIP-seq or RNA-seq experiments.	4/2/2014		detailed information
ALLPATHS-LG	49148	Linux	Illumina short reads assembly tool.	12/14/2011	12/5/2014	detailed information
AMOS	3.1.0	Linux	AMOS is a collection of tools and class interfaces for the assembly of DNA reads.	1/12/2013	1/14/2013	detailed information
analysis	0.8.4	Linux	C++ software for evolutionary genetic analysis.	9/2/2014		detailed information
Annovar	20130823	Linux	SNP/INDEL annotation.	12/13/2011	1/22/2015	detailed information
ATSAS	2.5.0-2	Linux	A program suite for small-angle scattering data analysis from biological macromolecules.	12/13/2011	1/31/2013	detailed information

Software



BioHPC Lab Software

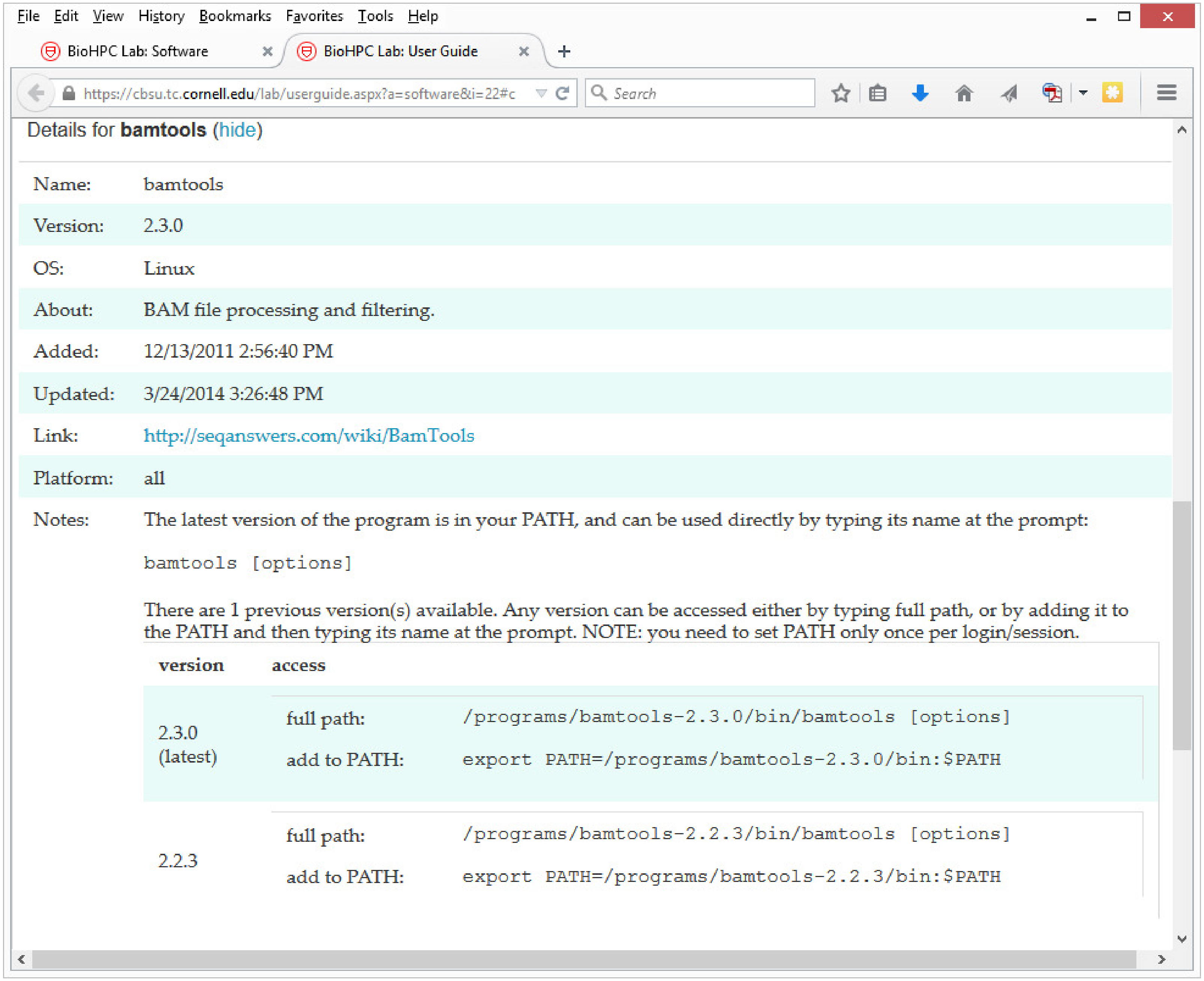
There is 179 software titles installed in BioHPC Lab. The sofware is available on all machines (unless stated otherwise in notes), complete list of programs is below, please click on a title to see details and instructions. Tabular list of software is available here

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 or gsMapper, a5, ABySS, Admixtools, Admixture, Alder, AlleleSeq, ALLPATHS-LG, AMOS, analysis, Annovar, ATSAS, Augustus, bamtools, BayeScan, Beagle, Beagle4, bedops, BEDtools, bgc, biobambam, Bioconductor, BioPerl, BioPython, Birdsuite, Bismark, blasr, BLAST, blast2go, BLAT, Boost, Bowtie, Bowtie2, BWA, CAP3, CBSU RNAseq, CEGMA, CLUSTALW, Cluster, Cufflinks, cutadapt, delly, diamond, Discovar, Discovar de novo, dsk, EIGENSOFT, EMBOSS, exonerate, eXpress, fastq_species_detector, FastQC, FASTX, flash, Flexible Adapter Remover, freebayes, GATK, GCTA, GEMMA, geneid, GeneMark-ES, GeneMarker, Genome STRiP, GenomeStudio (SoftGenetics), germline, GMAP, hapflk, HapSeq2, HMMER, HOTSPOT, HTSeq, HyPhy, iAssembler, IDBA-UD, IGV, impute2, InterProScan, jellyfish, JoinMap, julia, Kent source utilities, khmer, LocusZoom, LUCY, LUCY2, LUMPY, MACS, MACS2, MAFFT, MAKER, MAQ, MaSuRCA, mccortex, MetAMOS, MetaVelvet, MetaVelvet-SL, mira, miRDeep2, MISO (misopy), MixMapper, MKTest, MMSEQ, MrBayes, MSR-CA Genome Assembler, msstats, MSTMap, mugsy, MUMmer, ncftp, Nemo, Netbeans, NEURON, new_fugue, NextGenMap, ngsTools, Novoalign, NovoalignCS, Oases, Orthomcl, pandas, pandaseq, Panseq, PASTEC, PBJelly, PBSuite, PeakSplitter, PEAR, PennCNV, PHAST, PhyML, Picard, Pindel, plink, prinseq, progressiveCactus, QIIME, QUAST, R, RACA, RepeatMasker, RSEM, RSeQC, Samtools, scythe, SeqPrep, SHAPEIT, SHRiMP, sickle, SMRT Analysis, SOAPdenovo, SOAPdenovo-Trans, SOAPdenovo2, SPAdes, stacks, stampy, STAR, STRUCTURE, sutta, SVDetect, sweepsims, tabix, Tandem Repeats Finder (TRF), TASSEL 3, TASSEL 4, tcoffee, TopHat, traitRate, treemix, trimmomatic, Trinity, Trinotate, Variant Effect Predictor, vcf2diploid, vcftools, Velvet, VirusFinder 2, wgs-assembler (Celera), Wise2 (Genewise), yaha

Website credentials: user: jarekp [BioHPC Lab]
logout

Software



Details for **bamtools** (hide)

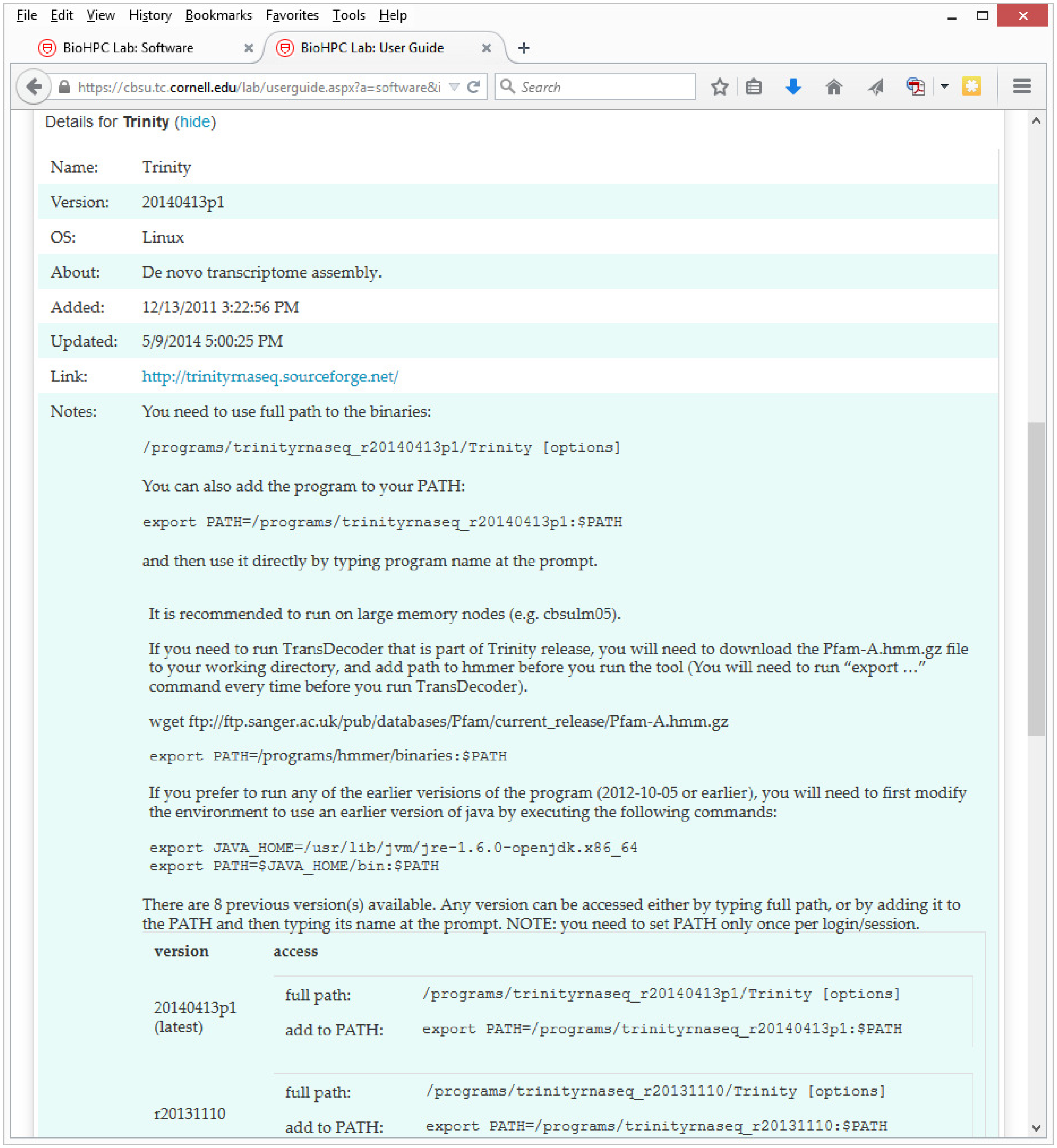
Name:	bamtools
Version:	2.3.0
OS:	Linux
About:	BAM file processing and filtering.
Added:	12/13/2011 2:56:40 PM
Updated:	3/24/2014 3:26:48 PM
Link:	http://seqanswers.com/wiki/BamTools
Platform:	all

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 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.

version	access
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** (hide)

Name:	Trinity
Version:	20140413p1
OS:	Linux
About:	De novo transcriptome assembly.
Added:	12/13/2011 3:22:56 PM
Updated:	5/9/2014 5:00:25 PM
Link:	http://trinityrnaseq.sourceforge.net/

Notes:

You need to use full path to the binaries:

```
/programs/trinityrnaseq_r20140413p1/Trinity [options]
```

You can also add the program to your PATH:

```
export PATH=/programs/trinityrnaseq_r20140413p1:$PATH
```

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

It is recommended to run on large memory nodes (e.g. cbsulm05).

If you need to run TransDecoder that is part of Trinity release, you will need to download the Pfam-A.hmm.gz file to your working directory, and add path to hmmer 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/Pfam/current_release/Pfam-A.hmm.gz

```
export PATH=/programs/hmmer/binaries:$PATH
```

If you prefer to run any of the earlier verisions of the program (2012-10-05 or earlier), you will need to first modify the environment to use an earlier version of java by executing the following commands:

```
export JAVA_HOME=/usr/lib/jvm/jre-1.6.0-openjdk.x86_64
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.

version	**access**	
20140413p1 (latest)	full path:	`/programs/trinityrnaseq_r20140413p1/Trinity [options]`
	add to PATH:	`export PATH=/programs/trinityrnaseq_r20140413p1:$PATH`
r20131110	full path:	`/programs/trinityrnaseq_r20131110/Trinity [options]`
	add to PATH:	`export PATH=/programs/trinityrnaseq_r20131110:$PATH`

Software

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.

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!*