Introduction to BioHPC Cloud

BioHPC Cloud Workshop

Jaroslaw Pillardy

Bioinformatics Facility
Institute of Biotechnology
Cornell University

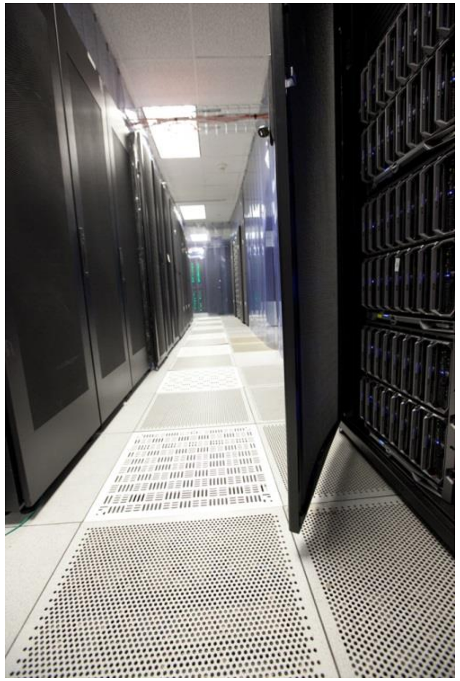
http://biohpc.cornell.edu/lab/lab.aspx

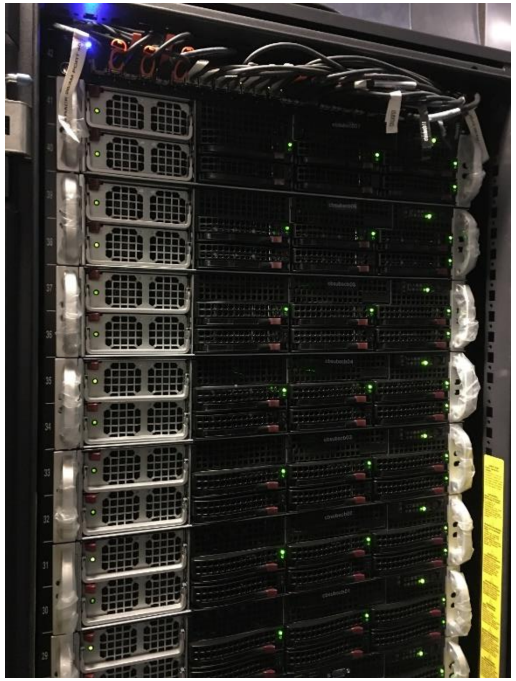
http://biohpc.cornell.edu/lab/doc/Introduction_to_BioHPC_Cloud_v10.pdf

BioHPC Cloud

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

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





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 most of the time, it is cheaper to host. You can still avoid problems related to maintaining your own hardware, software and resource allocation by hosting the machine with BioHPC. Access is restricted to your group, but otherwise all BioHPC resources are available on the server.

BioHPC Cloud: storage

- 1,440 TB of networked storage available in two volumes: 1,207 TB Lustre volume and 233 TB Gluster volume.
- Very robust and scalable architecture: cluster storage based on RAID6/RAID7 servers connected by Lustre or Gluster, new servers can be added to expand storage.
- Very affordable! $98 per TB per year
- Users get free storage allocations with the accounts

BioHPC Cloud: Backup service

- Users can choose backup options using BioHPC website (what to backup, how many versions etc)
- Backup servers (currently 459TB) are in different building (Weill Hall) for additional safety
- Backup service is NOT part of storage service. You can get storage, or backup or both, but you need to configure them separately. This way you only pay for what you need.

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, which depends on possible usage level and time investment required to deploy.
- 774 titles as of 8/28/2020
- Common genomic data is available locally in the Cloud: sequence and annotation databases, preformatted for common programs

http://biohpc.cornell.edu/lab/labsoftware.aspx

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	24-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)

BioHPC Cloud hardware: what is what

- CPU: an integrated circuit (a “chip”) containing computational hardware. May be more than one per server, typically 2-4.
- Core: a subunit of CPU, a part of CPU capable of independent computing. It shares CPU communication infrastructure and internal memory, but not computing hardware.
- Hyperthreaded Core (sometimes Thread): a CPU core subunit that shares parts of computing hardware (FPU etc) with another similar subunit. Typically there are two hyperthreaded cores per core.

BioHPC Cloud hardware infrastructure

"large gen2" servers

2	96-core, 512GB RAM, 12TB HDD, 1TB SSD (cbsulm12-13)
4	112-core, 512GB RAM, 12TB HDD, 1TB SSD (cbsulm14-17)
4	80-core, 512GB RAM, 10TB HDD, (cbsulm21-22,cbsulm27-28)
7	88-core, 512GB RAM, 7TB NVMe SSD (cbsulm18-20,cbsulm23-26)

"extra large" servers

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

"gpu-equipped gen2" servers

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

BioHPC Cloud hardware infrastructure

“networked storage: total 1,440 TB available in two volumes

233 TB Gluster cluster (4 servers)
1,207 TB Lustre cluster (10 servers)

login machines (cbsulogin, cbsulogin2, cbsulogin3)
3 12-core, 64-128GB RAM

BioHPC servers – big picture

Infrastructure servers

Lustre, Gluster, management

No direct user access
No reservations
Restricted network access

Rental servers

general, medium, large, extra large, GPU

Users must make reservations
Any BioHPC user can make reservation
Local storage cleaned after reservation ends
Cornell campus network only

Hosted servers

Purchased by groups or departments

Users must be added to appropriate groups
Only users approved by PIs can be added
Local storage persistent – no cleaning unless requested by owners
Cornell campus network only

Login servers

For data transfers and indirect access

Any BioHPC user can login
No local storage
No computing allowed
Open to the world network

Accessing BioHPC servers from outside Cornell campus

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

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

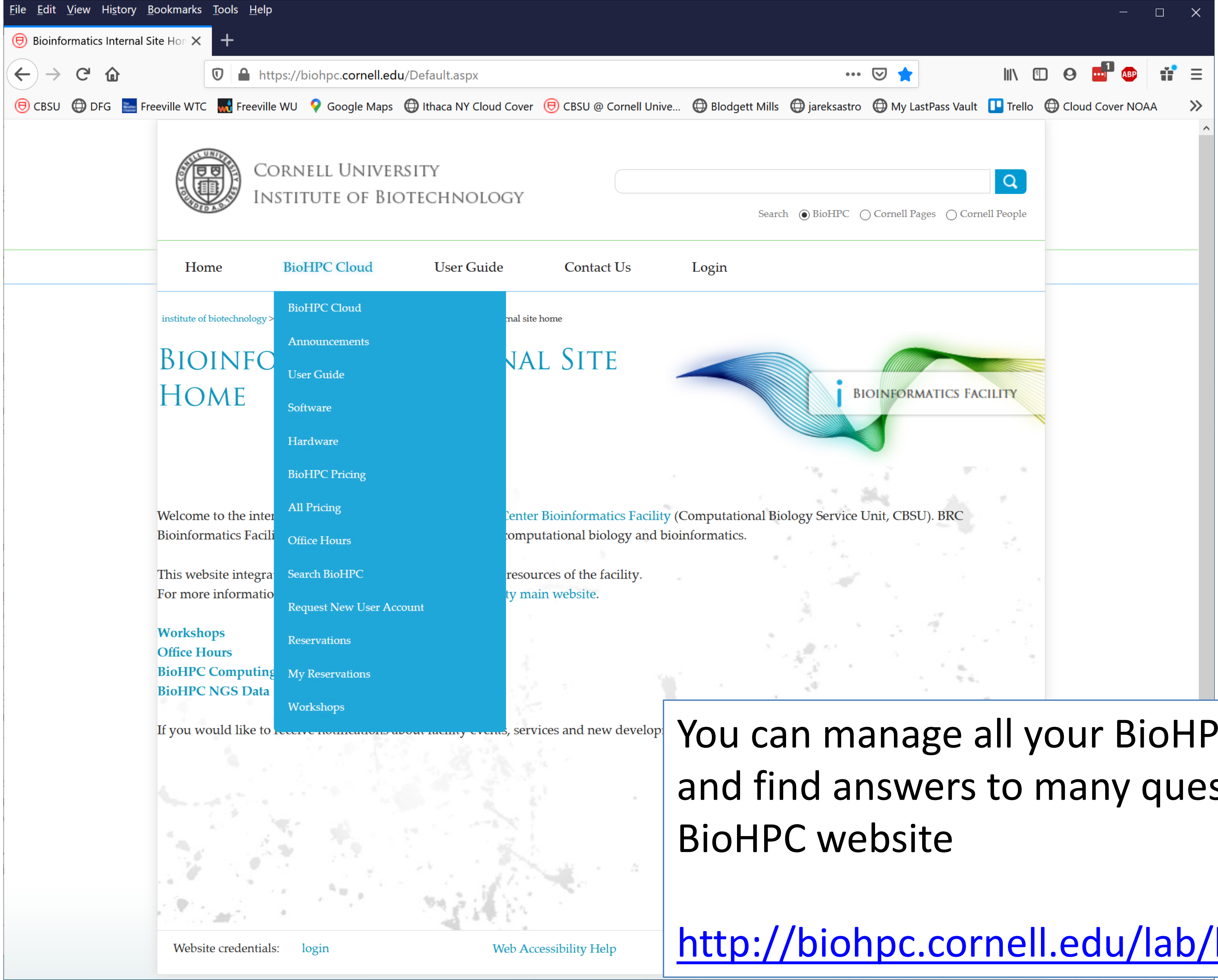
- Preferred method of outside access is Cornell VPN, available for all Cornelians from CIT

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

- Staff or faculty of Weill Cornell Medical College can request Cornell Netid

Setting up an account

- In order to get an account fill out account request form online https://biohpc.cornell.edu/NewUserRequest.aspx
- All Cornell employees, students and alumni are eligible
- Any collaborators of Cornell employees, students and alumni are eligible
- External user requests are considered on case-by-case basis, depending on BioHPC overall usage.



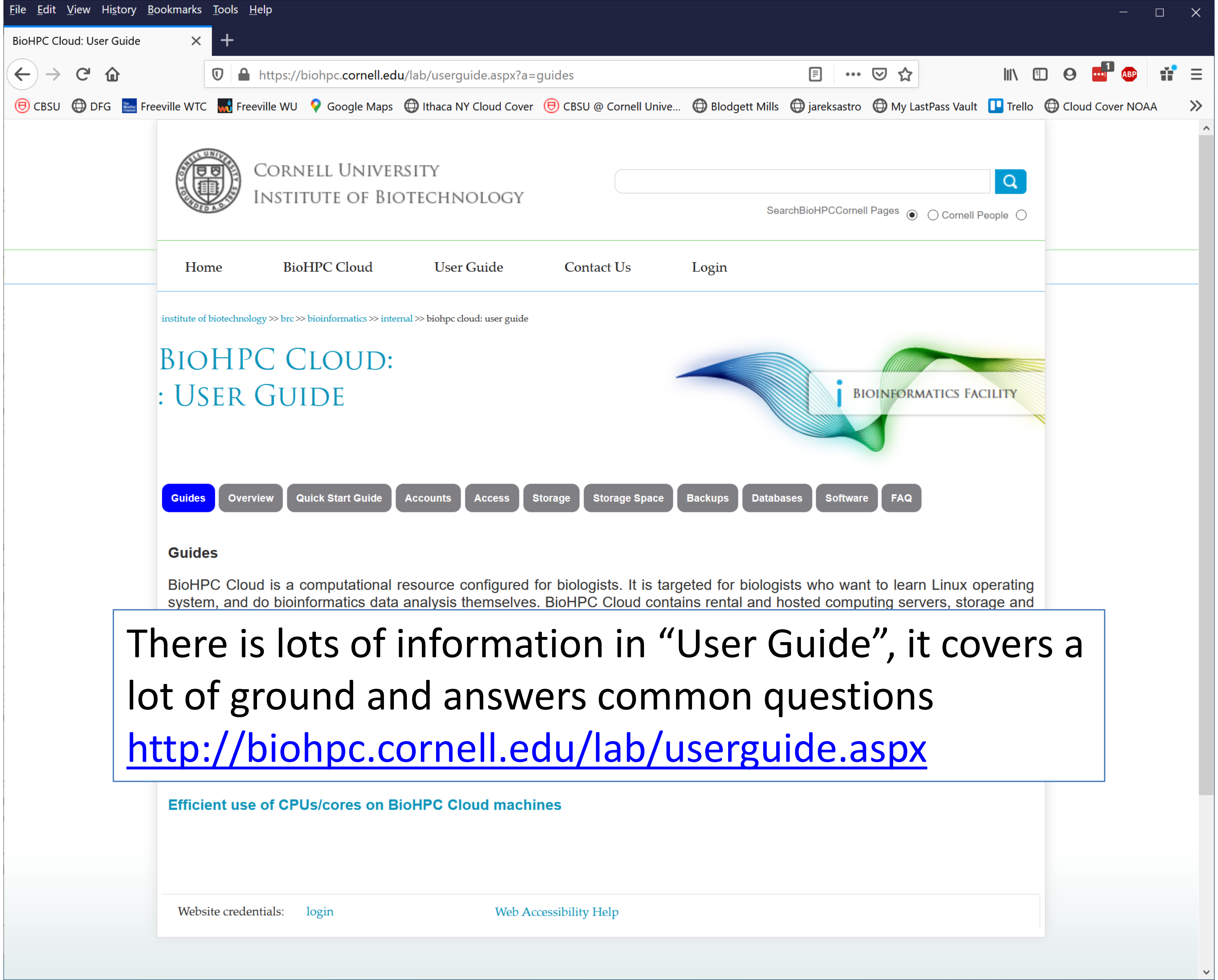
Welcome to the inter … Center Bioinformatics Facility (Computational Biology Service Unit, CBSU). BRC Bioinformatics Facili … computational biology and bioinformatics.

This website integra … resources of the facility.
For more informatio … ty main website.

If you would like to … , services and new develop

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

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



BIOHPC CLOUD:
: USER GUIDE

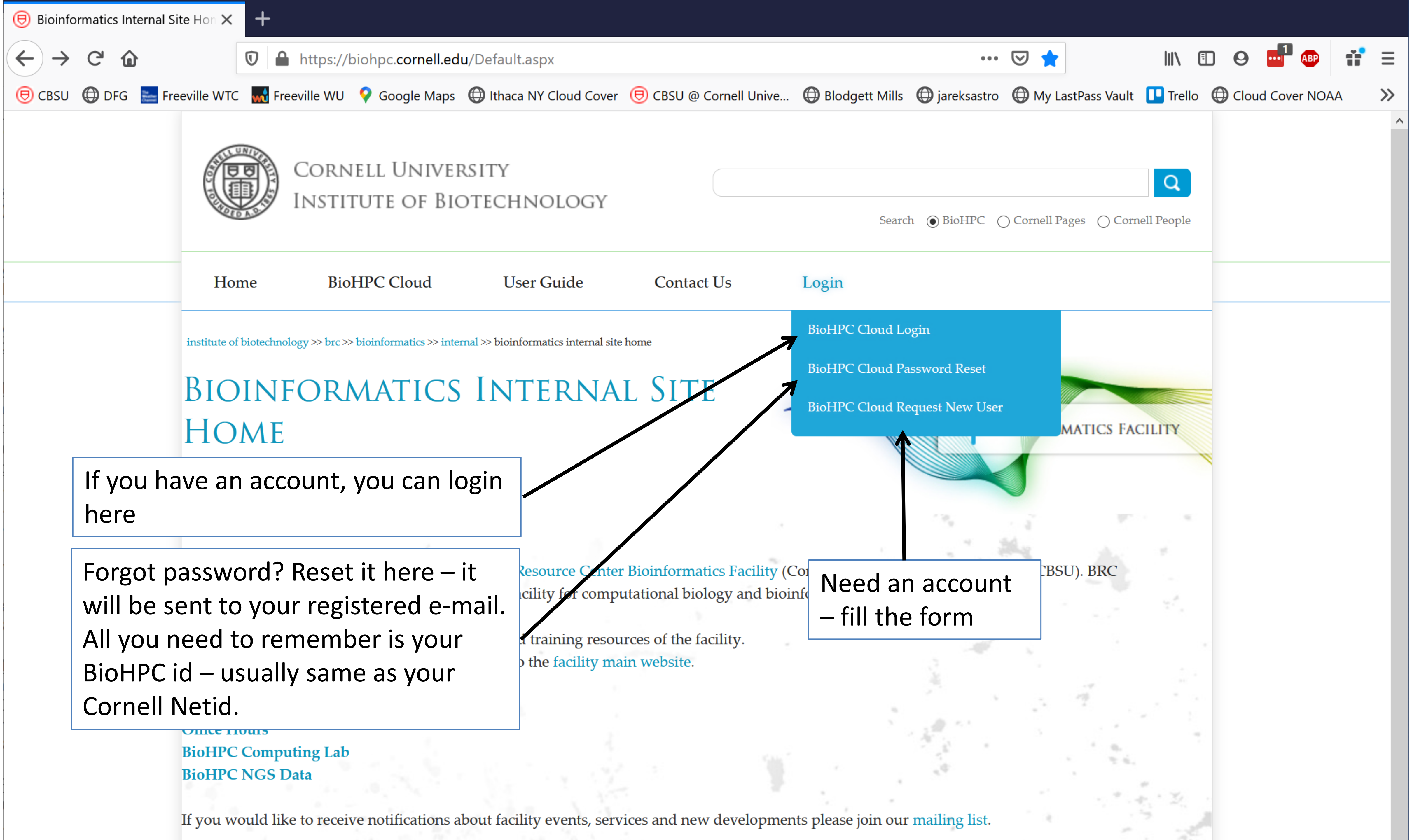
Guides

BioHPC Cloud is a computational resource configured for biologists. It is targeted for biologists who want to learn Linux operating system, and do bioinformatics data analysis themselves. BioHPC Cloud contains rental and hosted computing servers, storage and

Efficient use of CPUs/cores on BioHPC Cloud machines

Website credentials: login Web Accessibility Help

There is lots of information in “User Guide”, it covers a lot of ground and answers common questions
http://biohpc.cornell.edu/lab/userguide.aspx



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

BIOINFORMATICS INTERNAL SITE HOME

Resource Center Bioinformatics Facility (Co… CBSU). BRC
…cility for computational biology and bioinf…

…training resources of the facility.
… the facility main website.

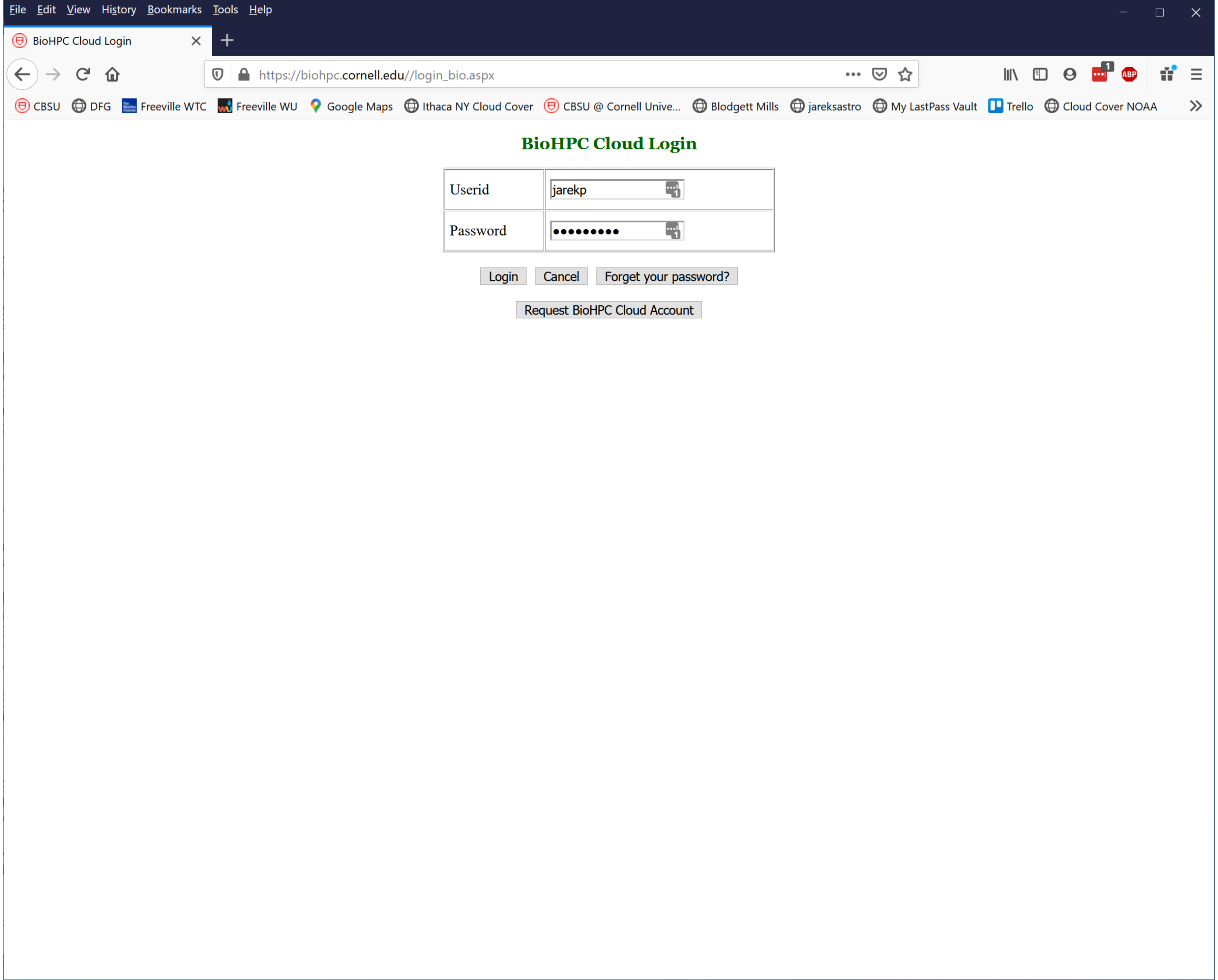
Office Hours
BioHPC Computing Lab
BioHPC NGS Data

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

If you have an account, you can login here

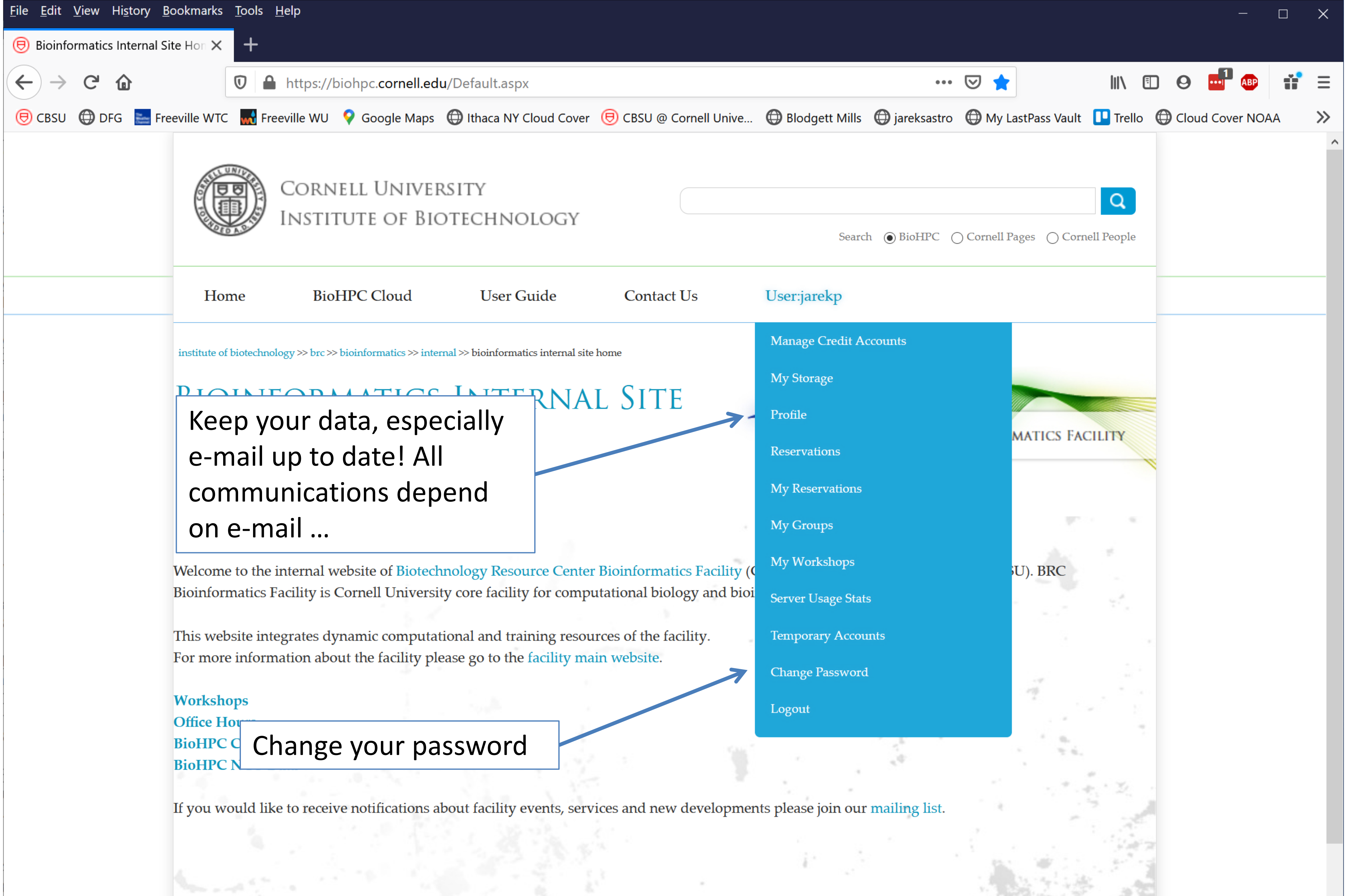
Forgot password? Reset it here – it will be sent to your registered e-mail. All you need to remember is your BioHPC id – usually same as your Cornell Netid.

Need an account – fill the form



BioHPC Cloud Login

Userid	jarekp
Password	●●●●●●●●●



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

BIOINFORMATICS INTERNAL SITE

Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (C U). BRC Bioinformatics Facility is Cornell University core facility for computational biology and bioi

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 Hou
BioHPC C
BioHPC N

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 password

Getting started with a new account

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

Credit Accounts

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

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

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

You can set up your own Credit Account and buy hours using Cornell Account, a credit card or a PO.

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://biohpc.cornell.edu/Lab/Pricing.aspx

Credit Accounts

CLASS	EQUIVALENT TO
purchase hours	1.000
general	3.052
medium gen1	1.960
large gen1	1.078
gpu-equipped gen1	1.674
extra large	0.750
large gen2	0.833
medium gen2	1.691
gpu-equipped gen2	0.990

What is the best server type to use?

Depends very much on the task at hand!

You need large memory when doing assembly and similar tasks.

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

Credit Accounts

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

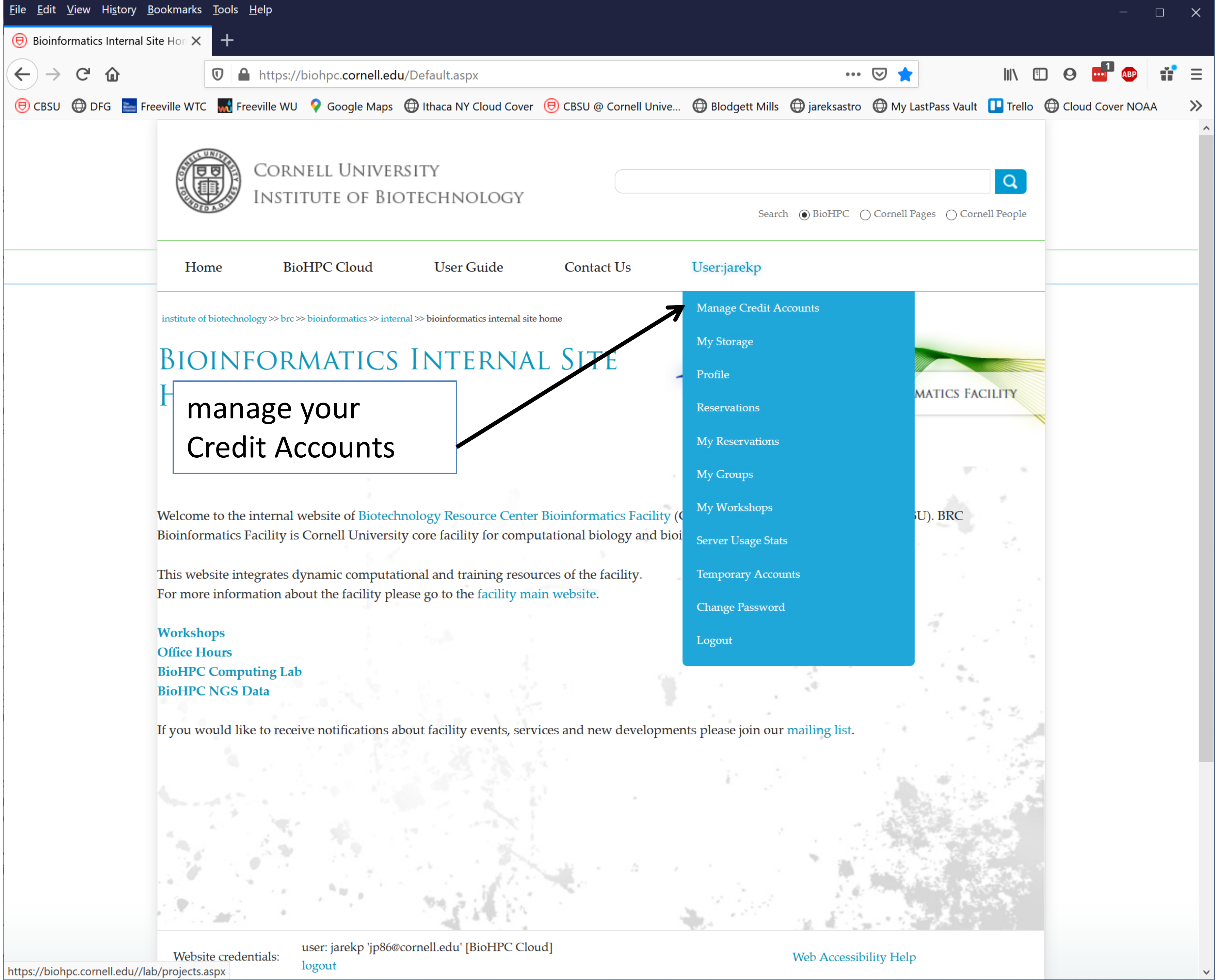
The more time you buy at a time the cheaper it gets.

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

medium gen2

This pricing applies to gen2 medium memory generally accessible workstations, i.e. cbsumm21-32 machines (40 cores; 256GB RAM; 8TB HDD).

purchase unit	current unit	hours	unit cost (Cornell)	cost per hour (Cornell)		unit cost (external)	cost per hour (external)	
				server	core		server	core
60 hours	101.5 hours	101.5 hours	$93.00	$0.92	$0.02 (40 cores)	$116.25	$1.15	$0.03 (40 cores)
200 hours	338.2 hours	338.2 hours	$309.00	$0.91	$0.02 (40 cores)	$386.25	$1.14	$0.03 (40 cores)
1 month	1.7 months	1,234.6 hours	$902.00	$0.73	$0.02 (40 cores)	$1,127.50	$0.91	$0.02 (40 cores)
6 months	10.1 months	7,407.5 hours	$4,326.00	$0.58	$0.01 (40 cores)	$5,407.50	$0.73	$0.02 (40 cores)
1 year	1.7 years	14,815.1 hours	$6,921.00	$0.47	$0.01 (40 cores)	$8,651.25	$0.58	$0.01 (40 cores)



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

BIOINFORMATICS INTERNAL SITE H

manage your Credit Accounts

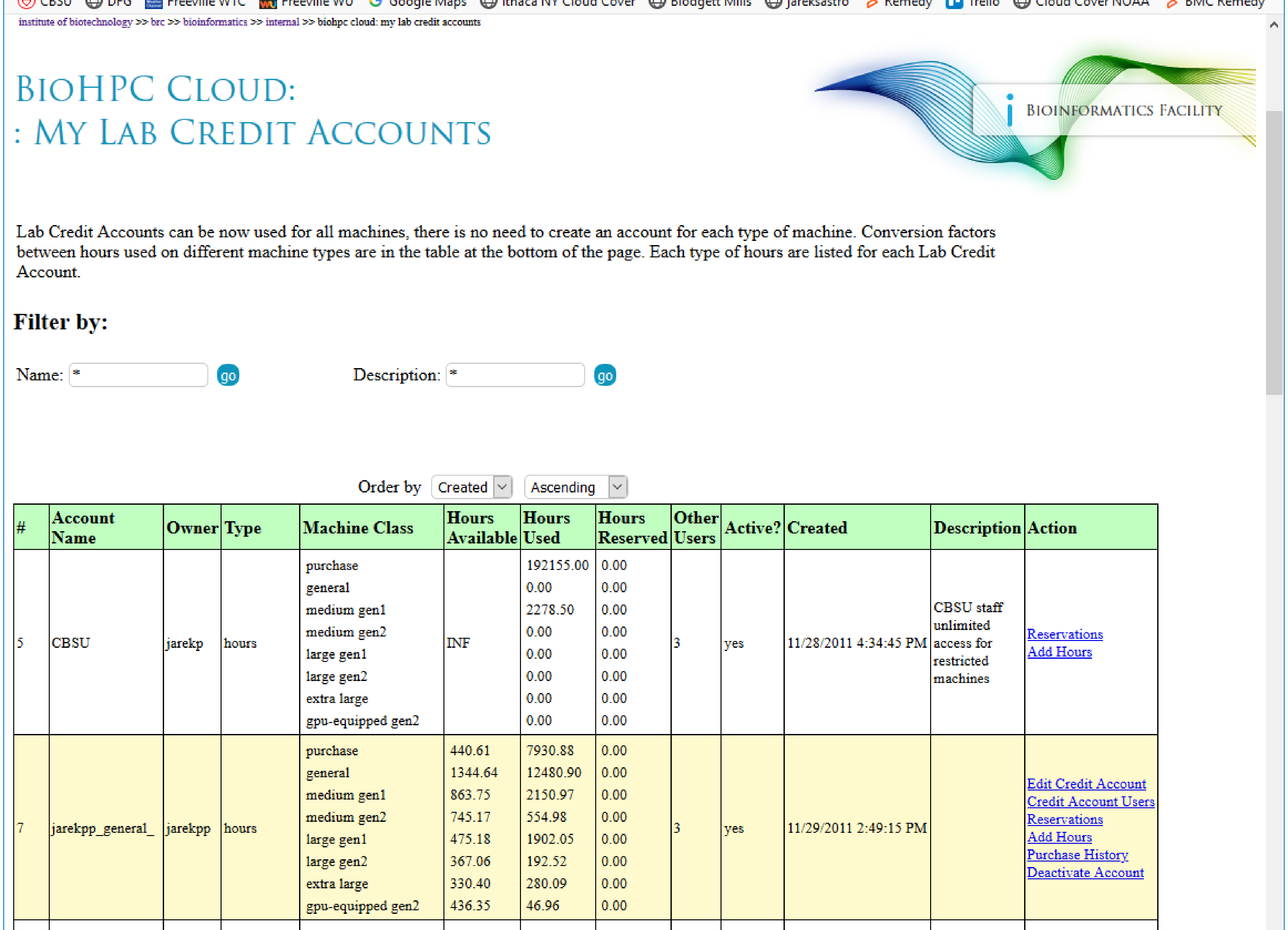
Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (C U). BRC Bioinformatics Facility is Cornell University core facility for computational biology and bioi

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

Workshops
Office Hours
BioHPC Computing Lab
BioHPC NGS Data

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

Website credentials: user: jarekp 'jp86@cornell.edu' [BioHPC Cloud]
logout



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

BioHPC Cloud:
: My Lab Credit Accounts

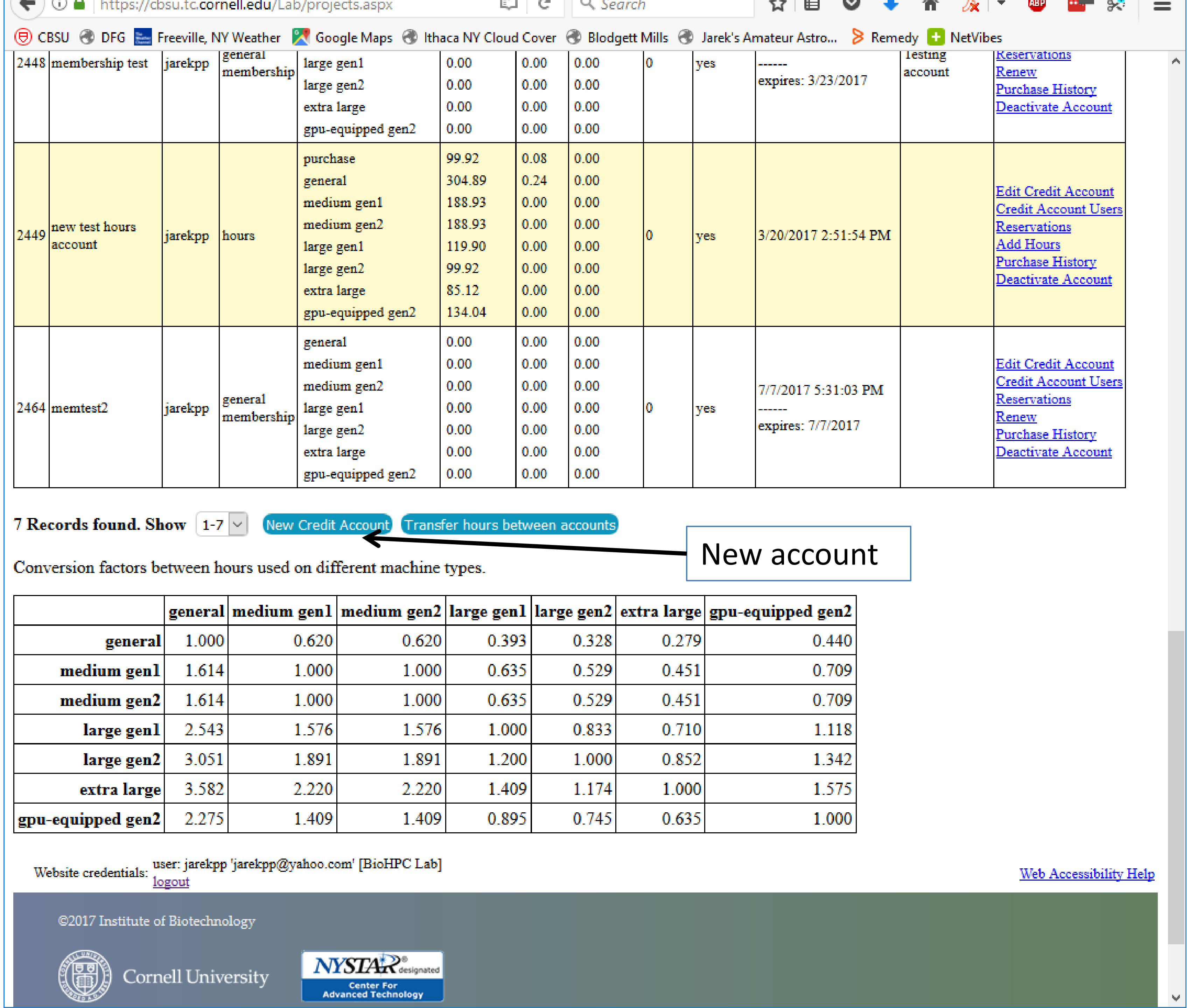
Lab Credit Accounts can be now used for all machines, there is no need to create an account for each type of machine. Conversion factors between hours used on different machine types are in the table at the bottom of the page. Each type of hours are listed for each Lab Credit Account.

Filter by:

Name: * go Description: * go

Order by Created Ascending

#	Account Name	Owner	Type	Machine Class	Hours Available	Hours Used	Hours Reserved	Other Users	Active?	Created	Description	Action
5	CBSU	jarekp	hours	purchase general medium gen1 medium gen2 large gen1 large gen2 extra large gpu-equipped gen2	INF	192155.00 0.00 2278.50 0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00	3	yes	11/28/2011 4:34:45 PM	CBSU staff unlimited access for restricted machines	Reservations Add Hours
7	jarekpp_general_	jarekpp	hours	purchase general medium gen1 medium gen2 large gen1 large gen2 extra large gpu-equipped gen2	440.61 1344.64 863.75 745.17 475.18 367.06 330.40 436.35	7930.88 12480.90 2150.97 554.98 1902.05 192.52 280.09 46.96	0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00	3	yes	11/29/2011 2:49:15 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account



2448	membership test	jarekpp	general membership	large gen1 large gen2 extra large gpu-equipped gen2	0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00	0	yes	------ expires: 3/23/2017	Testing account	Reservations Renew Purchase History Deactivate Account
2449	new test hours account	jarekpp	hours	purchase general medium gen1 medium gen2 large gen1 large gen2 extra large gpu-equipped gen2	99.92 304.89 188.93 188.93 119.90 99.92 85.12 134.04	0.08 0.24 0.00 0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00	0	yes	3/20/2017 2:51:54 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
2464	memtest2	jarekpp	general membership	general medium gen1 medium gen2 large gen1 large gen2 extra large gpu-equipped gen2	0.00 0.00 0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00 0.00 0.00	0.00 0.00 0.00 0.00 0.00 0.00 0.00	0	yes	7/7/2017 5:31:03 PM ------ expires: 7/7/2017		Edit Credit Account Credit Account Users Reservations Renew Purchase History Deactivate Account

7 Records found. Show 1-7

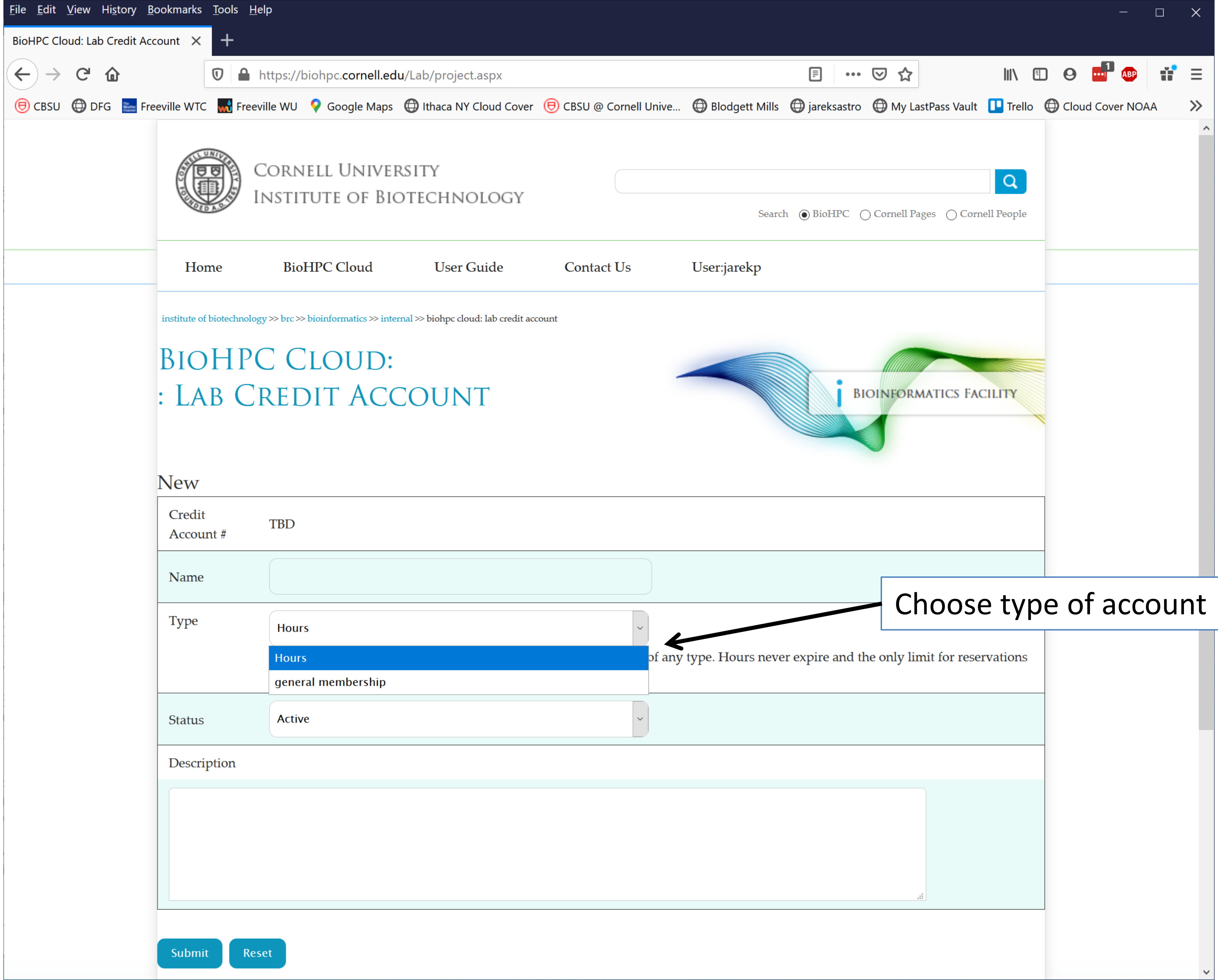
New account

Conversion factors between hours used on different machine types.

	general	medium gen1	medium gen2	large gen1	large gen2	extra large	gpu-equipped gen2
general	1.000	0.620	0.620	0.393	0.328	0.279	0.440
medium gen1	1.614	1.000	1.000	0.635	0.529	0.451	0.709
medium gen2	1.614	1.000	1.000	0.635	0.529	0.451	0.709
large gen1	2.543	1.576	1.576	1.000	0.833	0.710	1.118
large gen2	3.051	1.891	1.891	1.200	1.000	0.852	1.342
extra large	3.582	2.220	2.220	1.409	1.174	1.000	1.575
gpu-equipped gen2	2.275	1.409	1.409	0.895	0.745	0.635	1.000

Website credentials: user: jarekpp 'jarekpp@yahoo.com' [BioHPC Lab]
logout

Web Accessibility Help



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

BioHPC Cloud: : Lab Credit Account

New

Credit Account #	TBD
Name	
Type	Hours Hours general membership
Status	Active
Description	

of any type. Hours never expire and the only limit for reservations

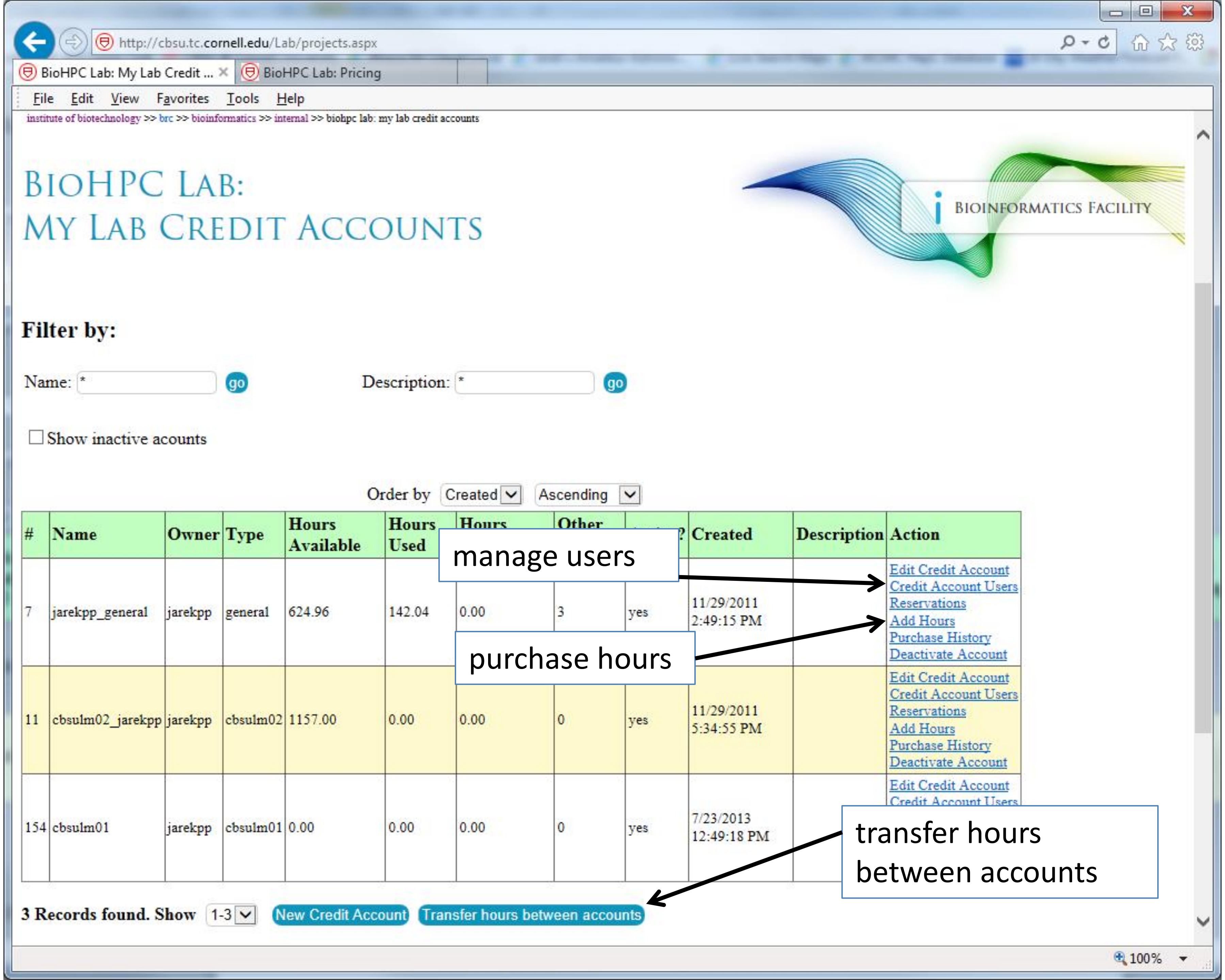
Choose type of account

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.

machine	maximum reservation length (rolling)		maximum combined reservation (rolling)		annual cost (Cornell)	annual cost (external)
	unit	hours	unit	hours		
general	1.0 weeks	168.0	1.0 weeks	168.0	$500.00	$600.00
medium gen1	1.0 days	24.0	1.0 days	24.0		



BioHPC Lab: My Lab Credit Accounts

Filter by:

Name: * go Description: * go

☐ Show inactive acounts

Order by Created Ascending

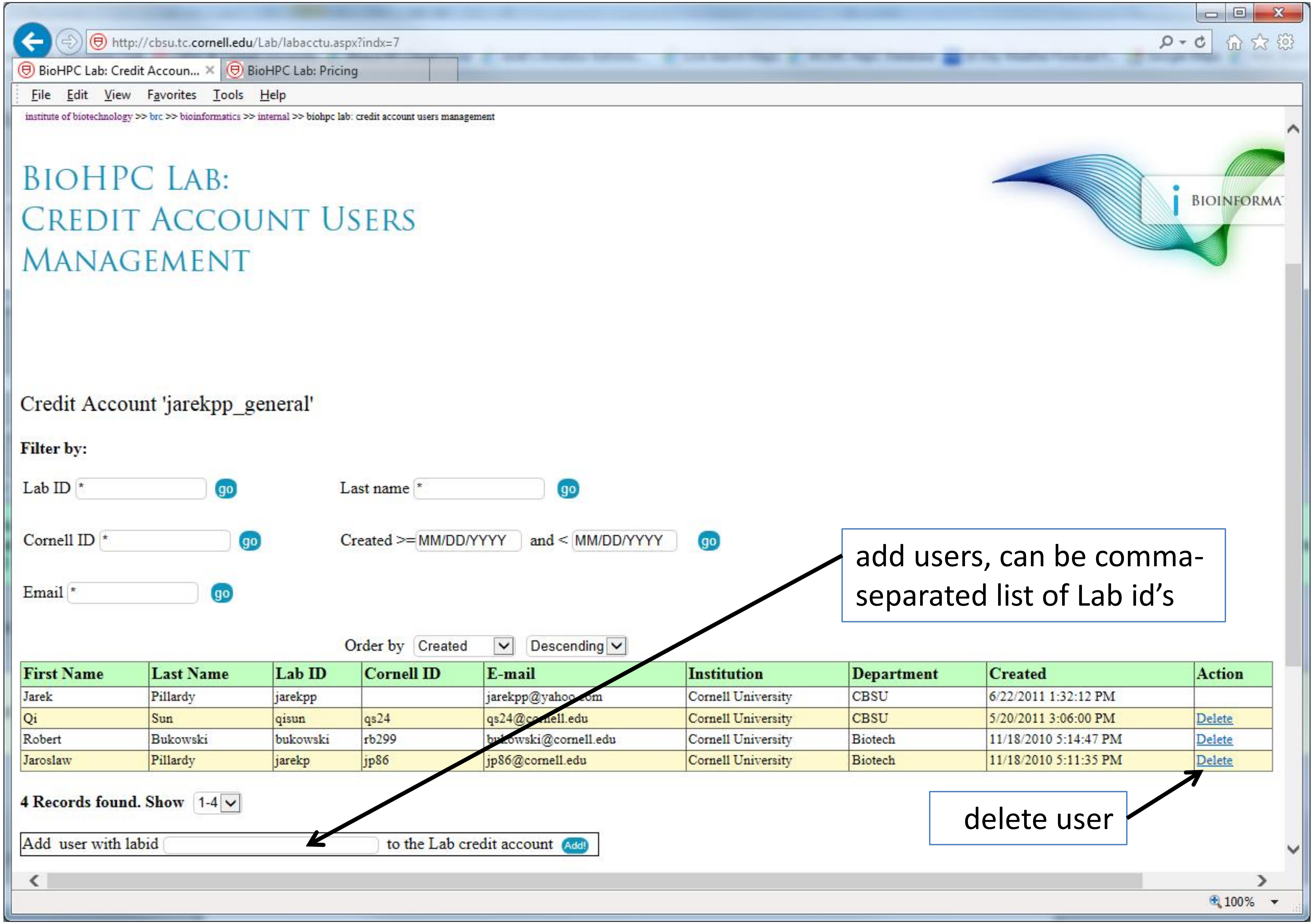
#	Name	Owner	Type	Hours Available	Hours Used	Hours	Other	?	Created	Description	Action
7	jarekpp_general	jarekpp	general	624.96	142.04	0.00	3	yes	11/29/2011 2:49:15 PM		Edit Credit Account Credit Account Users Reservations Add Hours Purchase History Deactivate Account
11	cbsulm02_jarekpp	jarekpp	cbsulm02	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 between accounts



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 *

Last name *

Cornell ID *

Created >= MM/DD/YYYY and < MM/DD/YYYY

Email *

Order by Created Descending

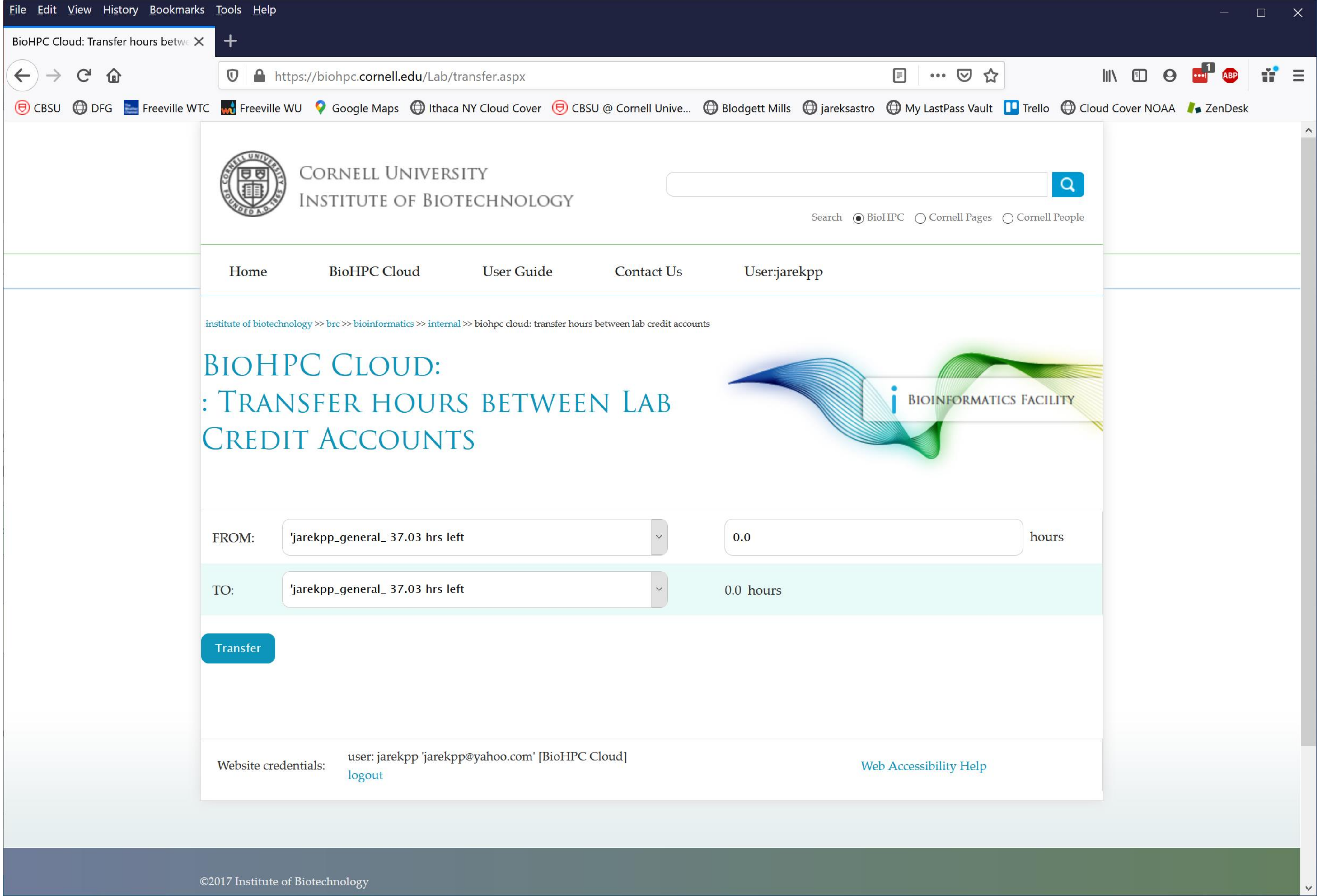
First Name	Last Name	Lab ID	Cornell ID	E-mail	Institution	Department	Created	Action
Jarek	Pillardy	jarekpp		jarekpp@yahoo.com	Cornell University	CBSU	6/22/2011 1:32:12 PM	
Qi	Sun	qisun	qs24	qs24@cornell.edu	Cornell University	CBSU	5/20/2011 3:06:00 PM	Delete
Robert	Bukowski	bukowski	rb299	bukowski@cornell.edu	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 users, can be comma-separated list of Lab id's

delete user



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

BioHPC Cloud:
: Transfer hours between Lab Credit Accounts

FROM: 'jarekpp_general_ 37.03 hrs left 0.0 hours

TO: 'jarekpp_general_ 37.03 hrs left 0.0 hours

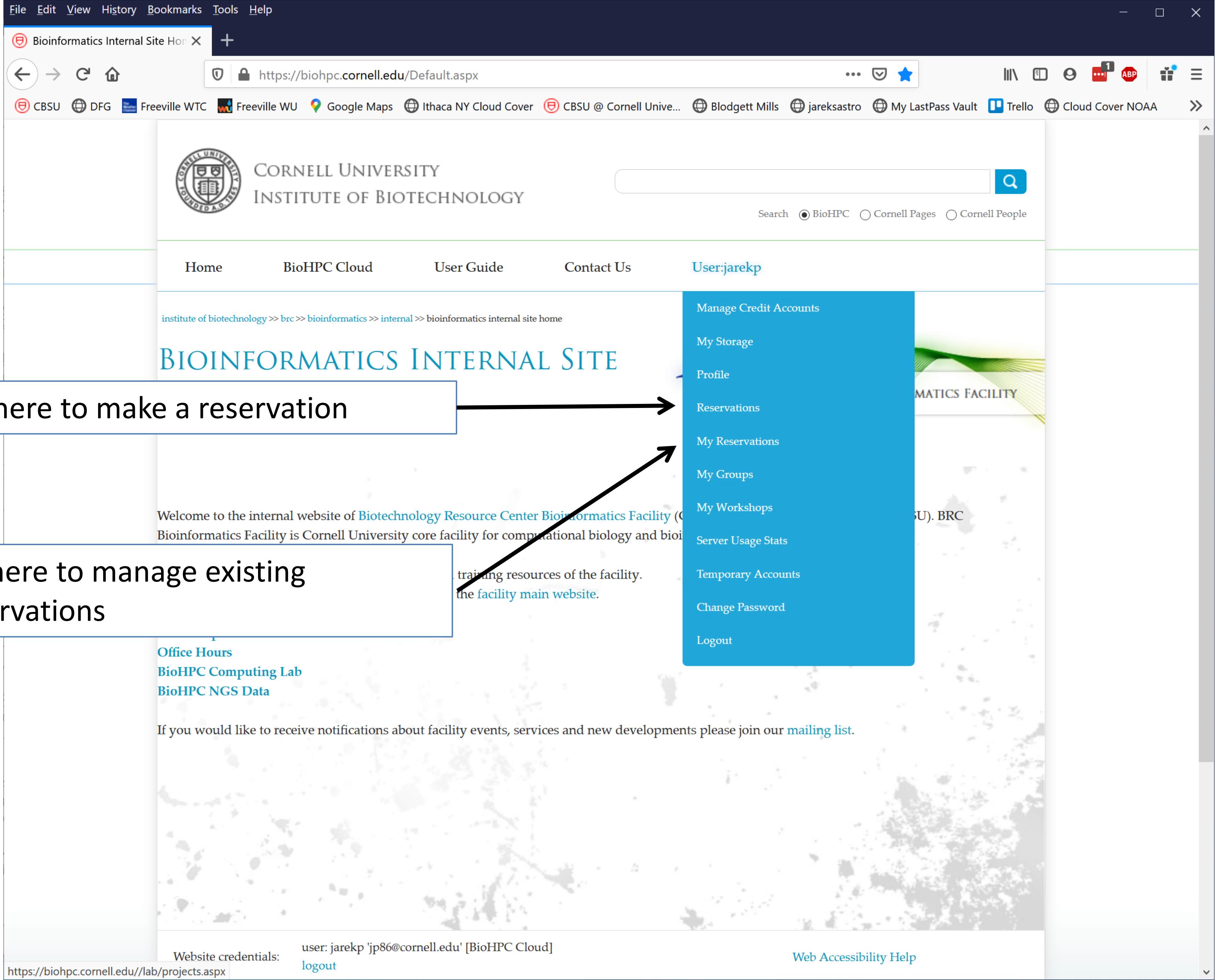
Website credentials: user: jarekpp 'jarekpp@yahoo.com' [BioHPC Cloud]
logout

Web Accessibility Help

©2017 Institute of Biotechnology

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 >> bioinformatics internal site home

BIOINFORMATICS INTERNAL SITE

Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (C U). BRC Bioinformatics Facility is Cornell University core facility for computational biology and bioi

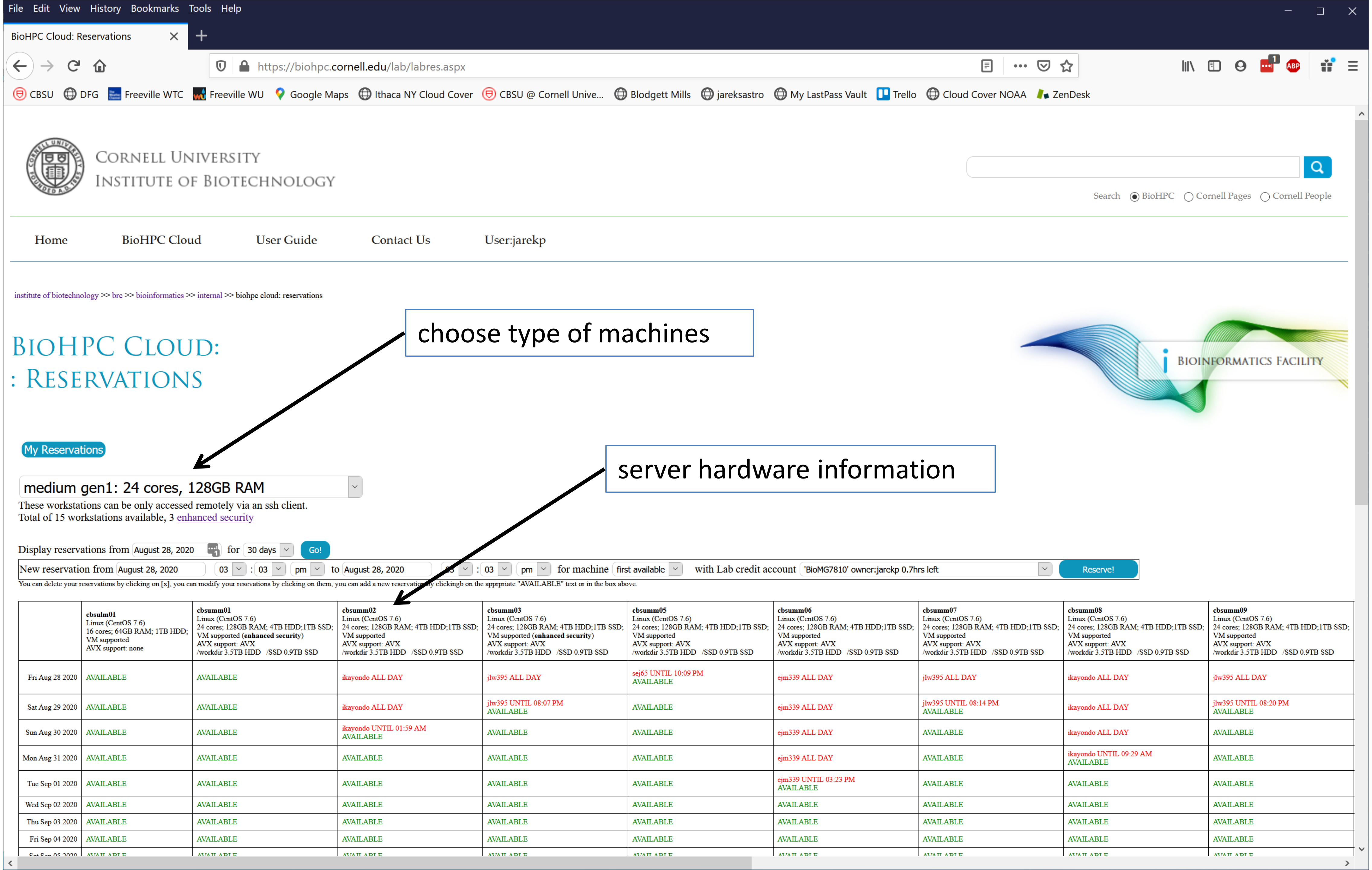
training resources of the facility.
the facility main website.

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

Website credentials: user: jarekp 'jp86@cornell.edu' [BioHPC Cloud]
logout

Go here to make a reservation

Go here to manage existing reservations



choose type of machines

server hardware information

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

BIOHPC CLOUD: : RESERVATIONS

medium gen1: 24 cores, 128GB RAM

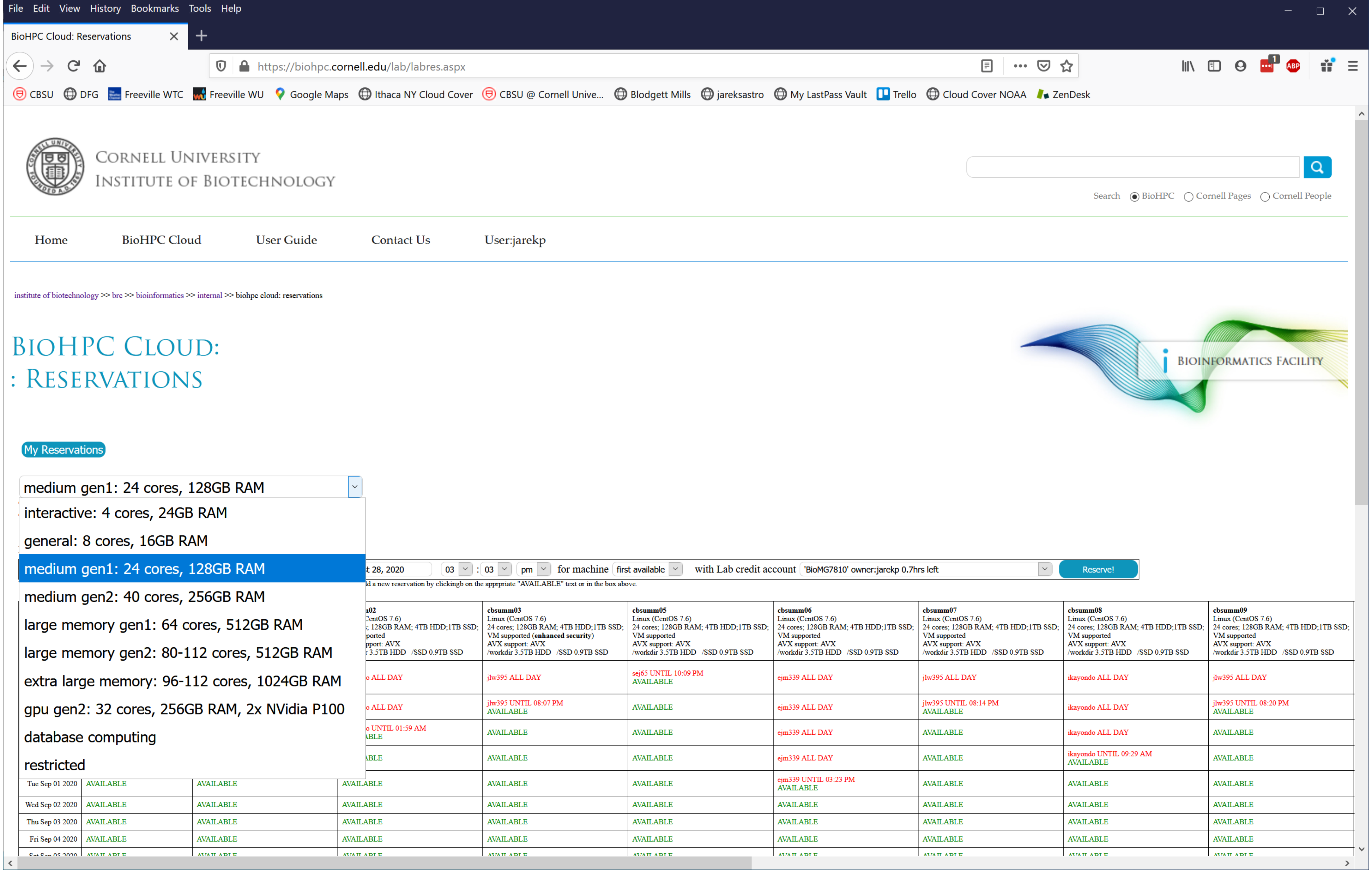
These workstations can be only accessed remotely via an ssh client.
Total of 15 workstations available, 3 enhanced security

Display reservations from August 28, 2020 for 30 days

New reservation from August 28, 2020 03 : 03 pm to August 28, 2020 03 : 03 pm for machine first available with Lab credit account 'BioMG7810' owner:jarekp 0.7hrs 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.

	cbsulm01 Linux (CentOS 7.6) 16 cores; 64GB RAM; 1TB HDD; VM supported AVX support: none	**cbsumm01** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm02** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm03** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm05** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm06** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm07** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm08** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm09** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD
Fri Aug 28 2020	AVAILABLE	AVAILABLE	ikayondo ALL DAY	jlw395 ALL DAY	sej65 UNTIL 10:09 PM AVAILABLE	ejm339 ALL DAY	jlw395 ALL DAY	ikayondo ALL DAY	jlw395 ALL DAY
Sat Aug 29 2020	AVAILABLE	AVAILABLE	ikayondo ALL DAY	jlw395 UNTIL 08:07 PM AVAILABLE	AVAILABLE	ejm339 ALL DAY	jlw395 UNTIL 08:14 PM AVAILABLE	ikayondo ALL DAY	jlw395 UNTIL 08:20 PM AVAILABLE
Sun Aug 30 2020	AVAILABLE	AVAILABLE	ikayondo UNTIL 01:59 AM AVAILABLE	AVAILABLE	AVAILABLE	ejm339 ALL DAY	AVAILABLE	ikayondo ALL DAY	AVAILABLE
Mon Aug 31 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ejm339 ALL DAY	AVAILABLE	ikayondo UNTIL 09:29 AM AVAILABLE	AVAILABLE
Tue Sep 01 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ejm339 UNTIL 03:23 PM AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Wed Sep 02 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Thu Sep 03 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Fri Sep 04 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Sat Sep 05 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE



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

BioHPC Cloud: : Reservations

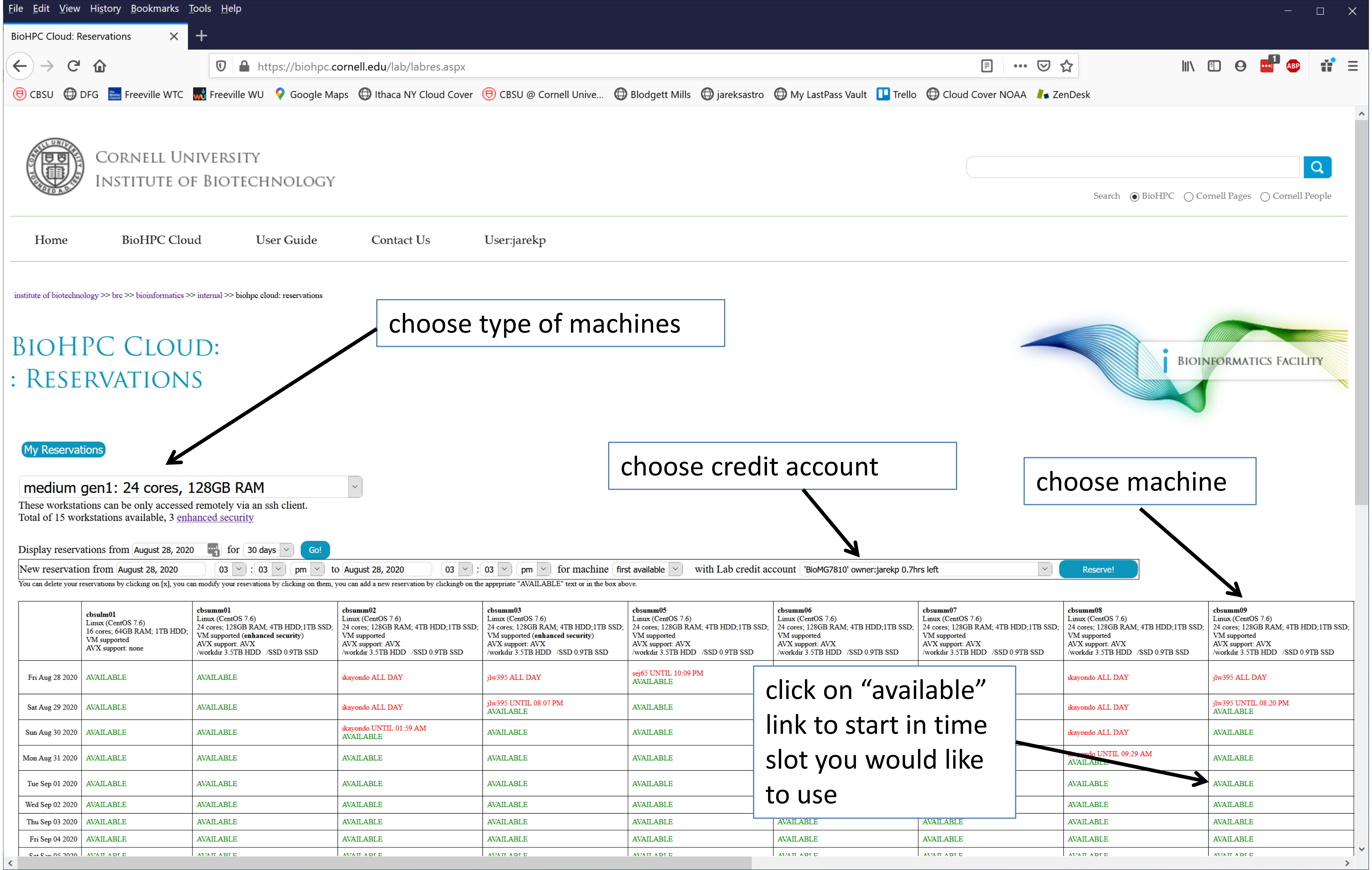
medium gen1: 24 cores, 128GB RAM

interactive: 4 cores, 24GB RAM
general: 8 cores, 16GB RAM
medium gen1: 24 cores, 128GB RAM
medium gen2: 40 cores, 256GB RAM
large memory gen1: 64 cores, 512GB RAM
large memory gen2: 80-112 cores, 512GB RAM
extra large memory: 96-112 cores, 1024GB RAM
gpu gen2: 32 cores, 256GB RAM, 2x NVidia P100
database computing
restricted

t 28, 2020 03 : 03 pm for machine first available with Lab credit account 'BioMG7810' owner:jarekp 0.7hrs left

ld a new reservation by clickingb on the apprpriate "AVAILABLE" text or in the box above.

			n02 CentOS 7.6) ; 128GB RAM; 4TB HDD;1TB SSD; ported pport: AVX r 3.5TB HDD /SSD 0.9TB SSD	**cbsumm03** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm05** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm06** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm07** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm08** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm09** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD
			o ALL DAY	jlw395 ALL DAY	sej65 UNTIL 10:09 PM AVAILABLE	ejm339 ALL DAY	jlw395 ALL DAY	ikayondo ALL DAY	jlw395 ALL DAY
			o ALL DAY	jlw395 UNTIL 08:07 PM AVAILABLE	AVAILABLE	ejm339 ALL DAY	jlw395 UNTIL 08:14 PM AVAILABLE	ikayondo ALL DAY	jlw395 UNTIL 08:20 PM AVAILABLE
			o UNTIL 01:59 AM ABLE	AVAILABLE	AVAILABLE	ejm339 ALL DAY	AVAILABLE	ikayondo ALL DAY	AVAILABLE
			ABLE	AVAILABLE	AVAILABLE	ejm339 ALL DAY	AVAILABLE	ikayondo UNTIL 09:29 AM AVAILABLE	AVAILABLE
Tue Sep 01 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ejm339 UNTIL 03:23 PM AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Wed Sep 02 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Thu Sep 03 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Fri Sep 04 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE



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

BioHPC Cloud:
: Reservations

choose type of machines

medium gen1: 24 cores, 128GB RAM

These workstations can be only accessed remotely via an ssh client.
Total of 15 workstations available, 3 enhanced security

choose credit account

choose machine

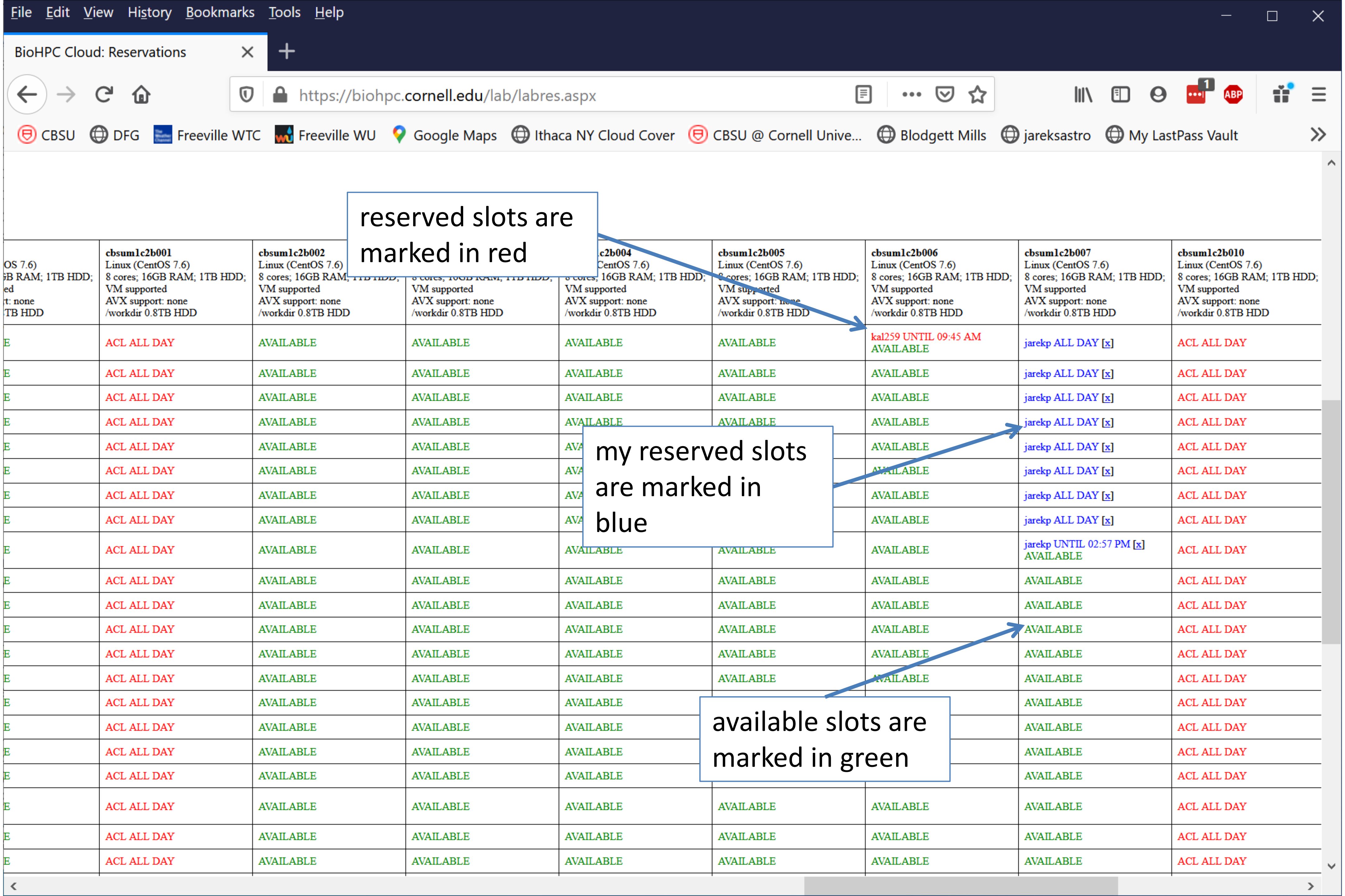
Display reservations from August 28, 2020 for 30 days

New reservation from August 28, 2020 03 : 03 pm to August 28, 2020 03 : 03 pm for machine first available with Lab credit account 'BioMG7810' owner:jarekp 0.7hrs 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.

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

	cbsulm01 Linux (CentOS 7.6) 16 cores; 64GB RAM; 1TB HDD; VM supported AVX support: none	**cbsumm01** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm02** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm03** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported (**enhanced security**) AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm05** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm06** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm07** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm08** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD	**cbsumm09** Linux (CentOS 7.6) 24 cores; 128GB RAM; 4TB HDD;1TB SSD; VM supported AVX support: AVX /workdir 3.5TB HDD /SSD 0.9TB SSD
Fri Aug 28 2020	AVAILABLE	AVAILABLE	ikayondo ALL DAY	jlw395 ALL DAY	sej65 UNTIL 10:09 PM AVAILABLE			ikayondo ALL DAY	jlw395 ALL DAY
Sat Aug 29 2020	AVAILABLE	AVAILABLE	ikayondo ALL DAY	jlw395 UNTIL 08:07 PM AVAILABLE	AVAILABLE			ikayondo ALL DAY	jlw395 UNTIL 08:20 PM AVAILABLE
Sun Aug 30 2020	AVAILABLE	AVAILABLE	ikayondo UNTIL 01:59 AM AVAILABLE	AVAILABLE	AVAILABLE			ikayondo ALL DAY	AVAILABLE
Mon Aug 31 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE			ikayondo UNTIL 09:29 AM AVAILABLE	AVAILABLE
Tue Sep 01 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE			AVAILABLE	AVAILABLE
Wed Sep 02 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE			AVAILABLE	AVAILABLE
Thu Sep 03 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Fri Sep 04 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE

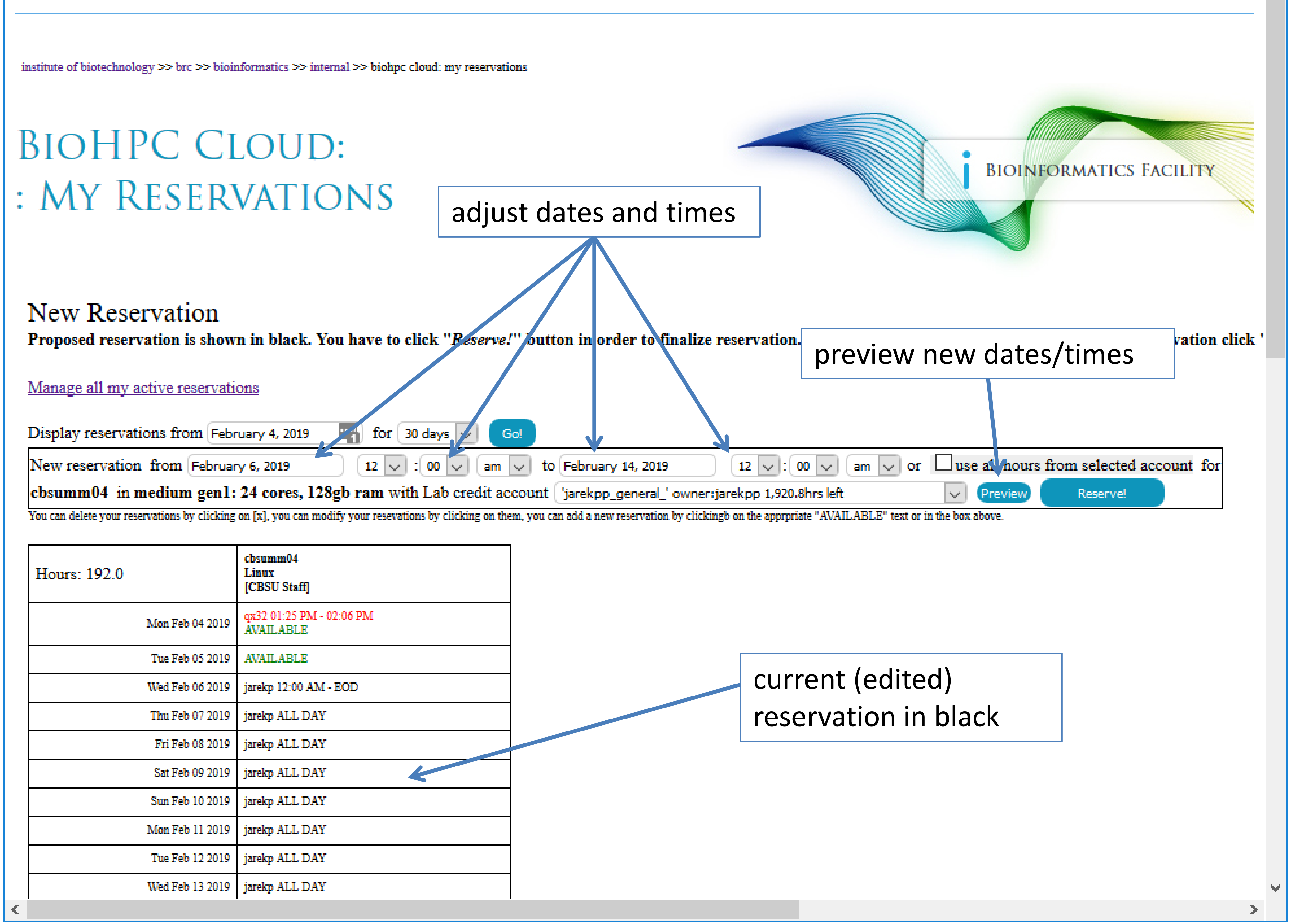


reserved slots are marked in red

my reserved slots are marked in blue

available slots are marked in green

OS 7.6) B RAM; 1TB HDD; ed t: none TB HDD	cbsumlc2b001 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	cbsumlc2b002 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	c2b004 CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	cbsumlc2b005 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	cbsumlc2b006 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	cbsumlc2b007 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD	cbsumlc2b010 Linux (CentOS 7.6) 8 cores; 16GB RAM; 1TB HDD; VM supported AVX support: none /workdir 0.8TB HDD
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	kal259 UNTIL 09:45 AM AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVA		AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVA		AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVA		AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVA		AVAILABLE	jarekp ALL DAY [x]	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	jarekp UNTIL 02:57 PM [x] AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE		AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE			AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE			AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE			AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY
E	ACL ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	ACL ALL DAY



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: my reservations

BIOHPC CLOUD: : MY RESERVATIONS

New Reservation

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

Manage all my active reservations

Display reservations from February 4, 2019 for 30 days Go!

New reservation from February 6, 2019 12 : 00 am to February 14, 2019 12 : 00 am or ☐ use all hours from selected account for **cbsumm04** in **medium gen1: 24 cores, 128gb ram** with Lab credit account 'jarekpp_general_' owner:jarekpp 1,920.8hrs left Preview Reserve!

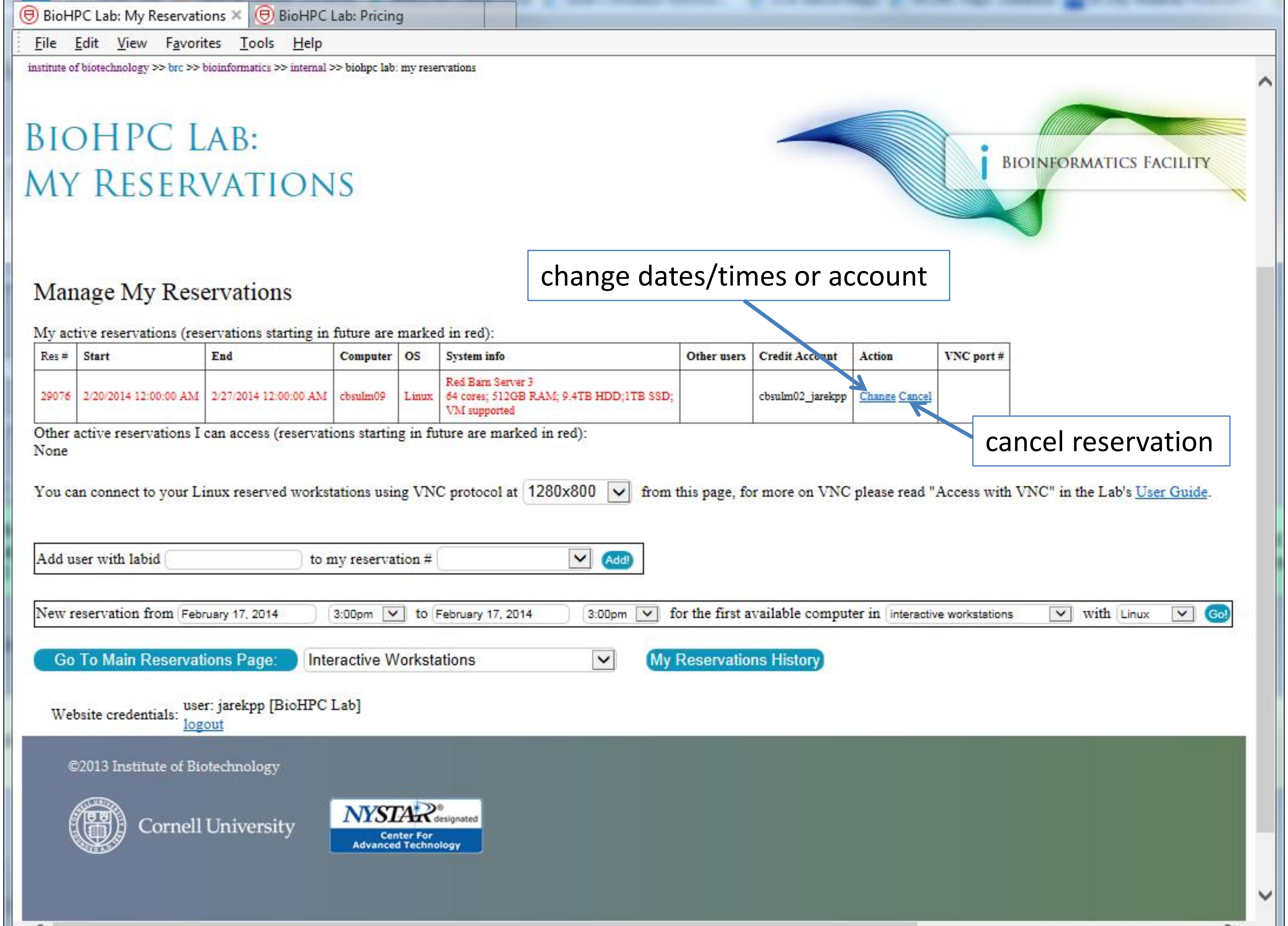
You can delete your reservations by clicking on [x], you can modify your resevations by clicking on them, you can add a new reservation by clickingb on the apprpriate "AVAILABLE" text or in the box above.

Hours: 192.0	cbsumm04 Linux [CBSU Staff]
Mon Feb 04 2019	qx32 01:25 PM - 02:06 PM AVAILABLE
Tue Feb 05 2019	AVAILABLE
Wed Feb 06 2019	jarekp 12:00 AM - EOD
Thu Feb 07 2019	jarekp ALL DAY
Fri Feb 08 2019	jarekp ALL DAY
Sat Feb 09 2019	jarekp ALL DAY
Sun Feb 10 2019	jarekp ALL DAY
Mon Feb 11 2019	jarekp ALL DAY
Tue Feb 12 2019	jarekp ALL DAY
Wed Feb 13 2019	jarekp ALL DAY

adjust dates and times

preview new dates/times

current (edited) reservation in black



BIOHPC LAB: MY RESERVATIONS

Manage My Reservations

change dates/times or account

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	

cancel reservation

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

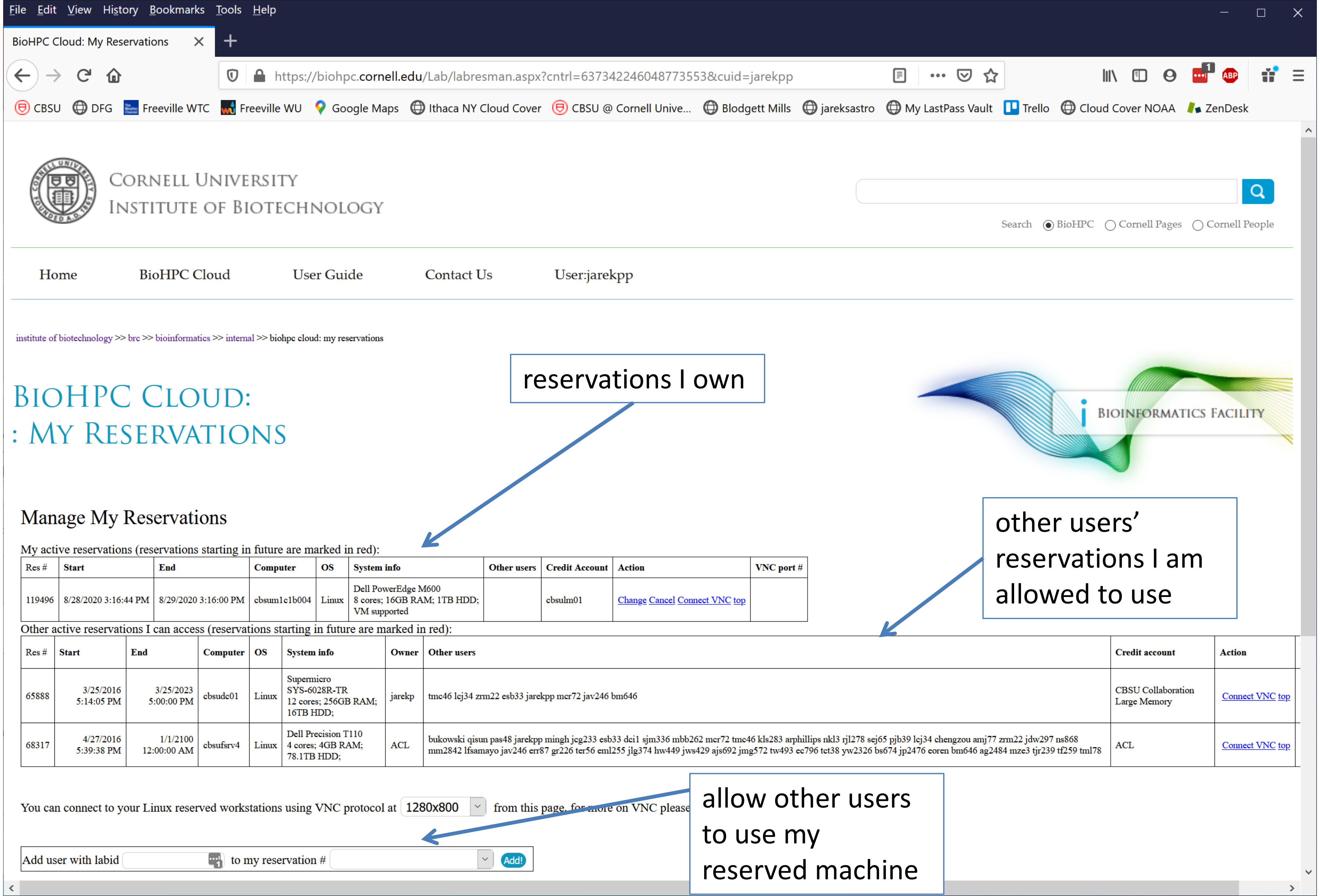
You can connect to your Linux reserved workstations using VNC protocol at 1280x800 from this page, for more on VNC please read "Access with VNC" in the Lab's User Guide.

Add user with labid ______ to my reservation # ______

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

Website credentials: user: jarekpp [BioHPC Lab]

©2013 Institute of Biotechnology



reservations I own

other users' reservations I am allowed to use

allow other users to use my reserved machine

institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: my reservations

BIOHPC CLOUD: : 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 #
119496	8/28/2020 3:16:44 PM	8/29/2020 3:16:00 PM	cbsum1c1b004	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported		cbsulm01	Change Cancel Connect VNC top	

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
65888	3/25/2016 5:14:05 PM	3/25/2023 5:00:00 PM	cbsudc01	Linux	Supermicro SYS-6028R-TR 12 cores; 256GB RAM; 16TB HDD;	jarekp	tmc46 lcj34 zrm22 esb33 jarekpp mcr72 jav246 bm646	CBSU Collaboration Large Memory	Connect VNC top
68317	4/27/2016 5:39:38 PM	1/1/2100 12:00:00 AM	cbsufsrv4	Linux	Dell Precision T110 4 cores; 4GB RAM; 78.1TB HDD;	ACL	bukowski qisun pas48 jarekpp mingh jcg233 esb33 dci1 sjm336 mbb262 mcr72 tmc46 kls283 arphillips nkl3 rjl278 sej65 pjb39 lcj34 chengzou amj77 zrm22 jdw297 ns868 mm2842 lfsamayo jav246 err87 gr226 ter56 eml255 jlg374 hw449 jws429 ajs692 jmg572 tw493 ec796 tct38 yw2326 bs674 jp2476 eoren bm646 ag2484 mze3 tjr239 tf259 tml78	ACL	Connect VNC top

You can connect to your Linux reserved workstations using VNC protocol at 1280x800 from this page, for more on VNC please

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

Reservations

You can only make a reservation if you have enough hours.

Your hours will be “reserved” when you make a reservation so you cannot use them twice – but not immediately deducted.

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

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

Reservations

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

- You will not be able to login to the machine anymore.
- If you are logged in you will stay logged in until the user from the next reservation logs in.
- Your programs will continue to run, as long as the machine is not used for 2 hours – this is to give you time to extend reservation if you need it.
- When a person that has the next reservation (now current) logs in all your programs and processes will be killed and you will be logged out. Same thing will happen after 2 hours past reservation

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
```

Reservations

What if I need a scheduler to submit jobs?

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

```
manage_slurm new machine1,machine2,...,machineN
```

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

https://biohpc.cornell.edu/lab/userguide.aspx?a=software&i=689#c

Getting started with a new account

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

Storage

- **Networked storage**

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

- **Local storage**

  Fast storage, especially on large memory and medium memory workstations. Small – between 1TB (general) and 12TB (large memory). *Designed to be used for computations*.

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 (/).

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

/home/jarekp/tmp/tmpfile

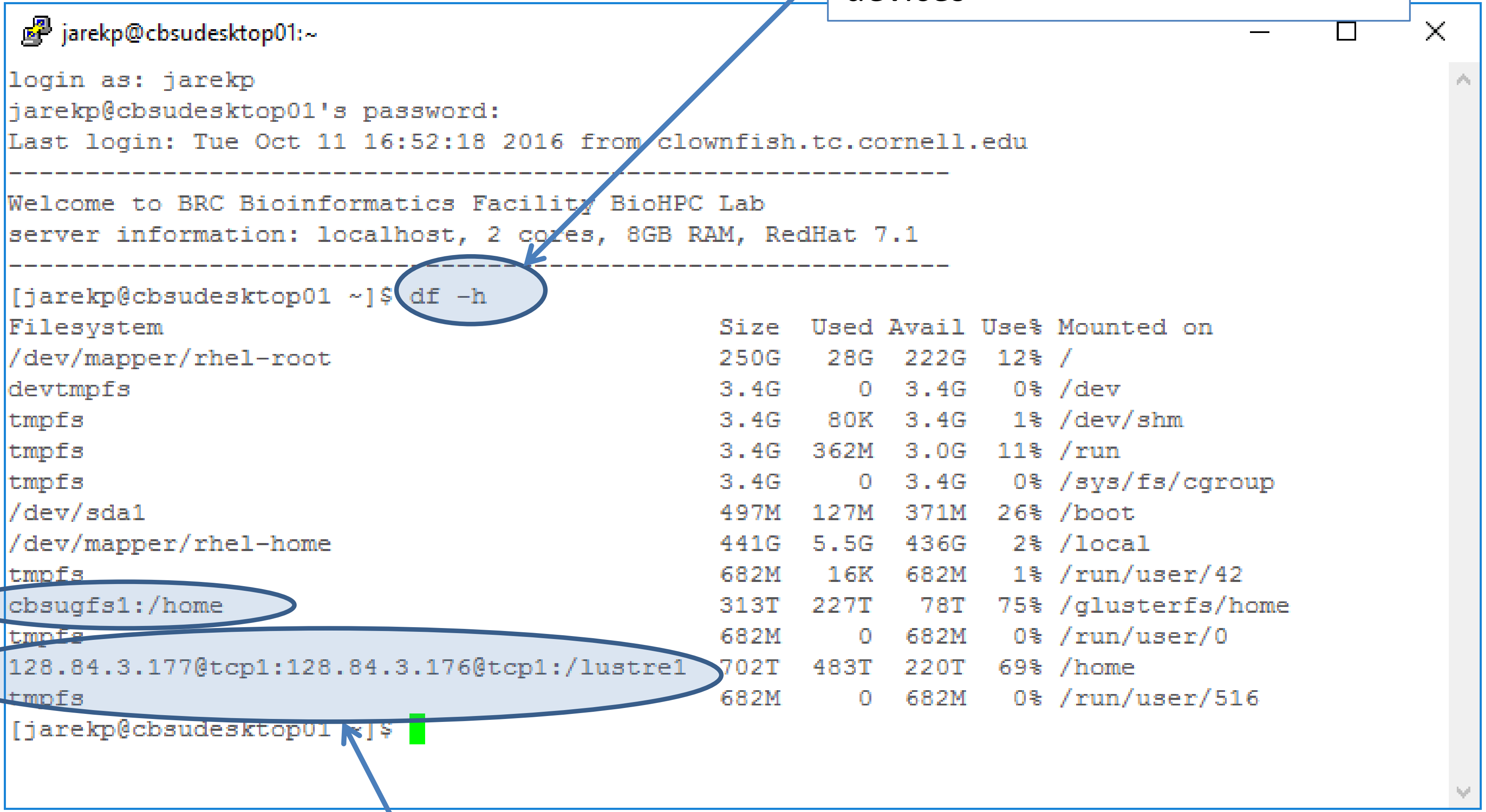
/programs/bin/labutils/endres.pl

Storage

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

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

Storage



```
login as: jarekp
jarekp@cbsudesktop01's password:
Last login: Tue Oct 11 16:52:18 2016 from clownfish.tc.cornell.edu
-------------------------------------------------------------
Welcome to BRC Bioinformatics Facility BioHPC Lab
server information: localhost, 2 cores, 8GB RAM, RedHat 7.1
-------------------------------------------------------------
[jarekp@cbsudesktop01 ~]$ df -h
Filesystem                                    Size  Used Avail Use% Mounted on
/dev/mapper/rhel-root                         250G   28G  222G  12% /
devtmpfs                                      3.4G     0  3.4G   0% /dev
tmpfs                                         3.4G   80K  3.4G   1% /dev/shm
tmpfs                                         3.4G  362M  3.0G  11% /run
tmpfs                                         3.4G     0  3.4G   0% /sys/fs/cgroup
/dev/sda1                                     497M  127M  371M  26% /boot
/dev/mapper/rhel-home                         441G  5.5G  436G   2% /local
tmpfs                                         682M   16K  682M   1% /run/user/42
cbsugfs1:/home                                313T  227T   78T  75% /glusterfs/home
tmpfs                                         682M     0  682M   0% /run/user/0
128.84.3.177@tcp1:128.84.3.176@tcp1:/lustre1  702T  483T  220T  69% /home
tmpfs                                         682M     0  682M   0% /run/user/516
[jarekp@cbsudesktop01 ~]$ 
```

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

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

Storage

- **Networked storage**

  /home

  /shared_data

  /programs

- **Local storage**

  /workdir

  /SSD

  /local

Local Storage Note

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

Local storage on <u>some hosted</u> servers (e.g. ECCO) is not persistent per owner request

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

Local directories are cleaned at 3am every night.

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

If the machine is used by other users, the files may be deleted on the first night …

/workdir full

What if my local storage /workdir is full on my reserved server?

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

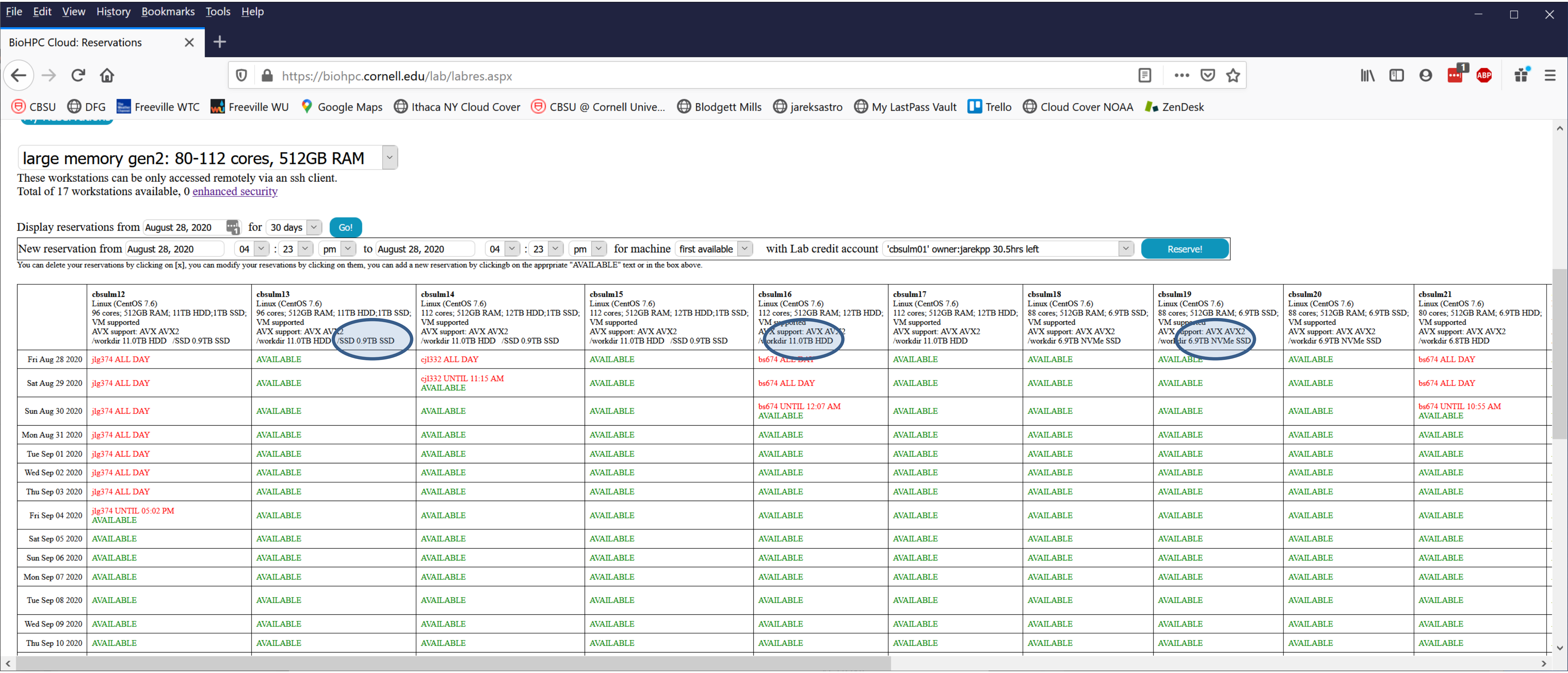
You can run this script yourself:

```
/programs/config/clean_workdir
```

This way 5-day rule will be ignored and all data from previous reservations will be deleted.

Local storage speed

Not all local storage is equally fast. Typically NVMe SSD is fastest, followed by regular SSD, then RAID HDD and lastly single HDD. Check top row on reservation page!



large memory gen2: 80-112 cores, 512GB RAM

These workstations can be only accessed remotely via an ssh client.
Total of 17 workstations available, 0 enhanced security

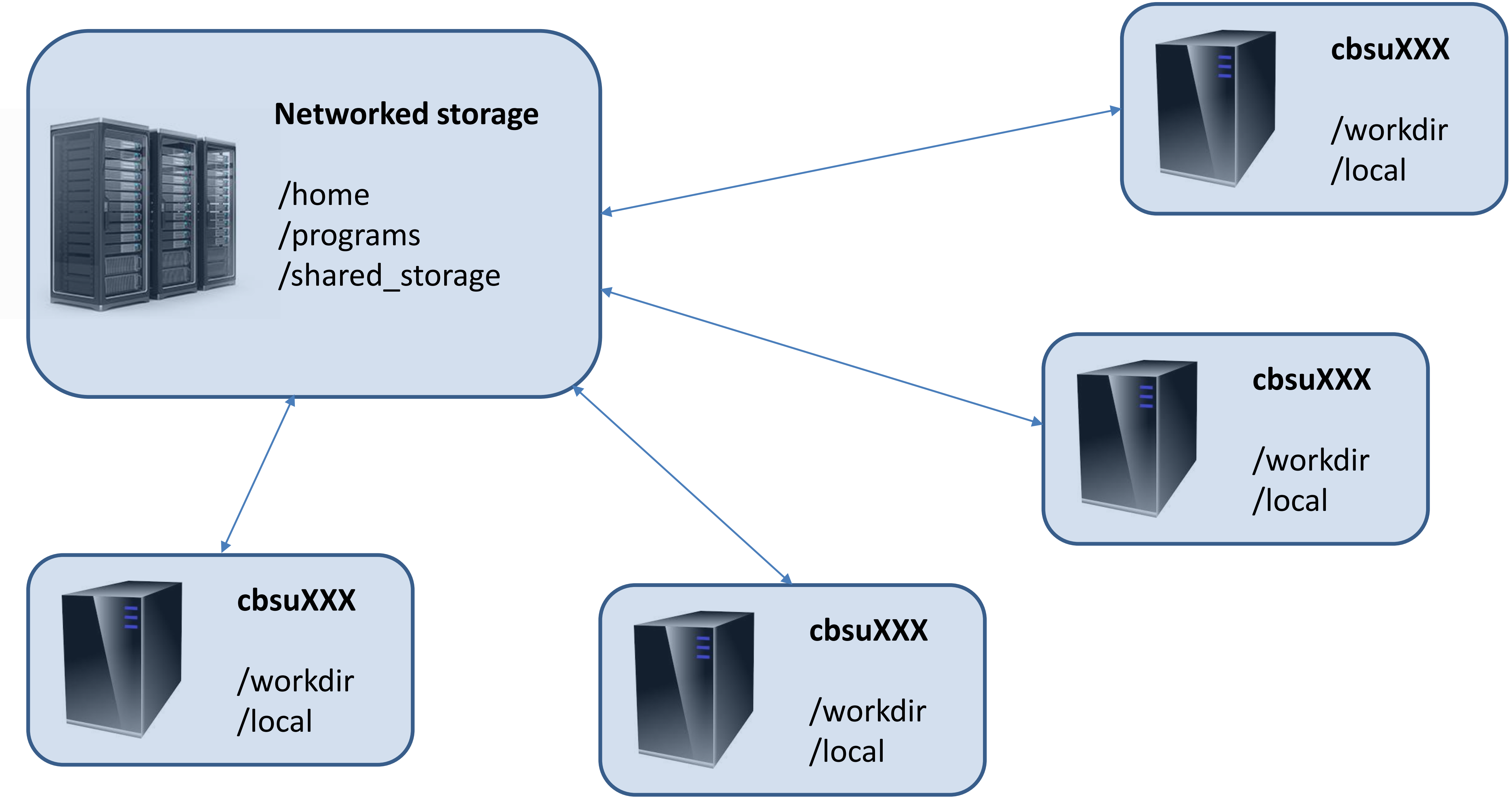
Display reservations from August 28, 2020 for 30 days Go!

New reservation from August 28, 2020 04 : 23 pm to August 28, 2020 04 : 23 pm for machine first available with Lab credit account 'cbsulm01' owner:jarekpp 30.5hrs left 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.

	cbsulm12 Linux (CentOS 7.6) 96 cores; 512GB RAM; 11TB HDD;1TB SSD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD /SSD 0.9TB SSD	**cbsulm13** Linux (CentOS 7.6) 96 cores; 512GB RAM; 11TB HDD;1TB SSD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD /SSD 0.9TB SSD	**cbsulm14** Linux (CentOS 7.6) 112 cores; 512GB RAM; 12TB HDD;1TB SSD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD /SSD 0.9TB SSD	**cbsulm15** Linux (CentOS 7.6) 112 cores; 512GB RAM; 12TB HDD;1TB SSD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD /SSD 0.9TB SSD	**cbsulm16** Linux (CentOS 7.6) 112 cores; 512GB RAM; 12TB HDD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD	**cbsulm17** Linux (CentOS 7.6) 112 cores; 512GB RAM; 12TB HDD; VM supported AVX support: AVX AVX2 /workdir 11.0TB HDD	**cbsulm18** Linux (CentOS 7.6) 88 cores; 512GB RAM; 6.9TB SSD; VM supported AVX support: AVX AVX2 /workdir 6.9TB NVMe SSD	**cbsulm19** Linux (CentOS 7.6) 88 cores; 512GB RAM; 6.9TB SSD; VM supported AVX support: AVX AVX2 /workdir 6.9TB NVMe SSD	**cbsulm20** Linux (CentOS 7.6) 88 cores; 512GB RAM; 6.9TB SSD; VM supported AVX support: AVX AVX2 /workdir 6.9TB NVMe SSD	**cbsulm21** Linux (CentOS 7.6) 80 cores; 512GB RAM; 6.9TB HDD; VM supported AVX support: AVX AVX2 /workdir 6.8TB HDD
Fri Aug 28 2020	jlg374 ALL DAY	AVAILABLE	cjl332 ALL DAY	AVAILABLE	bs674 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	bs674 ALL DAY
Sat Aug 29 2020	jlg374 ALL DAY	AVAILABLE	cjl332 UNTIL 11:15 AM AVAILABLE	AVAILABLE	bs674 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	bs674 ALL DAY
Sun Aug 30 2020	jlg374 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	bs674 UNTIL 12:07 AM AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	bs674 UNTIL 10:55 AM AVAILABLE
Mon Aug 31 2020	jlg374 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Tue Sep 01 2020	jlg374 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Wed Sep 02 2020	jlg374 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Thu Sep 03 2020	jlg374 ALL DAY	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Fri Sep 04 2020	jlg374 UNTIL 05:02 PM AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Sat Sep 05 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Sun Sep 06 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Mon Sep 07 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Tue Sep 08 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Wed Sep 09 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE
Thu Sep 10 2020	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE	AVAILABLE

Storage – big picture



Networked storage

/home
/programs
/shared_storage

cbsuXXX

/workdir
/local

cbsuXXX

/workdir
/local

cbsuXXX

/workdir
/local

cbsuXXX

/workdir
/local

Home directories

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

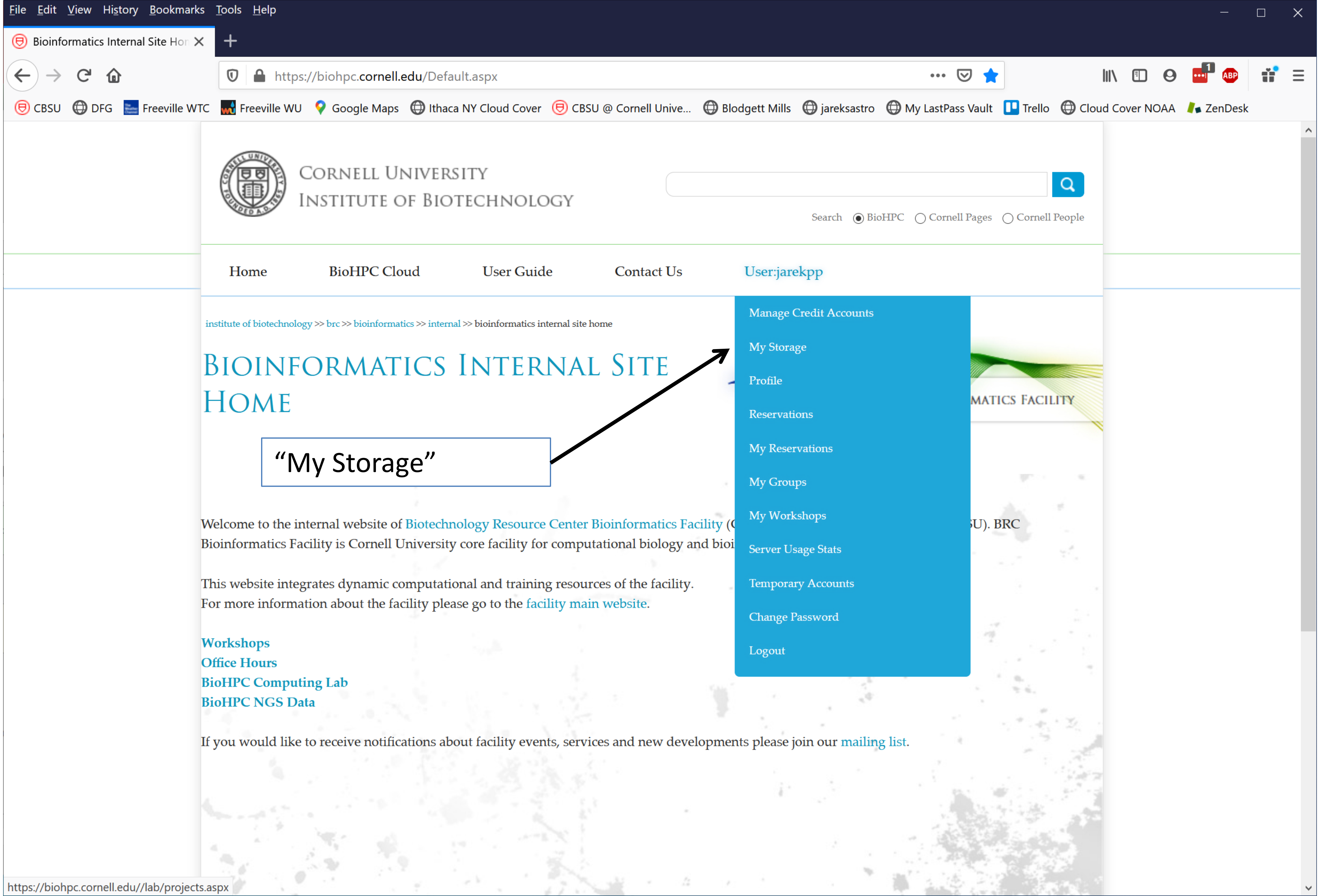
All home directories on BioHPC Cloud are networked and they are the same on all servers.

DO NOT RUN ANY COMPUTATIONS IN YOUR HOME DIRECTORY!

Copy your files to /workdir/mylabid first and run computations there!

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

You can always see your current limits and storage under “My Storage” menu. The storage info is updated daily during the night.



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

BIOINFORMATICS INTERNAL SITE HOME

“My Storage”

Welcome to the internal website of Biotechnology Resource Center Bioinformatics Facility (C U). BRC Bioinformatics Facility is Cornell University core facility for computational biology and bioi

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

Workshops
Office Hours
BioHPC Computing Lab
BioHPC NGS Data

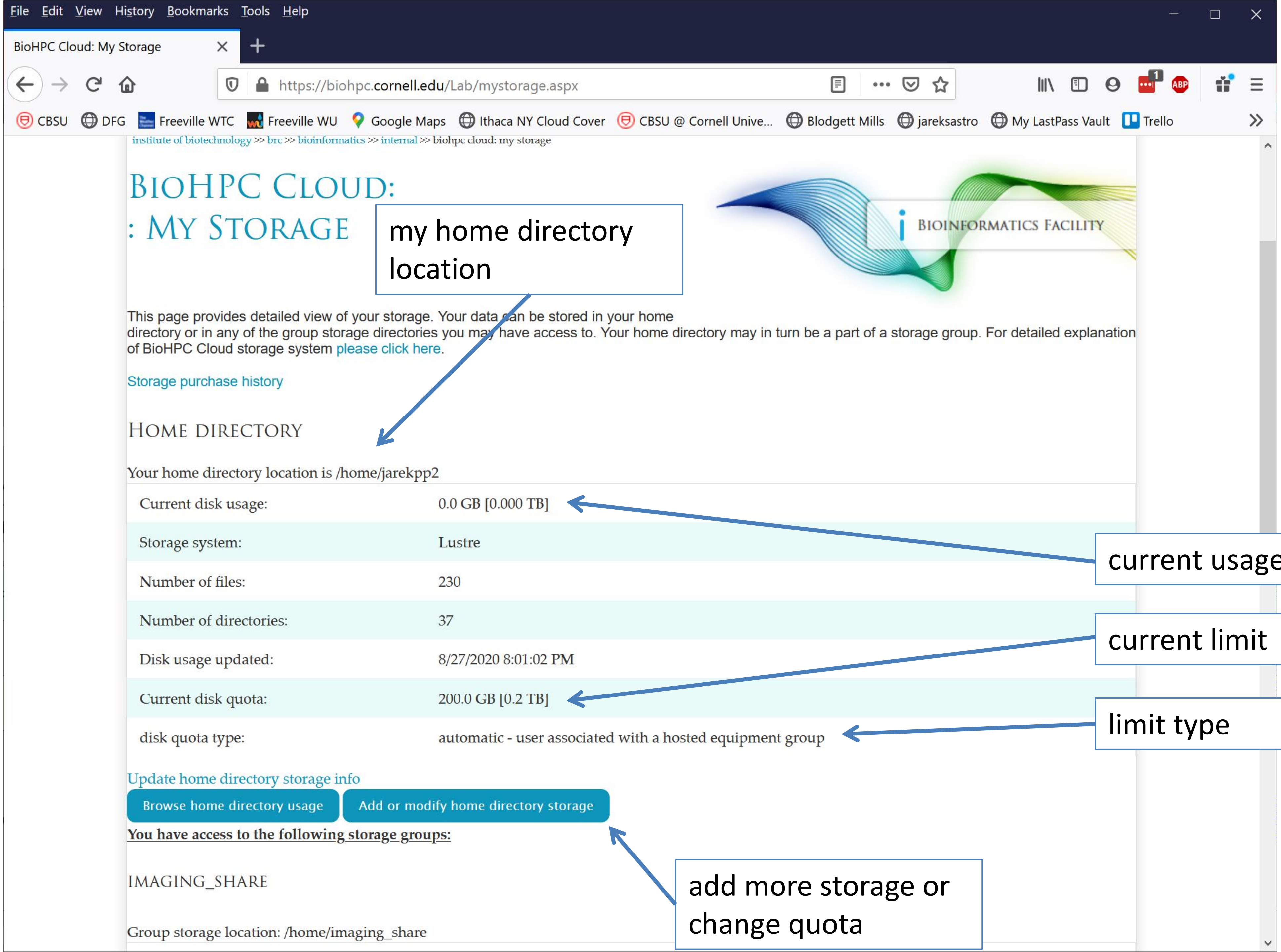
If you would like to receive notifications about facility events, services and new developments please join our mailing list.

Home storage quotas

If user DOES NOT have access to paid storage

- User is associated with an active Lab Credit Account. Home directory storage limit is 200 GB.
- User is associated with an active hosted hardware resource. Home directory storage limit is 200 GB.
- User is NOT associated with an active Lab Credit Account or hosted hardware. Home directory storage limit is 20 GB.

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



BIOHPC CLOUD: : MY STORAGE

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

Storage purchase history

HOME DIRECTORY

Your home directory location is /home/jarekpp2

Current disk usage:	0.0 GB [0.000 TB]
Storage system:	Lustre
Number of files:	230
Number of directories:	37
Disk usage updated:	8/27/2020 8:01:02 PM
Current disk quota:	200.0 GB [0.2 TB]
disk quota type:	automatic - user associated with a hosted equipment group

Update home directory storage info

You have access to the following storage groups:

IMAGING_SHARE

Group storage location: /home/imaging_share

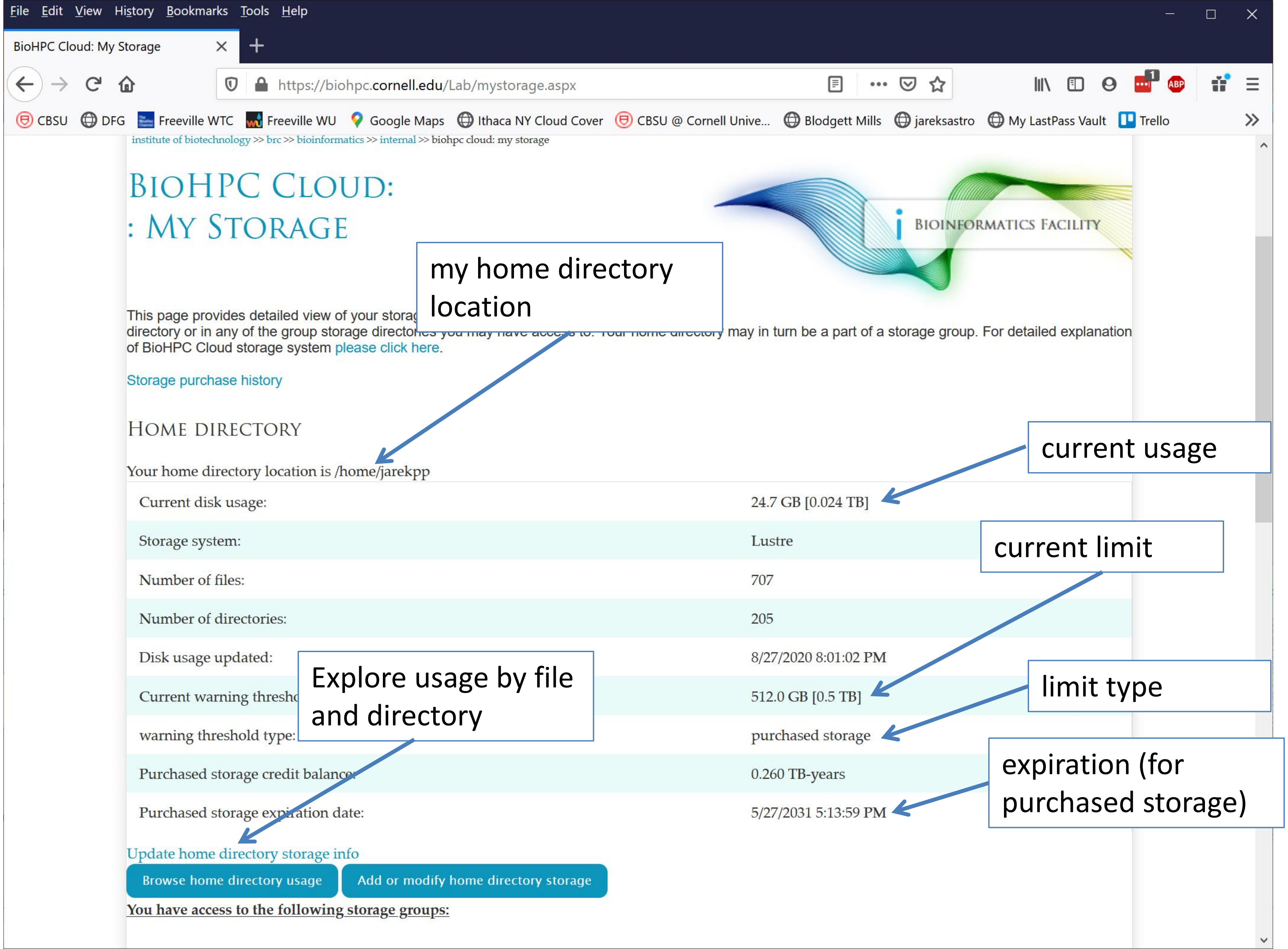
my home directory location

current usage

current limit

limit type

add more storage or change quota



BIOHPC CLOUD: : MY STORAGE

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

Storage purchase history

HOME DIRECTORY

Your home directory location is /home/jarekpp

Current disk usage:	24.7 GB [0.024 TB]
Storage system:	Lustre
Number of files:	707
Number of directories:	205
Disk usage updated:	8/27/2020 8:01:02 PM
Current warning thresho	512.0 GB [0.5 TB]
warning threshold type:	purchased storage
Purchased storage credit balance:	0.260 TB-years
Purchased storage expiration date:	5/27/2031 5:13:59 PM

Update home directory storage info

You have access to the following storage groups:

my home directory location

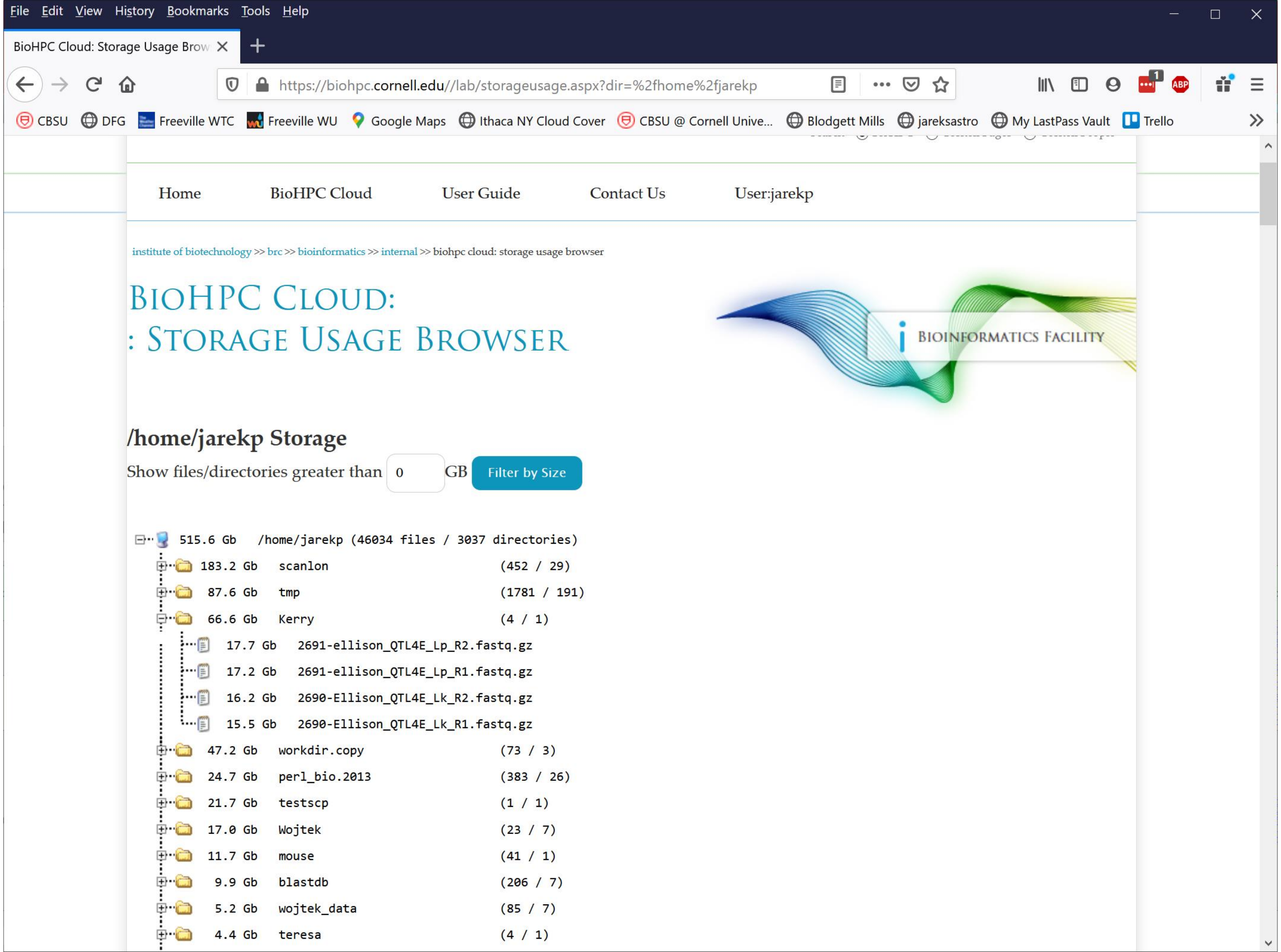
current usage

current limit

limit type

Explore usage by file and directory

expiration (for purchased storage)



https://biohpc.cornell.edu//lab/storageusage.aspx?dir=%2fhome%2fjarekp

institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: storage usage browser

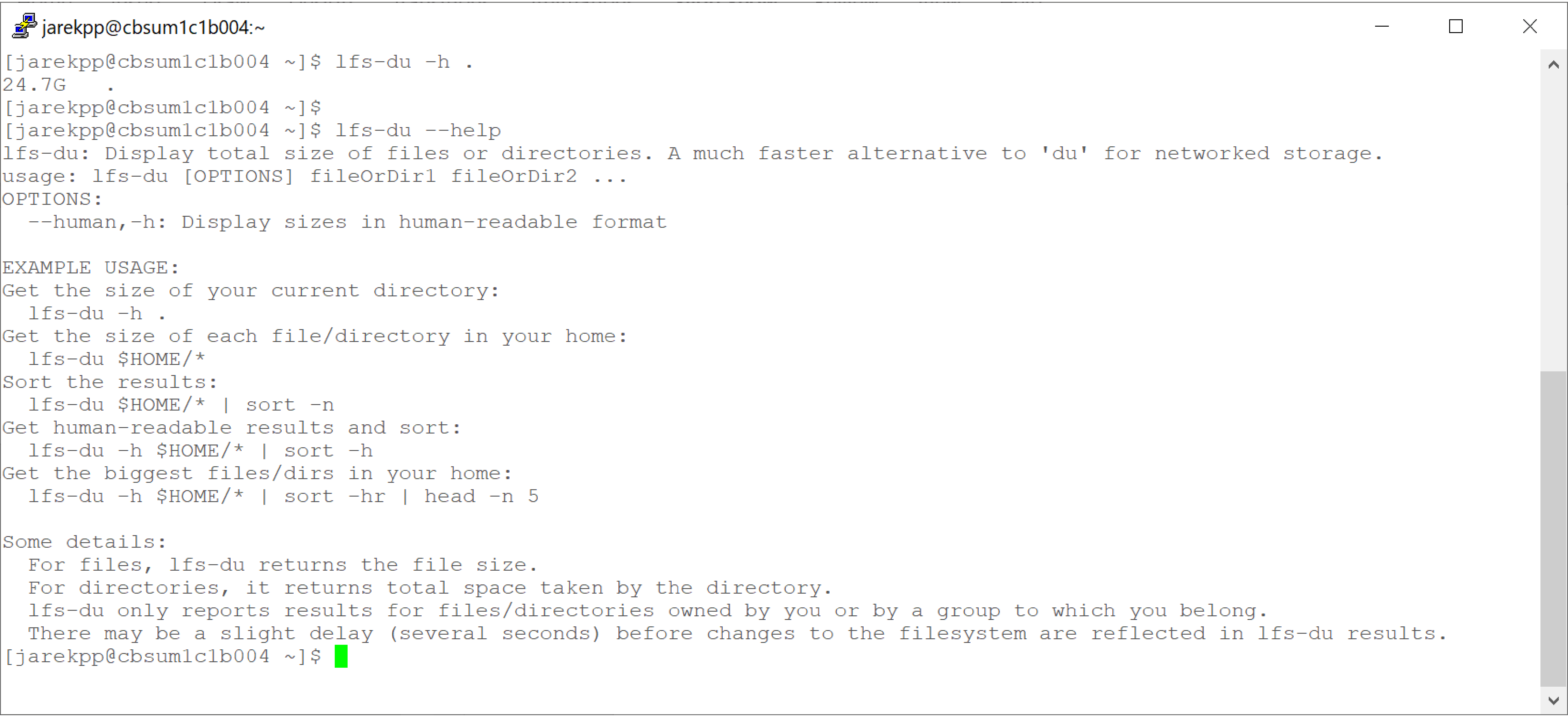
BIOHPC CLOUD:
: STORAGE USAGE BROWSER

/home/jarekp Storage

Show files/directories greater than 0 GB

```
515.6 Gb   /home/jarekp (46034 files / 3037 directories)
   183.2 Gb   scanlon                         (452 / 29)
    87.6 Gb   tmp                             (1781 / 191)
    66.6 Gb   Kerry                           (4 / 1)
       17.7 Gb   2691-ellison_QTL4E_Lp_R2.fastq.gz
       17.2 Gb   2691-ellison_QTL4E_Lp_R1.fastq.gz
       16.2 Gb   2690-Ellison_QTL4E_Lk_R2.fastq.gz
       15.5 Gb   2690-Ellison_QTL4E_Lk_R1.fastq.gz
    47.2 Gb   workdir.copy                    (73 / 3)
    24.7 Gb   perl_bio.2013                   (383 / 26)
    21.7 Gb   testscp                         (1 / 1)
    17.0 Gb   Wojtek                          (23 / 7)
    11.7 Gb   mouse                           (41 / 1)
     9.9 Gb   blastdb                         (206 / 7)
     5.2 Gb   wojtek_data                     (85 / 7)
     4.4 Gb   teresa                          (4 / 1)
```

lfs-du : fast command line tool to check sizes on **networked storage** (Lustre)



```
[jarekpp@cbsum1c1b004 ~]$ lfs-du -h .
24.7G   .
[jarekpp@cbsum1c1b004 ~]$
[jarekpp@cbsum1c1b004 ~]$ lfs-du --help
lfs-du: Display total size of files or directories. A much faster alternative to 'du' for networked storage.
usage: lfs-du [OPTIONS] fileOrDir1 fileOrDir2 ...
OPTIONS:
  --human,-h: Display sizes in human-readable format

EXAMPLE USAGE:
Get the size of your current directory:
  lfs-du -h .
Get the size of each file/directory in your home:
  lfs-du $HOME/*
Sort the results:
  lfs-du $HOME/* | sort -n
Get human-readable results and sort:
  lfs-du -h $HOME/* | sort -h
Get the biggest files/dirs in your home:
  lfs-du -h $HOME/* | sort -hr | head -n 5

Some details:
  For files, lfs-du returns the file size.
  For directories, it returns total space taken by the directory.
  lfs-du only reports results for files/directories owned by you or by a group to which you belong.
  There may be a slight delay (several seconds) before changes to the filesystem are reflected in lfs-du results.
[jarekpp@cbsum1c1b004 ~]$ 
```

Storage

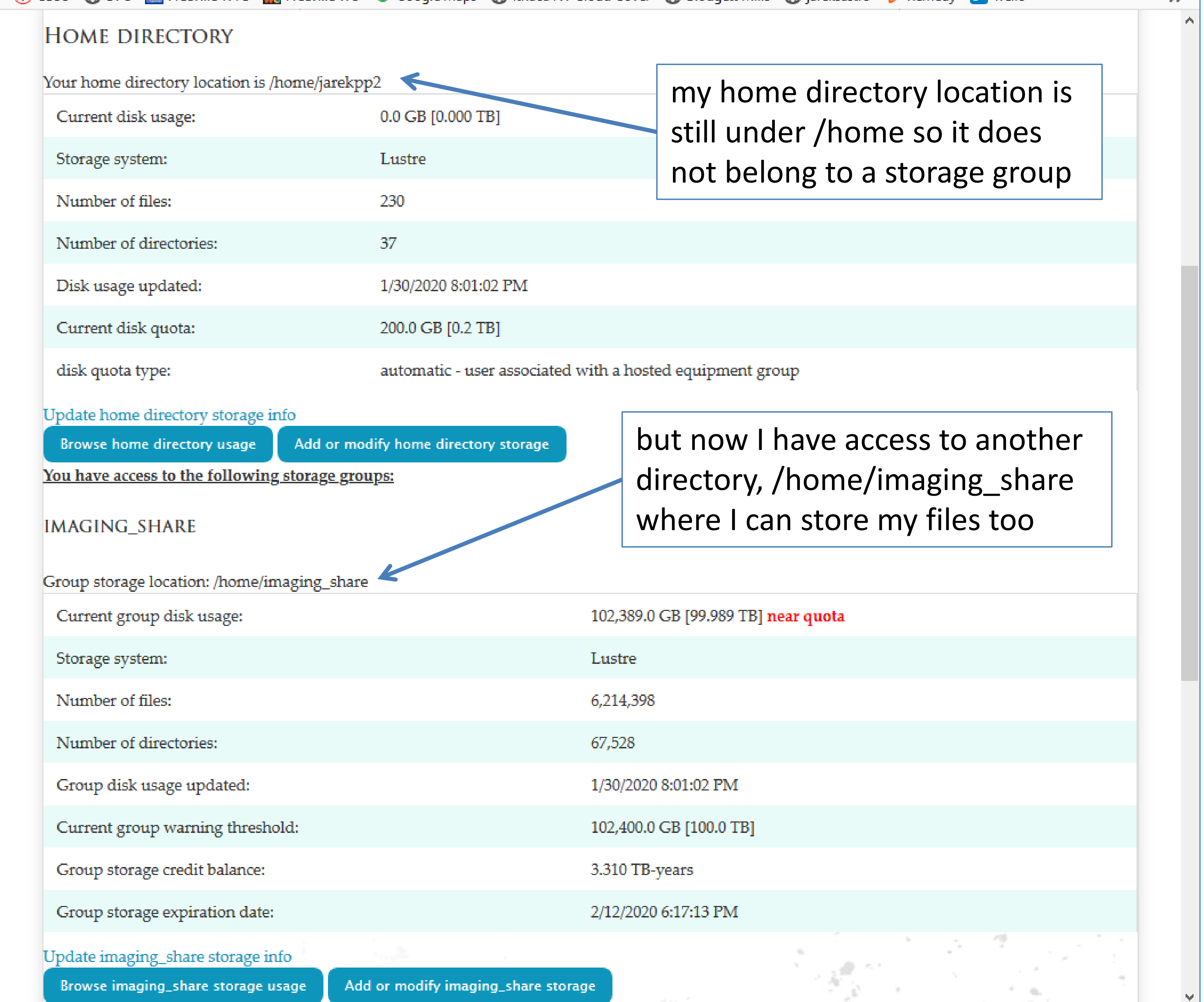
Extra storage can be purchased for $98 per TB per year.

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”



HOME DIRECTORY

Your home directory location is /home/jarekpp2

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

Update home directory storage info

You have access to the following storage groups:

IMAGING_SHARE

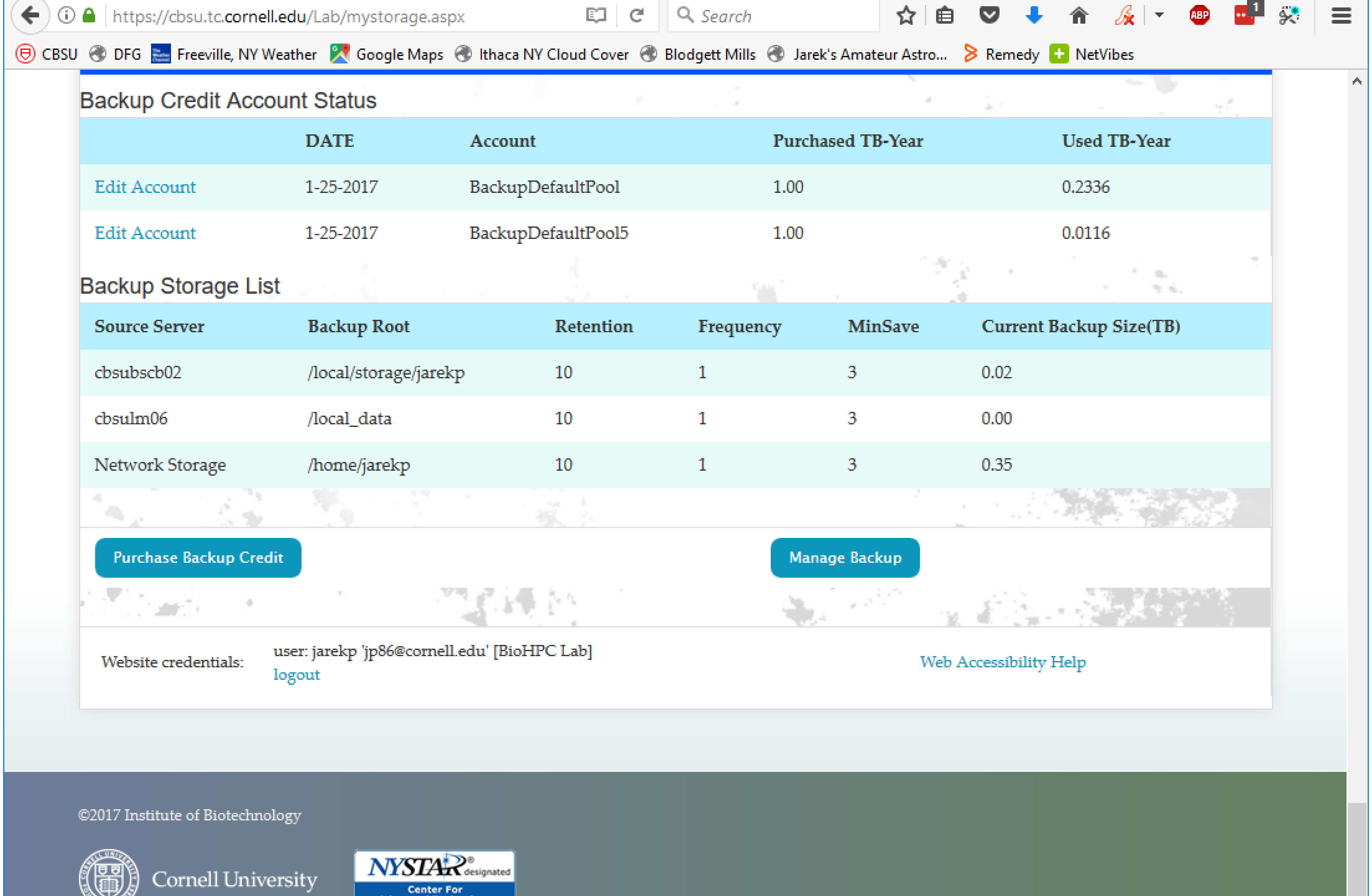
Group storage location: /home/imaging_share

Current group disk usage:	102,389.0 GB [99.989 TB] **near quota**
Storage system:	Lustre
Number of files:	6,214,398
Number of directories:	67,528
Group disk usage updated:	1/30/2020 8:01:02 PM
Current group warning threshold:	102,400.0 GB [100.0 TB]
Group storage credit balance:	3.310 TB-years
Group storage expiration date:	2/12/2020 6:17:13 PM

Update imaging_share storage info

my home directory location is still under /home so it does not belong to a storage group

but now I have access to another directory, /home/imaging_share where I can store my files too



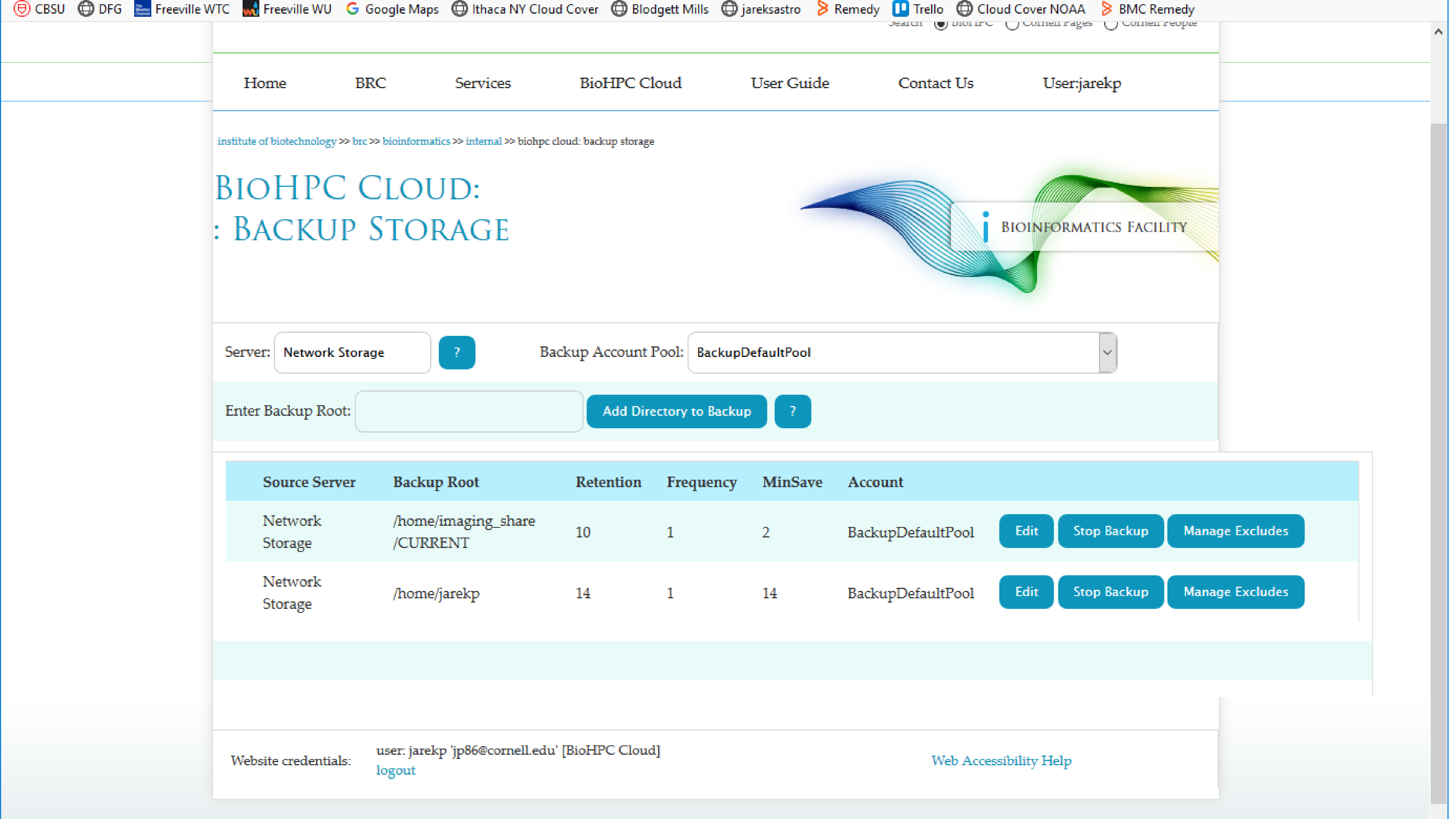
Backup Credit Account Status

	DATE	Account	Purchased TB-Year	Used TB-Year
Edit Account	1-25-2017	BackupDefaultPool	1.00	0.2336
Edit Account	1-25-2017	BackupDefaultPool5	1.00	0.0116

Backup Storage List

Source Server	Backup Root	Retention	Frequency	MinSave	Current Backup Size(TB)
cbsubscb02	/local/storage/jarekp	10	1	3	0.02
cbsulm06	/local_data	10	1	3	0.00
Network Storage	/home/jarekp	10	1	3	0.35

Website credentials: user: jarekp 'jp86@cornell.edu' [BioHPC Lab]
logout



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: backup storage

BIOHPC CLOUD:
: BACKUP STORAGE

Server: Network Storage

Backup Account Pool: BackupDefaultPool

Enter Backup Root:

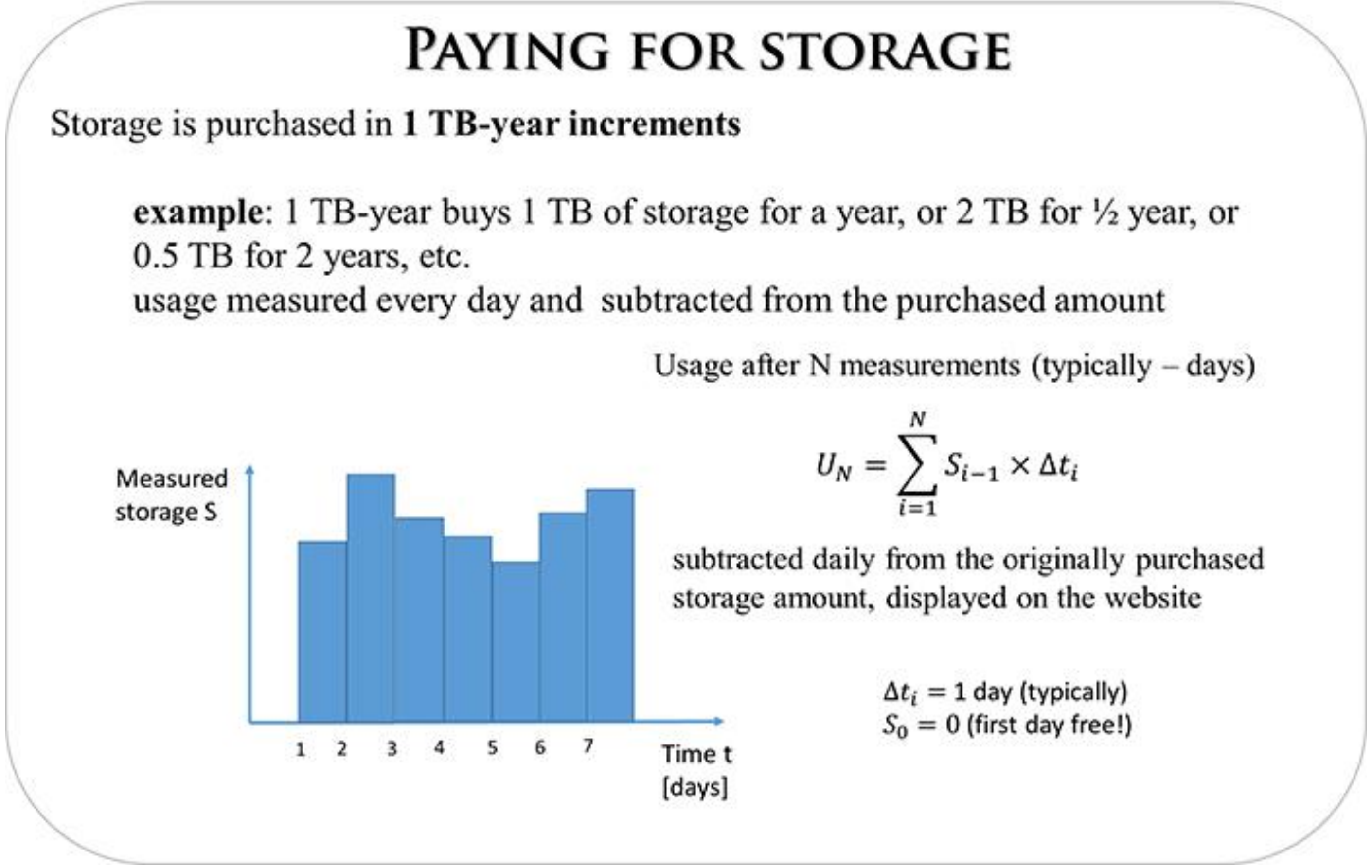
Source Server	Backup Root	Retention	Frequency	MinSave	Account			
Network Storage	/home/imaging_share /CURRENT	10	1	2	BackupDefaultPool	Edit	Stop Backup	Manage Excludes
Network Storage	/home/jarekp	14	1	14	BackupDefaultPool	Edit	Stop Backup	Manage Excludes

Website credentials: user: jarekp 'jp86@cornell.edu' [BioHPC Cloud]
logout

Web Accessibility Help

Storage

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



PAYING FOR STORAGE

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

example: 1 TB-year buys 1 TB of storage for a year, or 2 TB for ½ year, or 0.5 TB for 2 years, etc.
usage measured every day and subtracted from the purchased amount

Usage after N measurements (typically – days)

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

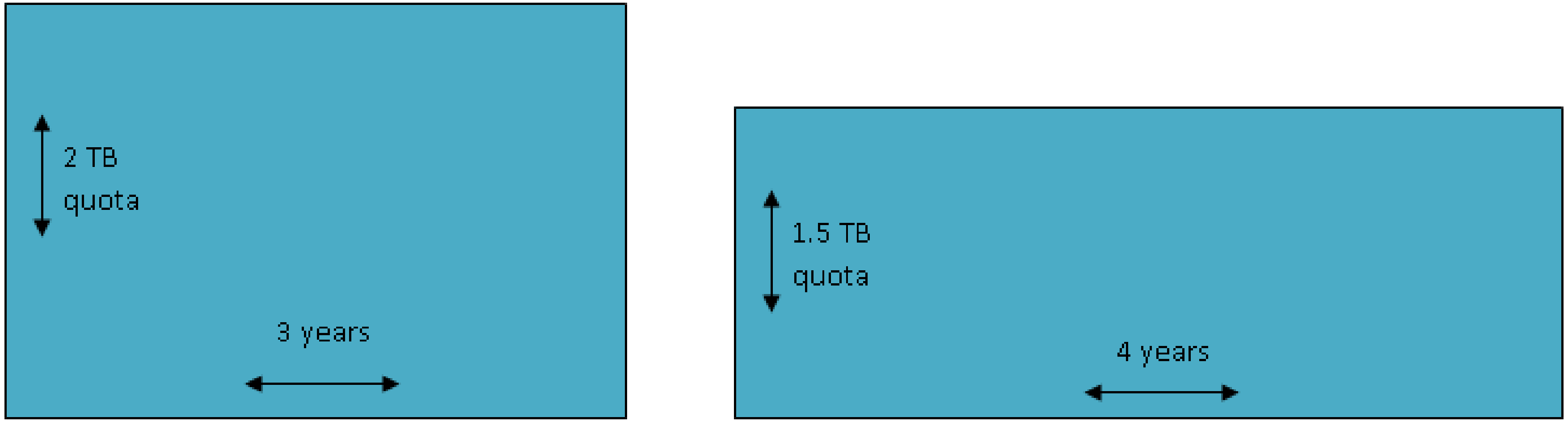
subtracted daily from the originally purchased storage amount, displayed on the website

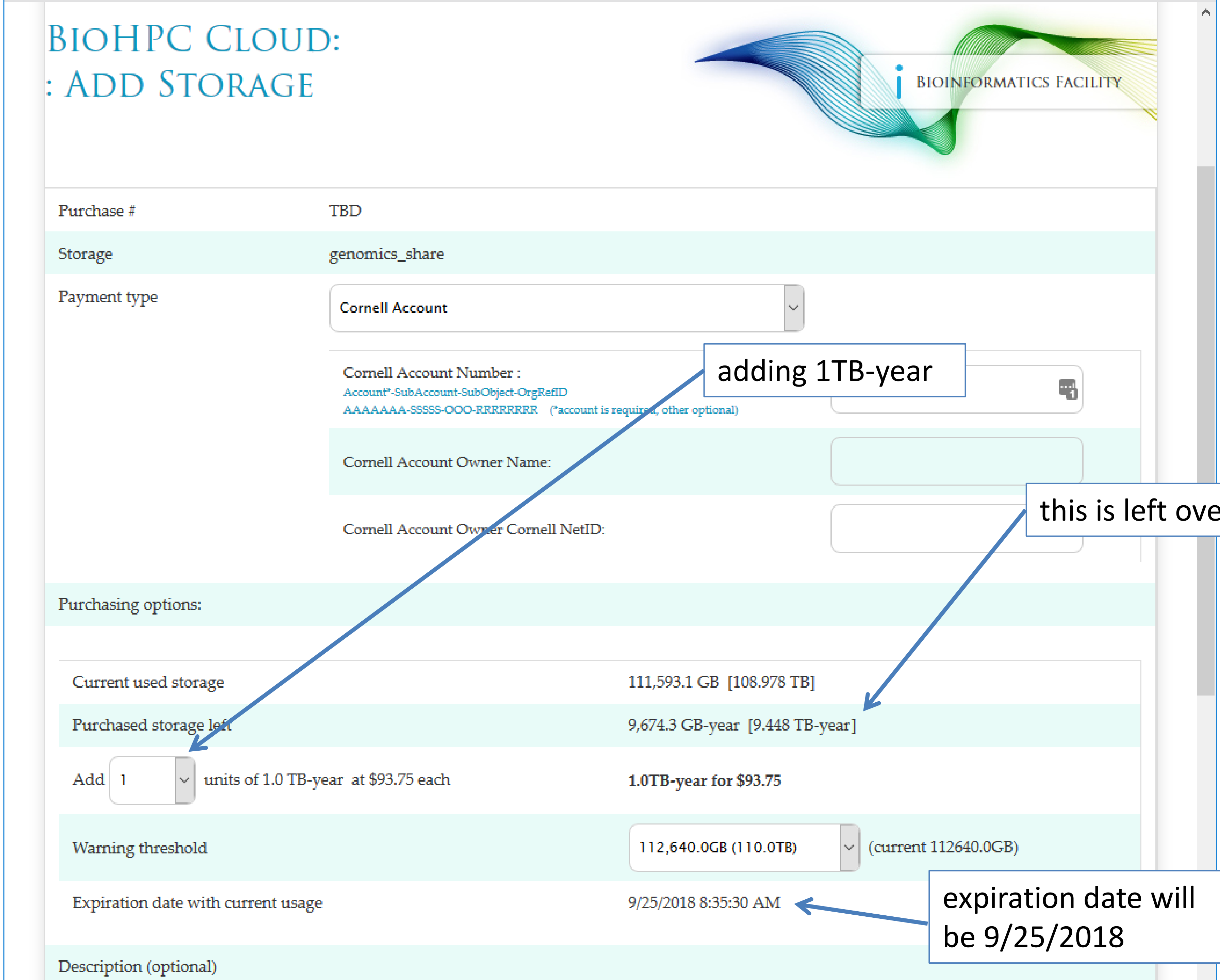
$\Delta t_i = 1$ day (typically)
$S_0 = 0$ (first day free!)

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.





BIOHPC CLOUD:
: ADD STORAGE

Purchase #	TBD
Storage	genomics_share
Payment type	Cornell Account

Cornell Account Number :
Account*-SubAccount-SubObject-OrgRefID
AAAAAAA-SSSSS-OOO-RRRRRRRR (*account is required, other optional)

Cornell Account Owner Name:

Cornell Account Owner Cornell NetID:

Purchasing options:

Current used storage	111,593.1 GB [108.978 TB]
Purchased storage left	9,674.3 GB-year [9.448 TB-year]
Add 1 units of 1.0 TB-year at $93.75 each	**1.0TB-year for $93.75**
Warning threshold	112,640.0GB (110.0TB) (current 112640.0GB)
Expiration date with current usage	9/25/2018 8:35:30 AM

Description (optional)

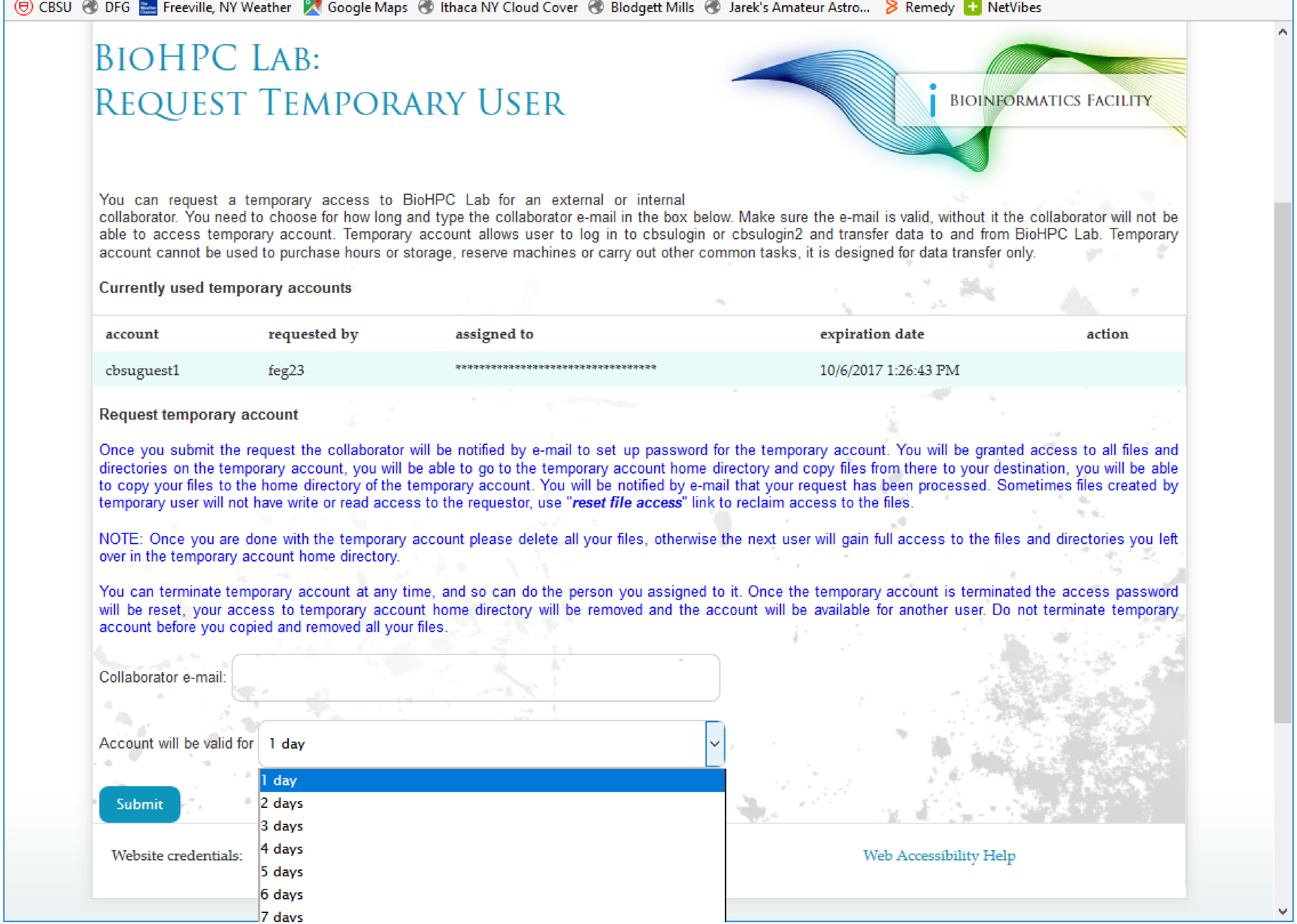
adding 1TB-year

this is left over from past

expiration date will be 9/25/2018

Sharing data with external collaborators

- Request temporary BioHPC account for data sharing. This allows external collaborators both inbound and outbound data transfer.
- Use Globus – you can share for reading any of your files or directories provided the other person has Globus account
- 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 cbsulogin or cbsulogin2 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

account	requested by	assigned to	expiration date	action
cbsuguest1	feg23	********************************	10/6/2017 1:26:43 PM	

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

Collaborator e-mail:

Account will be valid for 1 day

Website credentials:

Web Accessibility Help

Getting started with a new account

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

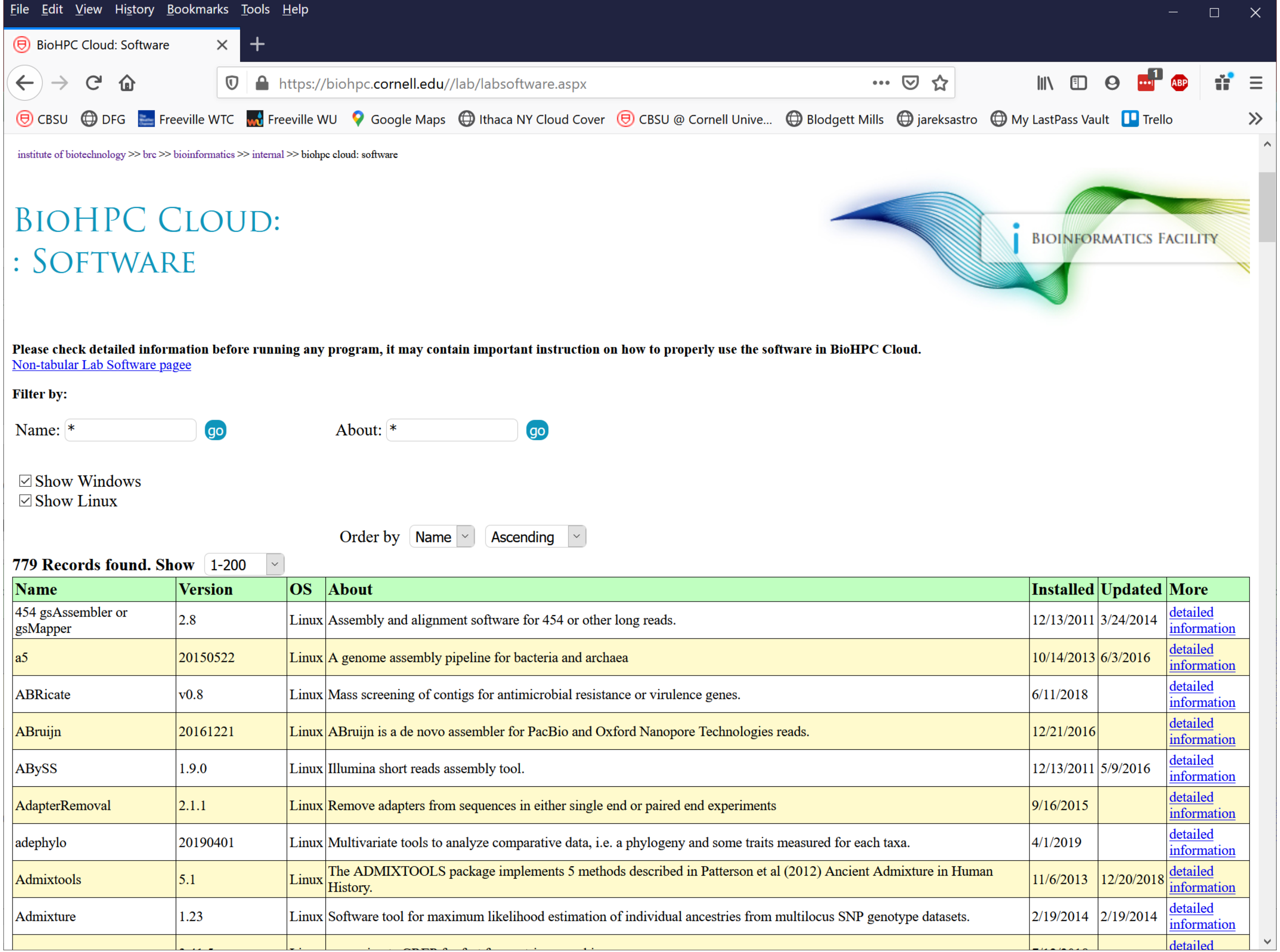
Software

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

https://biohpc.cornell.edu/lab/labsoftware.aspx

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

Some programs require special setup to run, some need to be run with full path, some are installed only on a subset of workstations.



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

BIOHPC CLOUD: : SOFTWARE

Please check detailed information before running any program, it may contain important instruction on how to properly use the software in BioHPC Cloud.

Non-tabular Lab Software pagee

Filter by:

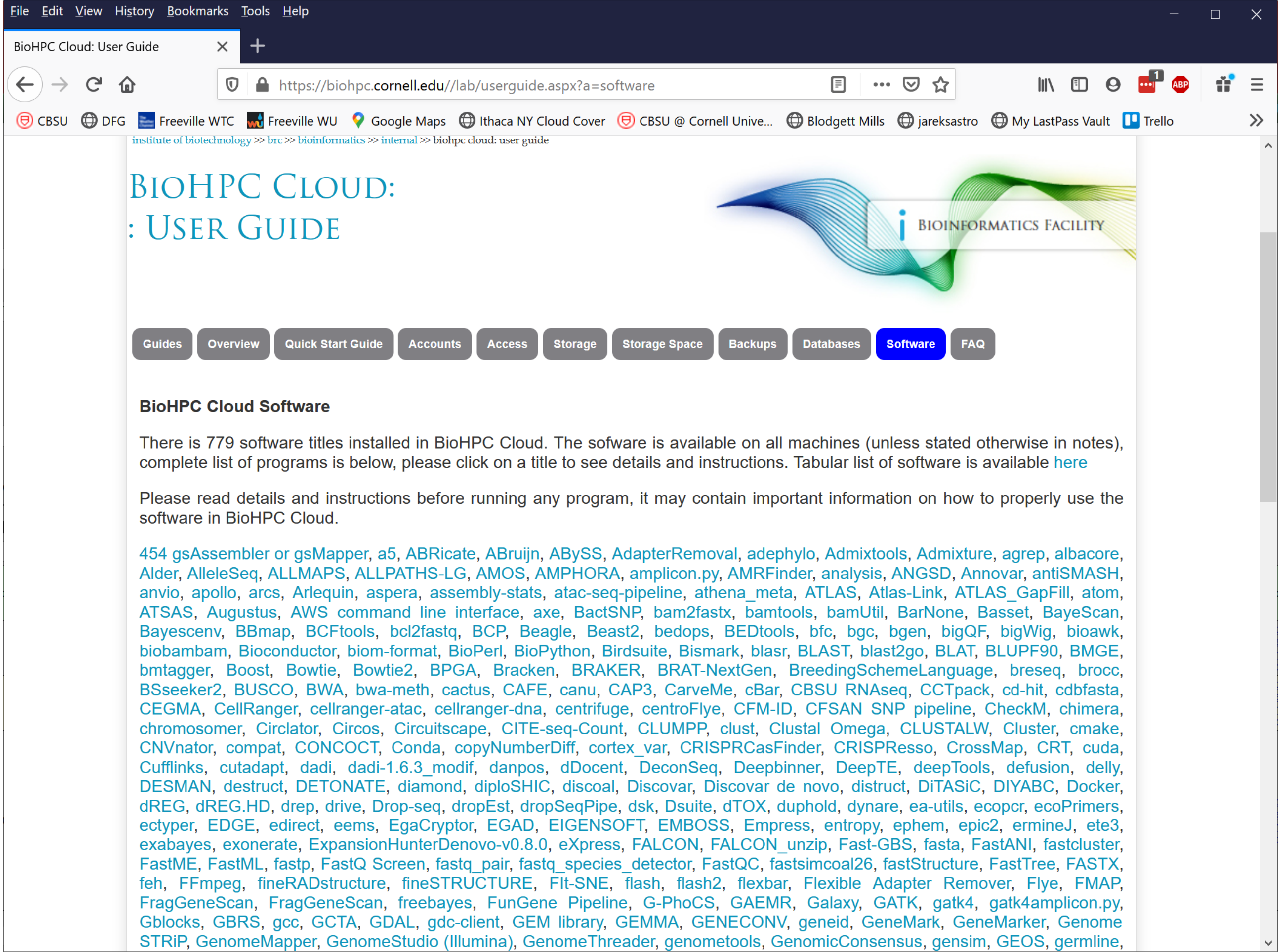
Name: * go About: * go

☑ Show Windows
☑ Show Linux

Order by Name Ascending

779 Records found. Show 1-200

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	20150522	Linux	A genome assembly pipeline for bacteria and archaea	10/14/2013	6/3/2016	detailed information
ABRicate	v0.8	Linux	Mass screening of contigs for antimicrobial resistance or virulence genes.	6/11/2018		detailed information
ABruijn	20161221	Linux	ABruijn is a de novo assembler for PacBio and Oxford Nanopore Technologies reads.	12/21/2016		detailed information
ABySS	1.9.0	Linux	Illumina short reads assembly tool.	12/13/2011	5/9/2016	detailed information
AdapterRemoval	2.1.1	Linux	Remove adapters from sequences in either single end or paired end experiments	9/16/2015		detailed information
adephylo	20190401	Linux	Multivariate tools to analyze comparative data, i.e. a phylogeny and some traits measured for each taxa.	4/1/2019		detailed information
Admixtools	5.1	Linux	The ADMIXTOOLS package implements 5 methods described in Patterson et al (2012) Ancient Admixture in Human History.	11/6/2013	12/20/2018	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



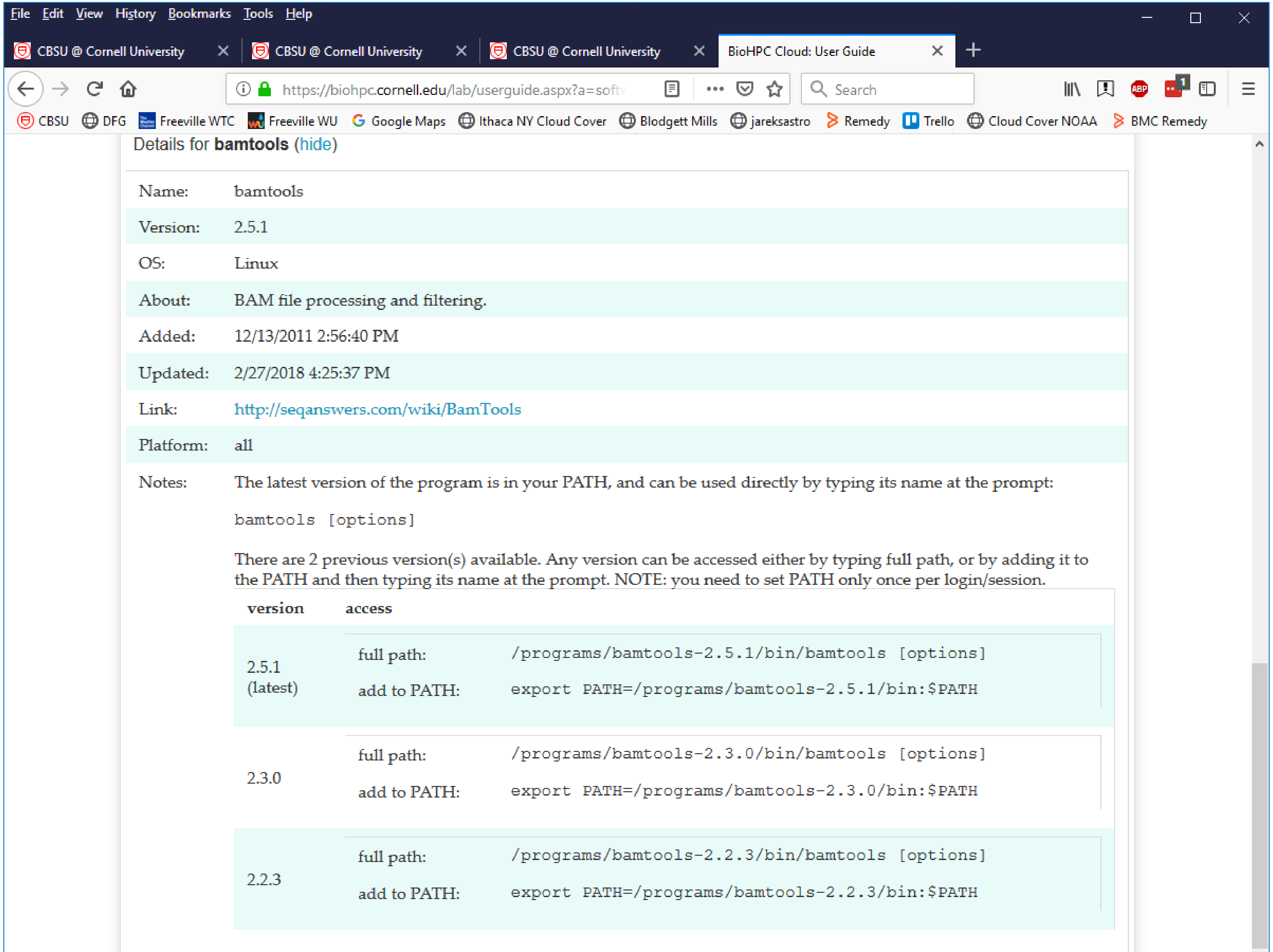
BIOHPC CLOUD: : USER GUIDE

BioHPC Cloud Software

There is 779 software titles installed in BioHPC Cloud. 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 Cloud.

454 gsAssembler or gsMapper, a5, ABRicate, ABruijn, ABySS, AdapterRemoval, adephylo, Admixtools, Admixture, agrep, albacore, Alder, AlleleSeq, ALLMAPS, ALLPATHS-LG, AMOS, AMPHORA, amplicon.py, AMRFinder, analysis, ANGSD, Annovar, antiSMASH, anvio, apollo, arcs, Arlequin, aspera, assembly-stats, atac-seq-pipeline, athena_meta, ATLAS, Atlas-Link, ATLAS_GapFill, atom, ATSAS, Augustus, AWS command line interface, axe, BactSNP, bam2fastx, bamtools, bamUtil, BarNone, Basset, BayeScan, Bayescenv, BBmap, BCFtools, bcl2fastq, BCP, Beagle, Beast2, bedops, BEDtools, bfc, bgc, bgen, bigQF, bigWig, bioawk, biobambam, Bioconductor, biom-format, BioPerl, BioPython, Birdsuite, Bismark, blasr, BLAST, blast2go, BLAT, BLUPF90, BMGE, bmtagger, Boost, Bowtie, Bowtie2, BPGA, Bracken, BRAKER, BRAT-NextGen, BreedingSchemeLanguage, breseq, brocc, BSseeker2, BUSCO, BWA, bwa-meth, cactus, CAFE, canu, CAP3, CarveMe, cBar, CBSU RNAseq, CCTpack, cd-hit, cdbfasta, CEGMA, CellRanger, cellranger-atac, cellranger-dna, centrifuge, centroFlye, CFM-ID, CFSAN SNP pipeline, CheckM, chimera, chromosomer, Circlator, Circos, Circuitscape, CITE-seq-Count, CLUMPP, clust, Clustal Omega, CLUSTALW, Cluster, cmake, CNVnator, compat, CONCOCT, Conda, copyNumberDiff, cortex_var, CRISPRCasFinder, CRISPResso, CrossMap, CRT, cuda, Cufflinks, cutadapt, dadi, dadi-1.6.3_modif, danpos, dDocent, DeconSeq, Deepbinner, DeepTE, deepTools, defusion, delly, DESMAN, destruct, DETONATE, diamond, diploSHIC, discoal, Discovar, Discovar de novo, distruct, DiTASiC, DIYABC, Docker, dREG, dREG.HD, drep, drive, Drop-seq, dropEst, dropSeqPipe, dsk, Dsuite, dTOX, duphold, dynare, ea-utils, ecopcr, ecoPrimers, ectyper, EDGE, edirect, eems, EgaCryptor, EGAD, EIGENSOFT, EMBOSS, Empress, entropy, ephem, epic2, ermineJ, ete3, exabayes, exonerate, ExpansionHunterDenovo-v0.8.0, eXpress, FALCON, FALCON_unzip, Fast-GBS, fasta, FastANI, fastcluster, FastME, FastML, fastp, FastQ Screen, fastq_pair, fastq_species_detector, FastQC, fastsimcoal26, fastStructure, FastTree, FASTX, feh, FFmpeg, fineRADstructure, fineSTRUCTURE, FIt-SNE, flash, flash2, flexbar, Flexible Adapter Remover, Flye, FMAP, FragGeneScan, FragGeneScan, freebayes, FunGene Pipeline, G-PhoCS, GAEMR, Galaxy, GATK, gatk4, gatk4amplicon.py, Gblocks, GBRS, gcc, GCTA, GDAL, gdc-client, GEM library, GEMMA, GENECONV, geneid, GeneMark, GeneMarker, Genome STRiP, GenomeMapper, GenomeStudio (Illumina), GenomeThreader, genometools, GenomicConsensus, gensim, GEOS, germline,



Details for **bamtools** (hide)

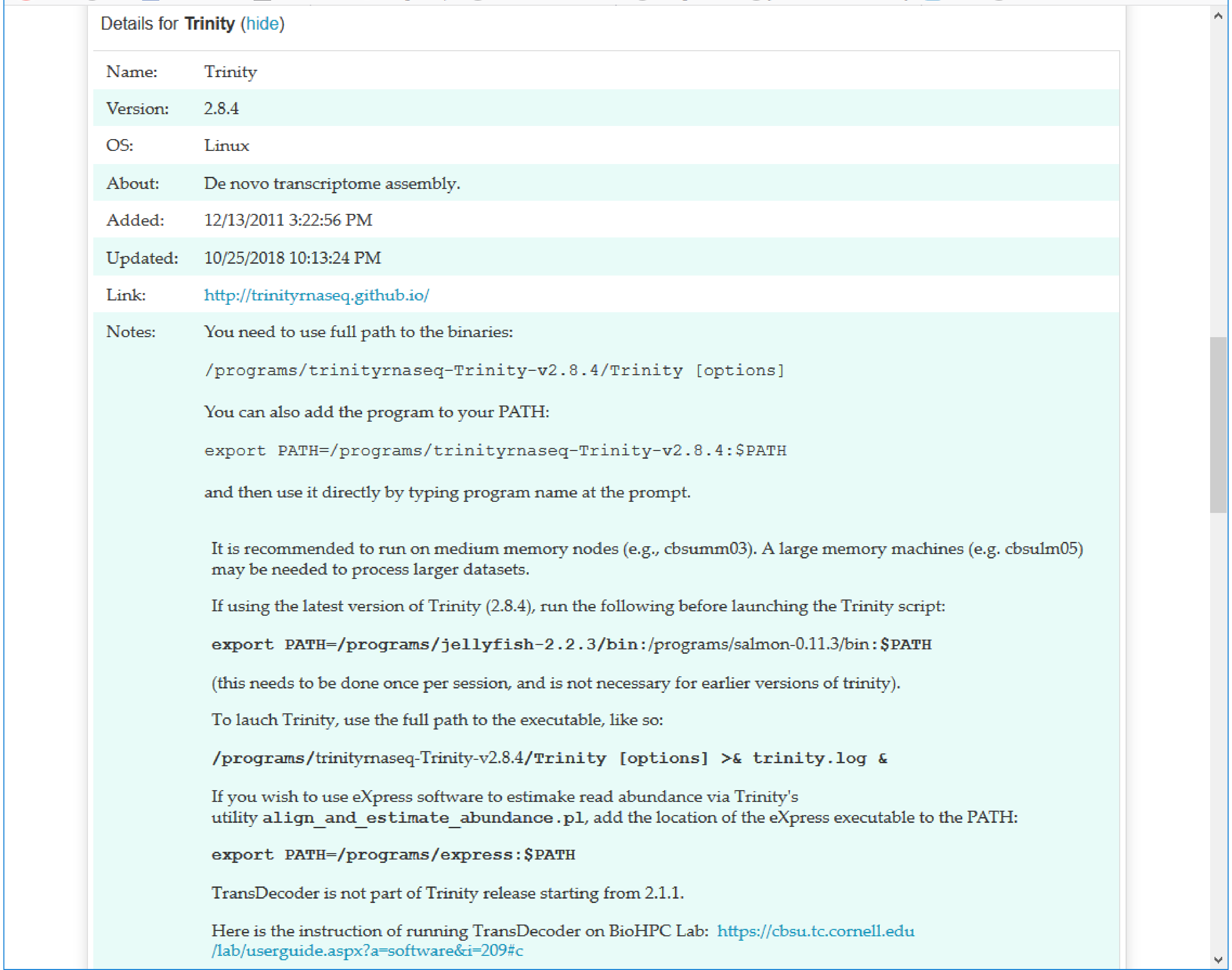
Name:	bamtools
Version:	2.5.1
OS:	Linux
About:	BAM file processing and filtering.
Added:	12/13/2011 2:56:40 PM
Updated:	2/27/2018 4:25:37 PM
Link:	http://seqanswers.com/wiki/BamTools
Platform:	all

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

```
bamtools [options]
```

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

version	access	
2.5.1 (latest)	full path:	`/programs/bamtools-2.5.1/bin/bamtools [options]`
	add to PATH:	`export PATH=/programs/bamtools-2.5.1/bin:$PATH`
2.3.0	full path:	`/programs/bamtools-2.3.0/bin/bamtools [options]`
	add to PATH:	`export PATH=/programs/bamtools-2.3.0/bin:$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:	2.8.4
OS:	Linux
About:	De novo transcriptome assembly.
Added:	12/13/2011 3:22:56 PM
Updated:	10/25/2018 10:13:24 PM
Link:	http://trinityrnaseq.github.io/

Notes:

You need to use full path to the binaries:

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

You can also add the program to your PATH:

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

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

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

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

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

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

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

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

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

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

TransDecoder is not part of Trinity release starting from 2.1.1.

Here is the instruction of running TransDecoder on BioHPC Lab: https://cbsu.tc.cornell.edu/lab/userguide.aspx?a=software&i=209#c

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 BioHPC program page.

Want to install software yourself?

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

- You can use Conda to install software in a separate environment.

- You can download and use Docker image. You can install software inside Docker container.

⇒“Software installation on Linux”

⇒“Using Docker in BioHPC Cloud “

Getting started with a new account

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

Connecting to BioHPC Cloud machines

Text-based connection: ssh (Secure SHell)

GUI (graphical) connection: X-Windows or VNC

SSH

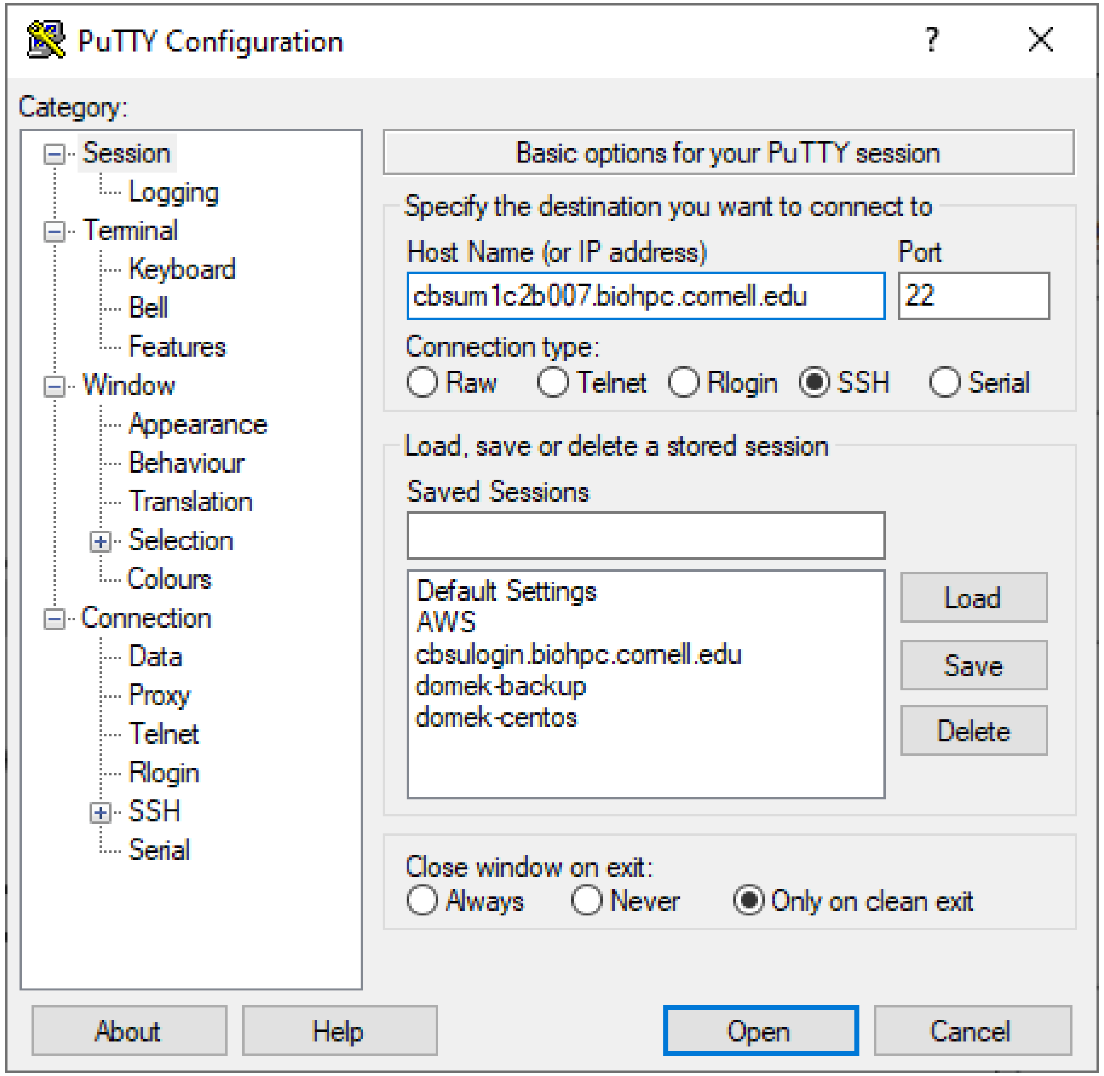
SSH is the simplest and most useful connection.

It is text based, no graphics.

Details are explored in “Linux for Biologists” workshop, but we will cover the basics in this workshop.

SSH - Windows

- Install PuTTY – open source SSH package for Windows
- Start PuTTY
- Type fully qualified server name you want to connect to, e.g. cbsu1c2b007.biohpc.cornell.edu



PuTTY Configuration

Basic options for your PuTTY session

Specify the destination you want to connect to

Host Name (or IP address): cbsum1c2b007.biohpc.cornell.edu

Port: 22

Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

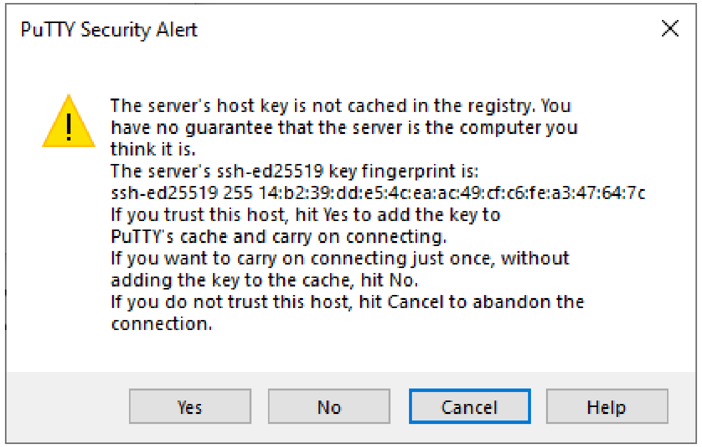
Load, save or delete a stored session

Saved Sessions

Default Settings
AWS
cbsulogin.biohpc.cornell.edu
domek-backup
domek-centos

Close window on exit: ○ Always ○ Never ● Only on clean exit

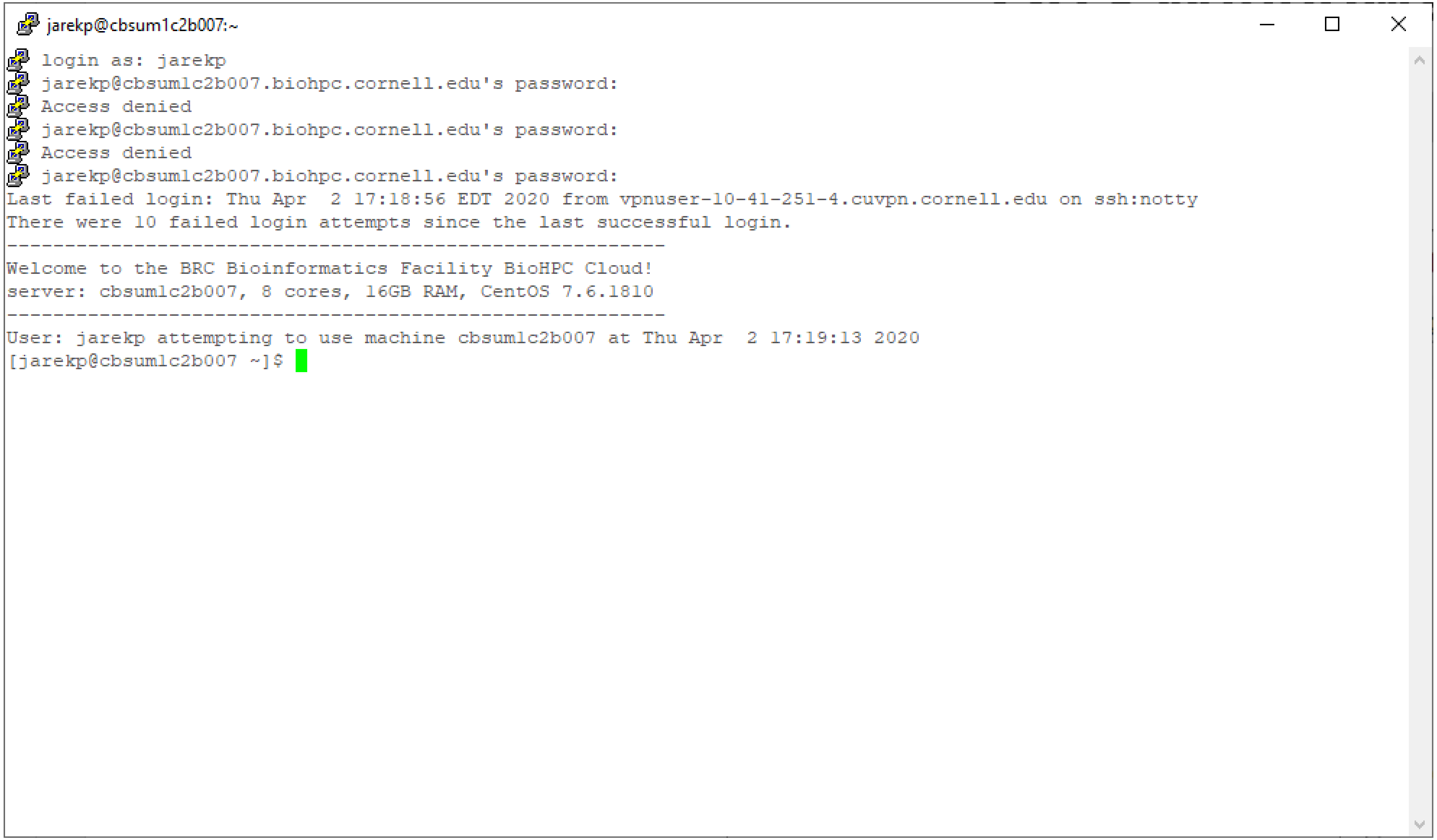
- Click “Open”



PuTTY Security Alert

The server's host key is not cached in the registry. You have no guarantee that the server is the computer you think it is.
The server's ssh-ed25519 key fingerprint is:
ssh-ed25519 255 14:b2:39:dd:e5:4c:ea:ac:49:cf:c6:fe:a3:47:64:7c
If you trust this host, hit Yes to add the key to PuTTY's cache and carry on connecting.
If you want to carry on connecting just once, without adding the key to the cache, hit No.
If you do not trust this host, hit Cancel to abandon the connection.

SSH - Windows

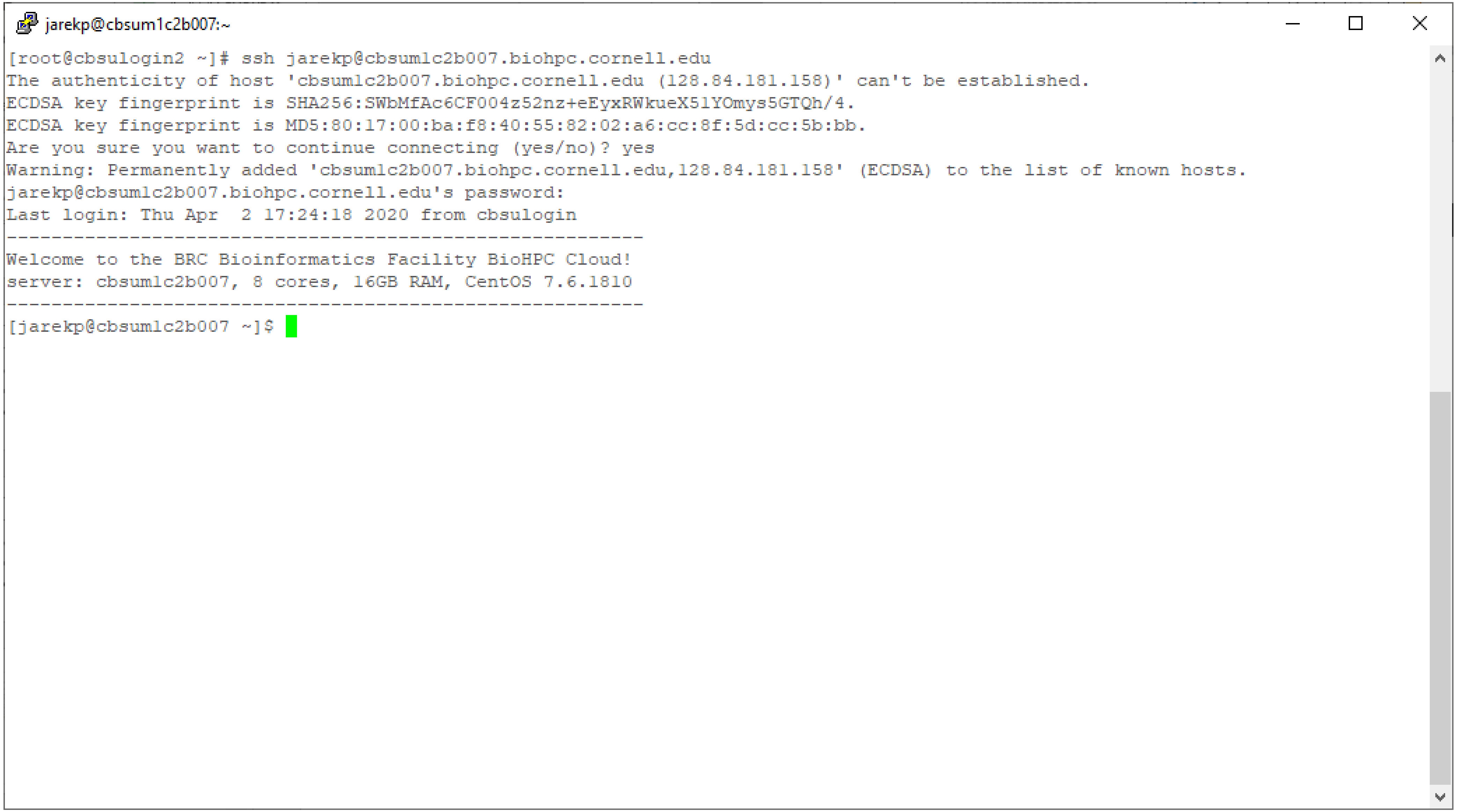


```
login as: jarekp
jarekp@cbsum1c2b007.biohpc.cornell.edu's password:
Access denied
jarekp@cbsum1c2b007.biohpc.cornell.edu's password:
Access denied
jarekp@cbsum1c2b007.biohpc.cornell.edu's password:
Last failed login: Thu Apr  2 17:18:56 EDT 2020 from vpnuser-10-41-251-4.cuvpn.cornell.edu on ssh:notty
There were 10 failed login attempts since the last successful login.
---------------------------------------------------------
Welcome to the BRC Bioinformatics Facility BioHPC Cloud!
server: cbsum1c2b007, 8 cores, 16GB RAM, CentOS 7.6.1810
---------------------------------------------------------
User: jarekp attempting to use machine cbsum1c2b007 at Thu Apr  2 17:19:13 2020
[jarekp@cbsum1c2b007 ~]$
```

SSH – Mac and Linux

- Open a terminal (command window)
- Type
  ssh userid@cbsuxxx.biohpc.cornell.edu
- Provide password when prompted
- Example:
  ssh jarekp@cbsum1c2b007.biohpc.cornell.edu

SSH – Mac and Linux



```
[root@cbsulogin2 ~]# ssh jarekp@cbsumlc2b007.biohpc.cornell.edu
The authenticity of host 'cbsumlc2b007.biohpc.cornell.edu (128.84.181.158)' can't be established.
ECDSA key fingerprint is SHA256:SWbMfAc6CF004z52nz+eEyxRWkueX51YOmys5GTQh/4.
ECDSA key fingerprint is MD5:80:17:00:ba:f8:40:55:82:02:a6:cc:8f:5d:cc:5b:bb.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'cbsumlc2b007.biohpc.cornell.edu,128.84.181.158' (ECDSA) to the list of known hosts.
jarekp@cbsumlc2b007.biohpc.cornell.edu's password:
Last login: Thu Apr  2 17:24:18 2020 from cbsulogin
---------------------------------------------------------
Welcome to the BRC Bioinformatics Facility BioHPC Cloud!
server: cbsumlc2b007, 8 cores, 16GB RAM, CentOS 7.6.1810
---------------------------------------------------------
[jarekp@cbsumlc2b007 ~]$
```

Logging in to CBSU machines from outside of Cornell

Two ways to connect from outside:

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

- ☐ Log in to cbsulogin.biohpc.cornell.edu (or cbsulogin2 or cbsulogin3):

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

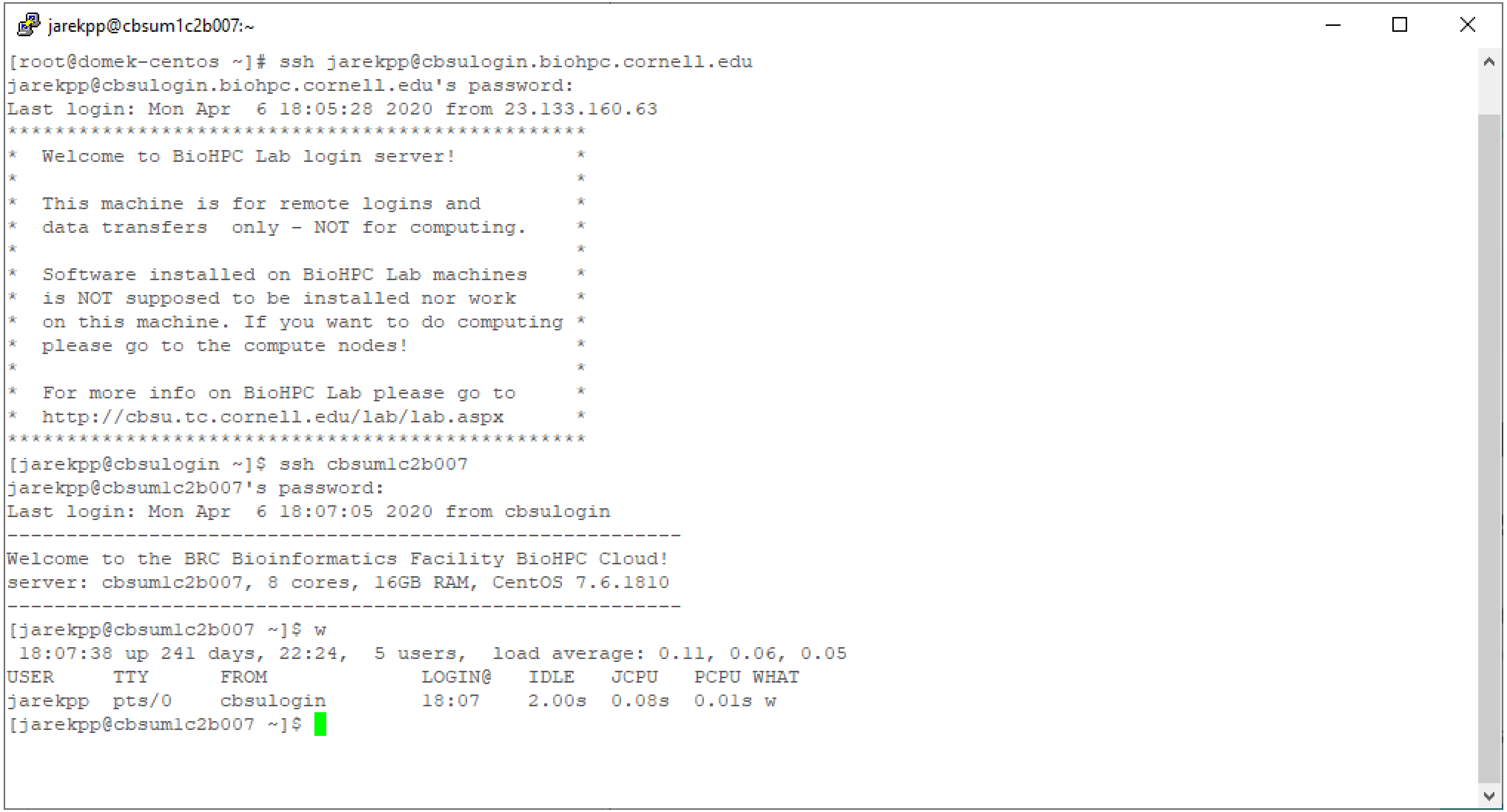
  Once logged in to cbsulogin, ssh further to your reserved machine

  ssh jarekp@cbsum1c2b007.biohpc.cornell.edu

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

https://biohpc.cornell.edu/lab/doc/BioHPCLabexternal.pdf

SSH – connect from outside without VPN



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

Logging out of a Linux machine

- While in terminal window, type **exit** or **Ctrl-D** **-** this will close the current terminal window

Exercise: connect to your assigned workstations using ssh

- Find your assigned machine on the list on workshop page
  https://biohpc.cornell.edu/ww/machines.aspx?i=139

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

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

- Once connected check storage configuration (`df -h`) and who else is there (`w`).

Connecting to BioHPC Cloud

GUI (graphical) connection: VNC

VNC is a remote desktop connection for Linux. It will open full graphical environment the same way one is created when logging in via a graphical console.

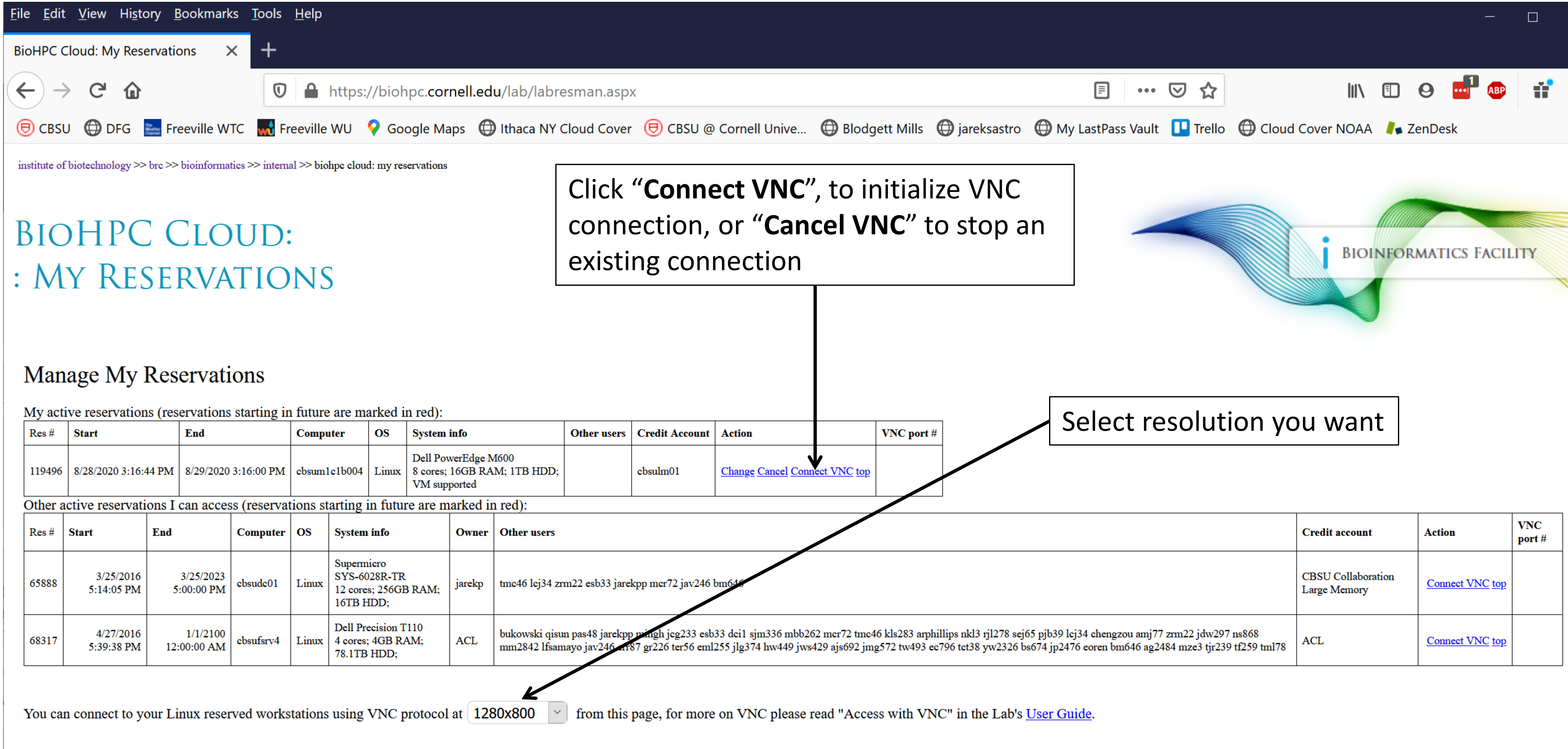
Please note that it is MUCH slower way to connect since lots of graphical data needs to be sent over the network. SSH is much faster and should be used when no graphics is needed.

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

VNC stands for Virtual Network Computing

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

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



institute of biotechnology >> brc >> bioinformatics >> internal >> biohpc cloud: my reservations

BIOHPC CLOUD:
: MY RESERVATIONS

Click “**Connect VNC**”, to initialize VNC connection, or “**Cancel VNC**” to stop an existing connection

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 #
119496	8/28/2020 3:16:44 PM	8/29/2020 3:16:00 PM	cbsum1c1b004	Linux	Dell PowerEdge M600 8 cores; 16GB RAM; 1TB HDD; VM supported		cbsulm01	Change Cancel Connect VNC top	

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

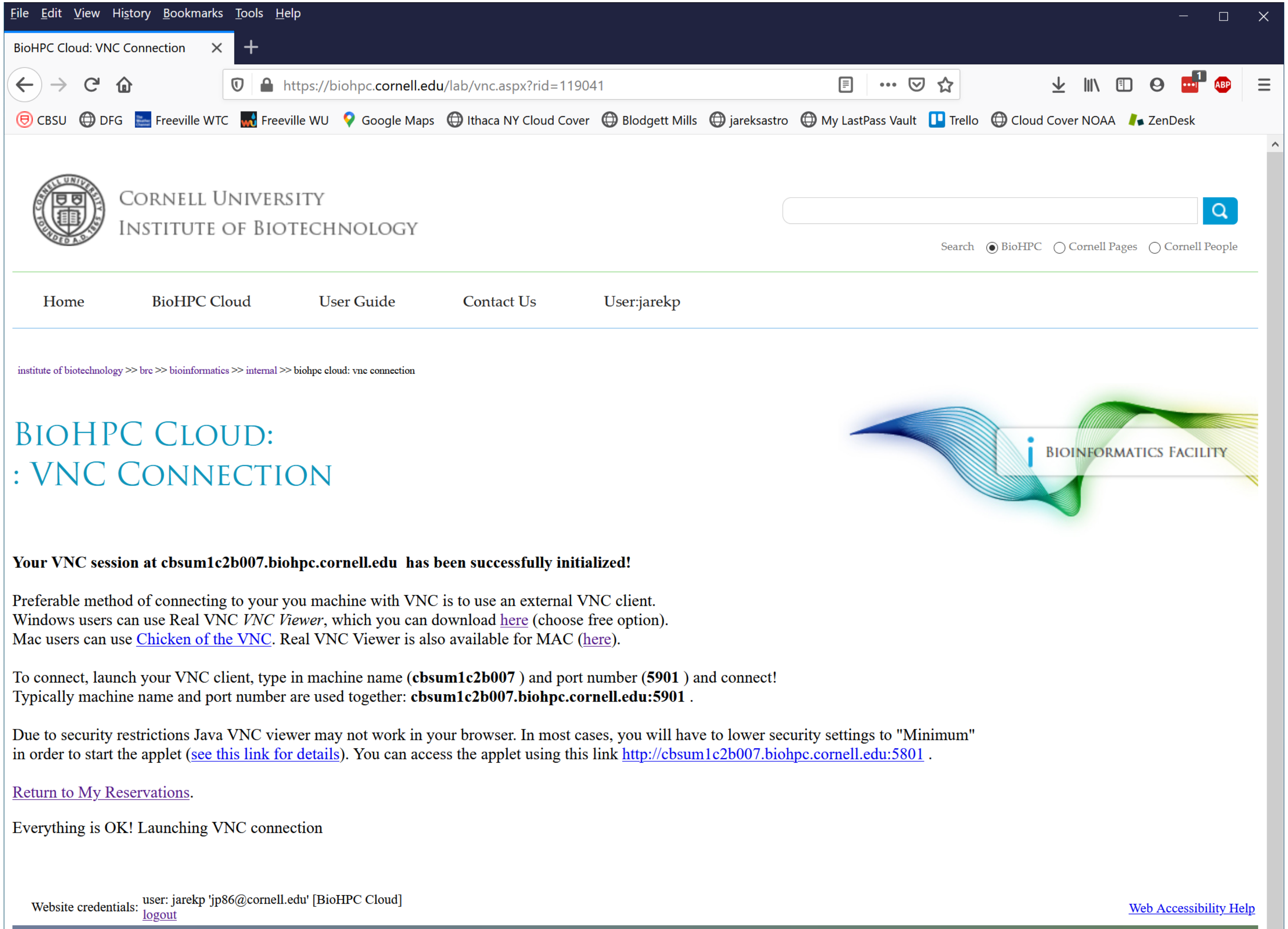
Res #	Start	End	Computer	OS	System info	Owner	Other users	Credit account	Action	VNC port #
65888	3/25/2016 5:14:05 PM	3/25/2023 5:00:00 PM	cbsudc01	Linux	Supermicro SYS-6028R-TR 12 cores; 256GB RAM; 16TB HDD;	jarekp	tmc46 lcj34 zrm22 esb33 jarekpp mcr72 jav246 bm646	CBSU Collaboration Large Memory	Connect VNC top	
68317	4/27/2016 5:39:38 PM	1/1/2100 12:00:00 AM	cbsufsrv4	Linux	Dell Precision T110 4 cores; 4GB RAM; 78.1TB HDD;	ACL	bukowski qisun pas48 jarekpp mingh jcg233 esb33 dci1 sjm336 mbb262 mcr72 tmc46 kls283 arphillips nkl3 rjl278 sej65 pjb39 lcj34 chengzou amj77 zrm22 jdw297 ns868 mm2842 lfsamayo jav246 crf87 gr226 ter56 eml255 jlg374 hw449 jws429 ajs692 jmg572 tw493 ec796 tct38 yw2326 bs674 jp2476 eoren bm646 ag2484 mze3 tjr239 tf259 tml78	ACL	Connect VNC top	

Select resolution you want

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

Logging in to a Linux workstation

(GUI)



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

BIOHPC CLOUD:
: VNC CONNECTION

Your VNC session at cbsum1c2b007.biohpc.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 (**cbsum1c2b007**) and port number (**5901**) and connect!
Typically machine name and port number are used together: **cbsum1c2b007.biohpc.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://cbsum1c2b007.biohpc.cornell.edu:5801 .

Return to My Reservations.

Everything is OK! Launching VNC connection

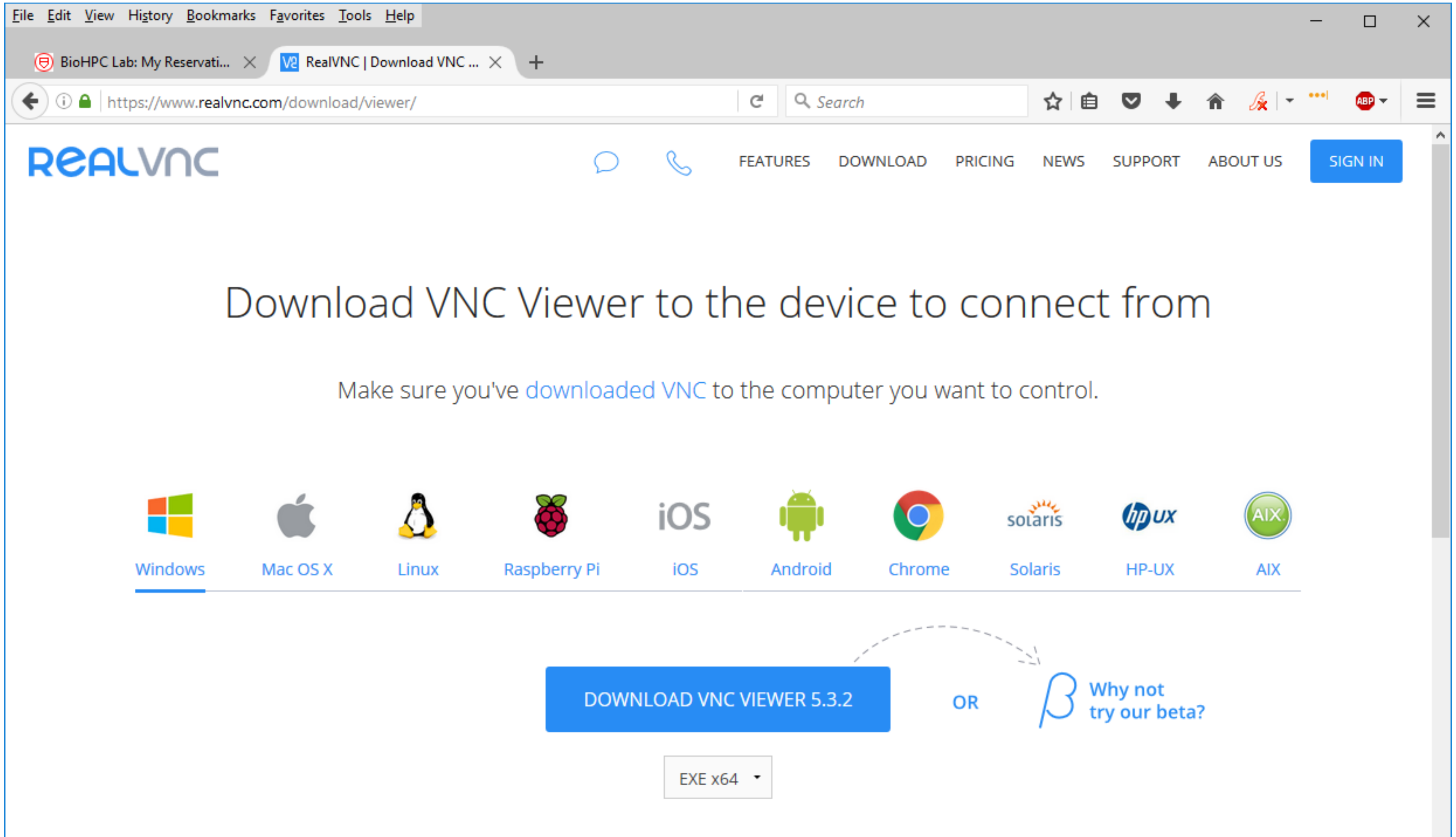
Website credentials: user: jarekp 'jp86@cornell.edu' [BioHPC Cloud]
logout

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.

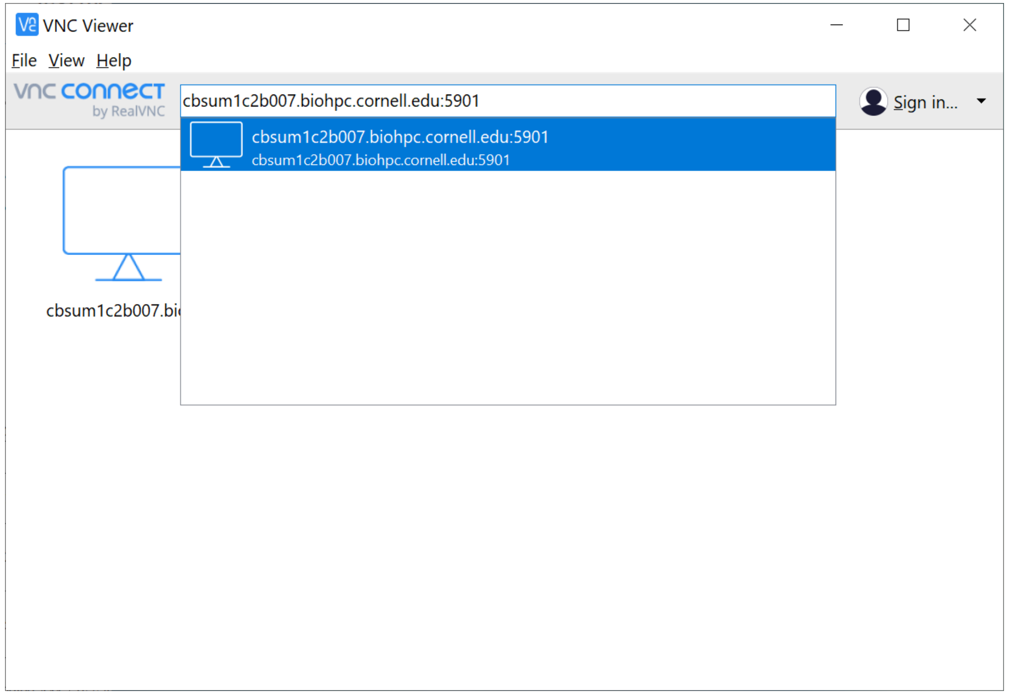


Download VNC Viewer to the device to connect from

Make sure you've downloaded VNC to the computer you want to control.

Logging in to a Linux workstation

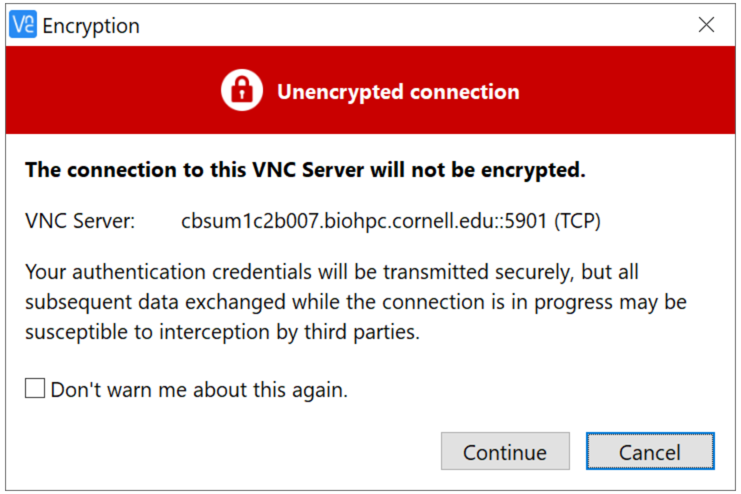
(GUI)



cbsum1c2b007.biohpc.cornell.edu:5901

cbsum1c2b007.biohpc.cornell.edu:5901
cbsum1c2b007.biohpc.cornell.edu:5901

cbsum1c2b007.bi

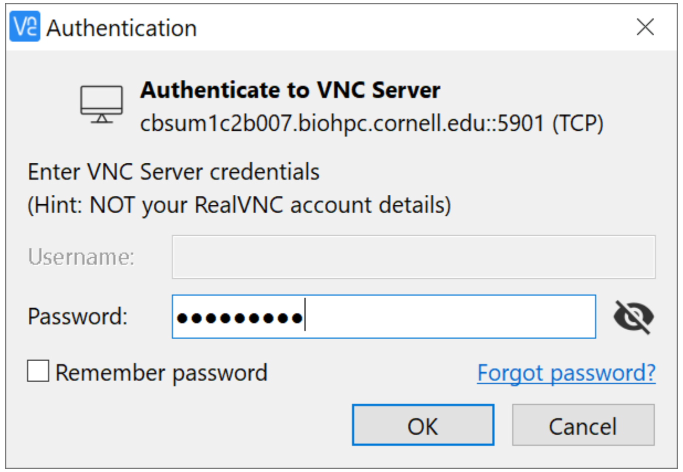


The connection to this VNC Server will not be encrypted.

VNC Server: cbsum1c2b007.biohpc.cornell.edu::5901 (TCP)

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.



Authenticate to VNC Server
cbsum1c2b007.biohpc.cornell.edu::5901 (TCP)

Enter VNC Server credentials
(Hint: NOT your RealVNC account details)

Username:

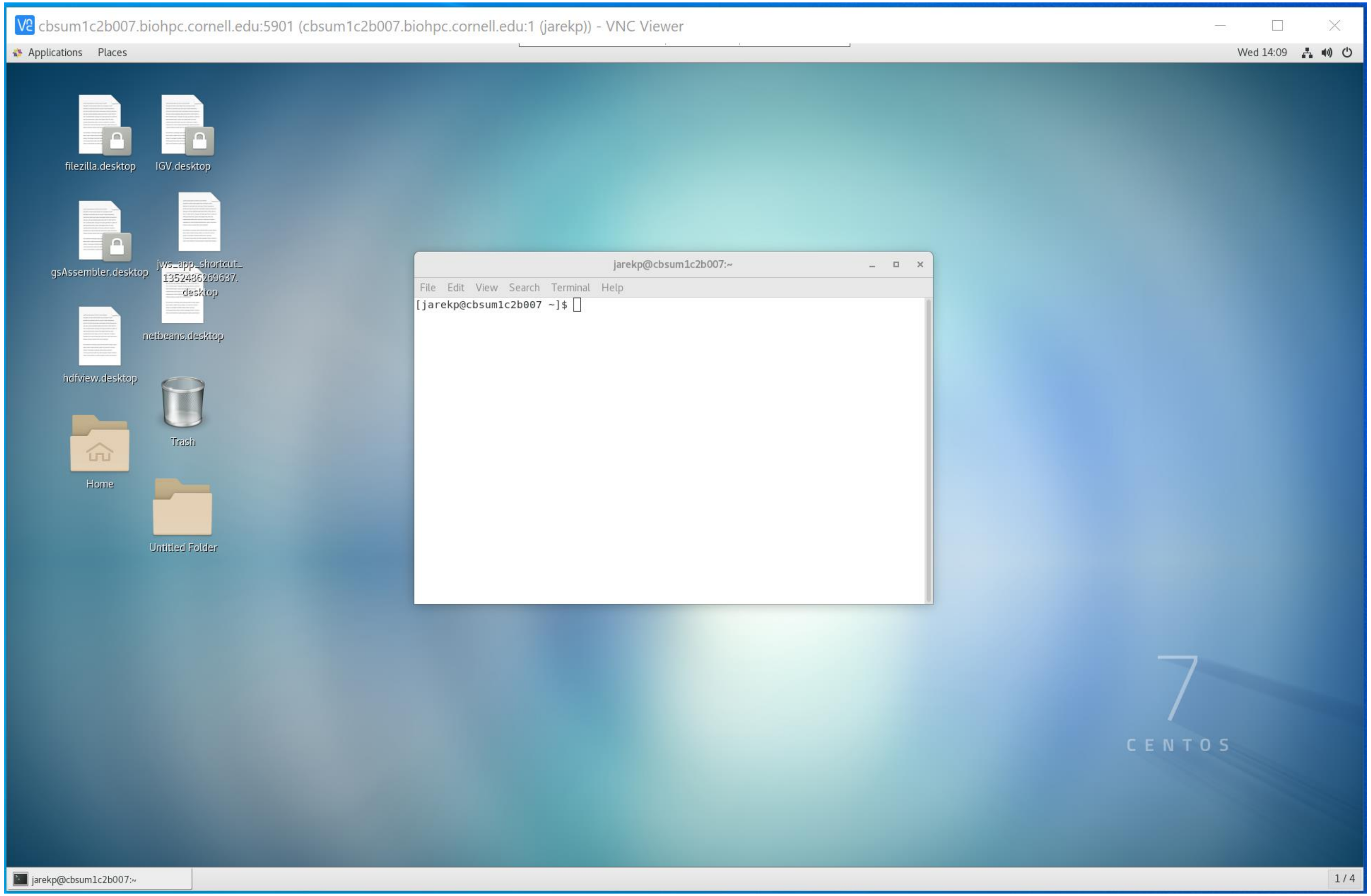
Password: •••••••••

☐ Remember password

Logging in to a Linux workstation via VNC client

(GUI)

Right-click anywhere within desktop, select **Open Terminal** …. or
…. click **Applications -> Accessories -> Terminal**



```
[jarekp@cbsum1c2b007 ~]$ 
```

Connecting to BioHPC Cloud: VNC

VNC sessions are *persistent*.

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

If you need to reset the session you need to use “Cancel VNC” link followed by “Connect VNC”.

Equivalent to Windows Remote Desktop.

Connecting with VNC form external network without VPN Mac and Linux

- Enable your VNC connection first (see slide 94)

- Open local terminal window on your Mac or Linux computer

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

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

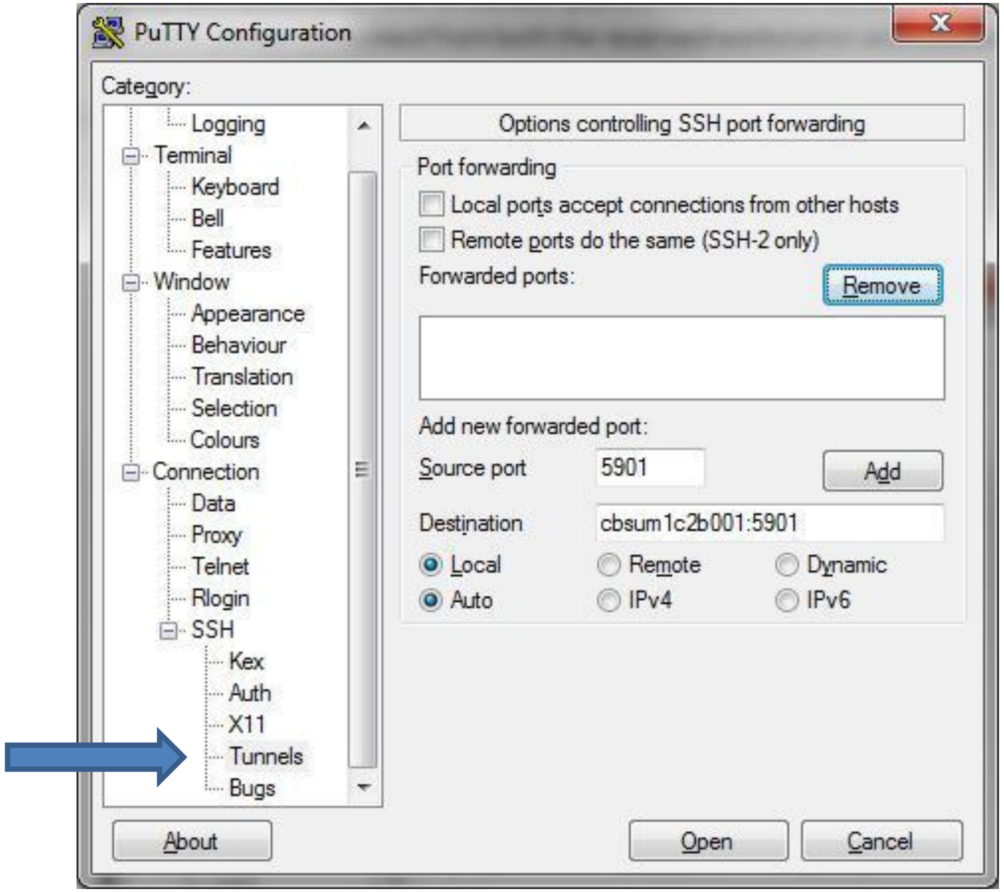
- Now you can connect to your VNC by typing localhost:5901 in your VNC Viewer software.

Connecting with VNC form external network without VPN
Windows

- Enable your VNC connection first (see slide 94). Note what is your VNC port.

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

- On the left panel scroll down to Connection -> SSH -> Tunnels



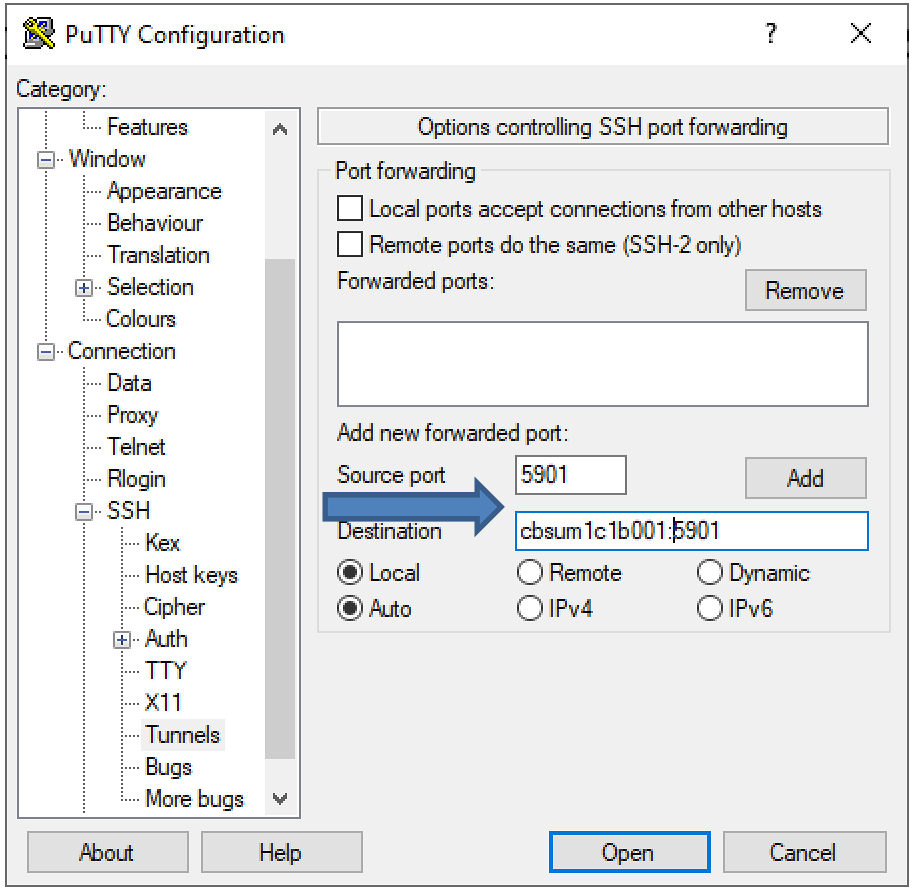
☐ Local ports accept connections from other hosts
☐ Remote ports do the same (SSH-2 only)

Source port: 5901
Destination: cbsum1c2b001:5901

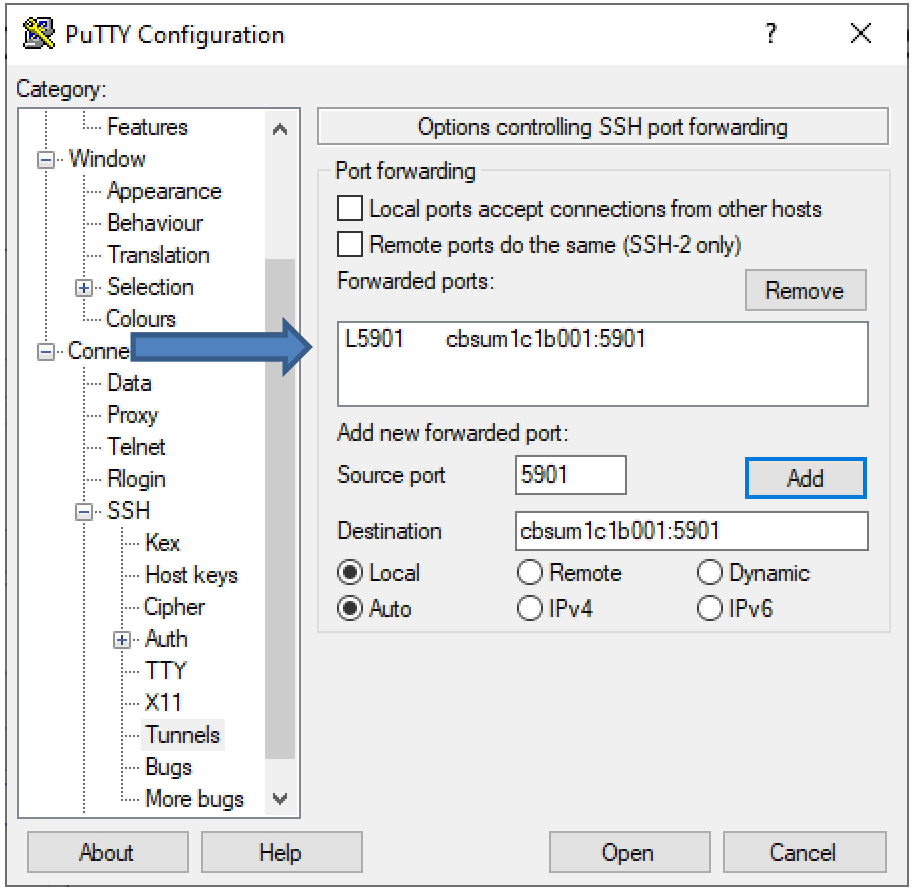
● Local ○ Remote ○ Dynamic
● Auto ○ IPv4 ○ IPv6

Connecting with VNC form external network without VPN Windows

Enable your VNC connection first (see slide 82). Note what is your VNC port. Type the port as shown below with the destinations server name and click Add. Now you can connect to your VNC by typing localhost:5901 in your VNC Viewer software.



☐ Local ports accept connections from other hosts
☐ Remote ports do the same (SSH-2 only)
Forwarded ports:
Add new forwarded port:
Source port: 5901
Destination: cbsum1c1b001:5901
● Local ○ Remote ○ Dynamic
● Auto ○ IPv4 ○ IPv6



☐ Local ports accept connections from other hosts
☐ Remote ports do the same (SSH-2 only)
Forwarded ports:
L5901 cbsum1c1b001:5901
Add new forwarded port:
Source port: 5901
Destination: cbsum1c1b001:5901
● Local ○ Remote ○ Dynamic
● Auto ○ IPv4 ○ IPv6

Connecting with VNC form external network without VPN Windows

Connect to the login node (e.g. cbsulogin.biohpc.cornell.edu). Once connected, type the command to connect to the destination server (cbsum1c2b001 in this example)

```
ssh cbsum1c2b001
```

which will connect you to your reserved workstation (you will be asked password again). Remember to substitute example workstation name with your reserved workstation name! Now you will be able to work on your server with command line software and X-Windows software. Once you are done, please disconnect from both the reserved workstation and cbsulogin (press “Ctrl-D” twice).

Exercise: connect to your assigned workstation using VNC

- Go to “My Reservations” page
  http://biohpc.cornell.edu/lab/lab.aspx , log in, click on “My Reservations” menu link
- Choose resolution (depends on your monitor)
- Click on “Connect VNC”
- Follow prompts
- Open terminal window in the VNC desktop by right-click on the desktop background and choosing “Open Terminal”.
- Disconnect (close browser windows) and then reconnect. Is the session still alive?

Connecting to BioHPC Cloud

GUI (graphical) connection: X-Windows

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

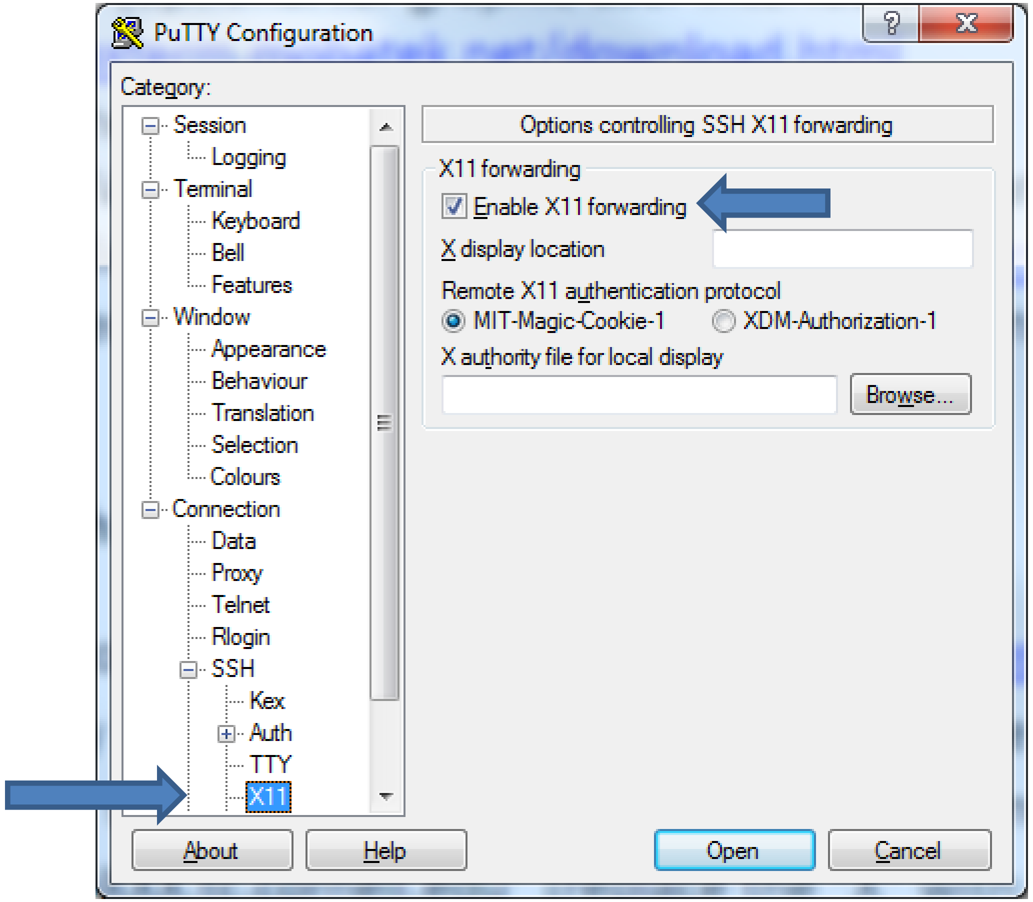
Often faster than VNC

Connecting to BioHPC Cloud: X-Windows 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



☑ Enable X11 forwarding

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

Connecting to BioHPC Cloud: X-Windows Windows

- Start your GUI (graphical) software in terminal window. For testing you may try eog (Linux image viewer) or firefox.
- You may need to allow MobaXterm connections through Windows firewall.
- New window with your GUI program will appear. The program will physically run on the BioHPC Cloud machine, but it will display graphics on your local computer.

Connecting to BioHPC Cloud: X-Windows
Linux or Mac

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

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

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

- NOTE: Mac machines no longer have X11 as a part of their system, it must be installed by user, XQuartz is what we know works

Connect to BioHPC Cloud: X-Windows from external network no VPN Windows

- Make sure X11 forwarding is enabled (see slide 105).
- Connect to cbsulogin.biohpc.cornell.edu (or cbsulogin2 or cbsulogin3)
- From cbsulogin connect to your target server (cbsuxxx) with the following command

  ```
  ssh -Y cbsuxxx
  ```

  (replace cbsuxxx with you target server name)
- Make sure MobaXTerm is running and start your X11 application

Connect to BioHPC Cloud: X-Windows from external network no VPN Linux or Mac

- Open local terminal window on your Mac or Linux computer

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

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

- NOTE: Mac machines no longer have X11 as a part of their system, it must be installed by user, XQuatrz is what we know works

Exercise: connect to your assigned workstation with X-Windows

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

```
eog /home/jarekp/picture1.jpg
```

Getting started with a new account

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

Transferring data

- sftp or scp (secure file transfer protocol)

  Transfer can be done to and from BioHPC machines.

- Globus

  Transfer can be done to and from BioHPC machines.

- wget

  Transfer can be done to BioHPC machines only.

File transfer: wget

from web- and ftp sites to lab workstations

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

- **Examples:**

```
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.biohpc.cornell.edu/Sequencing/showseqfile.aspx?cntrl=646698859&laneid=487&mode=http&file=e_coli_1000_1.fq"
```

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

Exercise: download BLOSUM100 with wget

- Connect to your workstation using ssh
- Type wget command

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

File transfer: sftp

between PC or Mac and a BioHPC workstation

On Windows PC: install and use your favorite **sftp client** program, such as

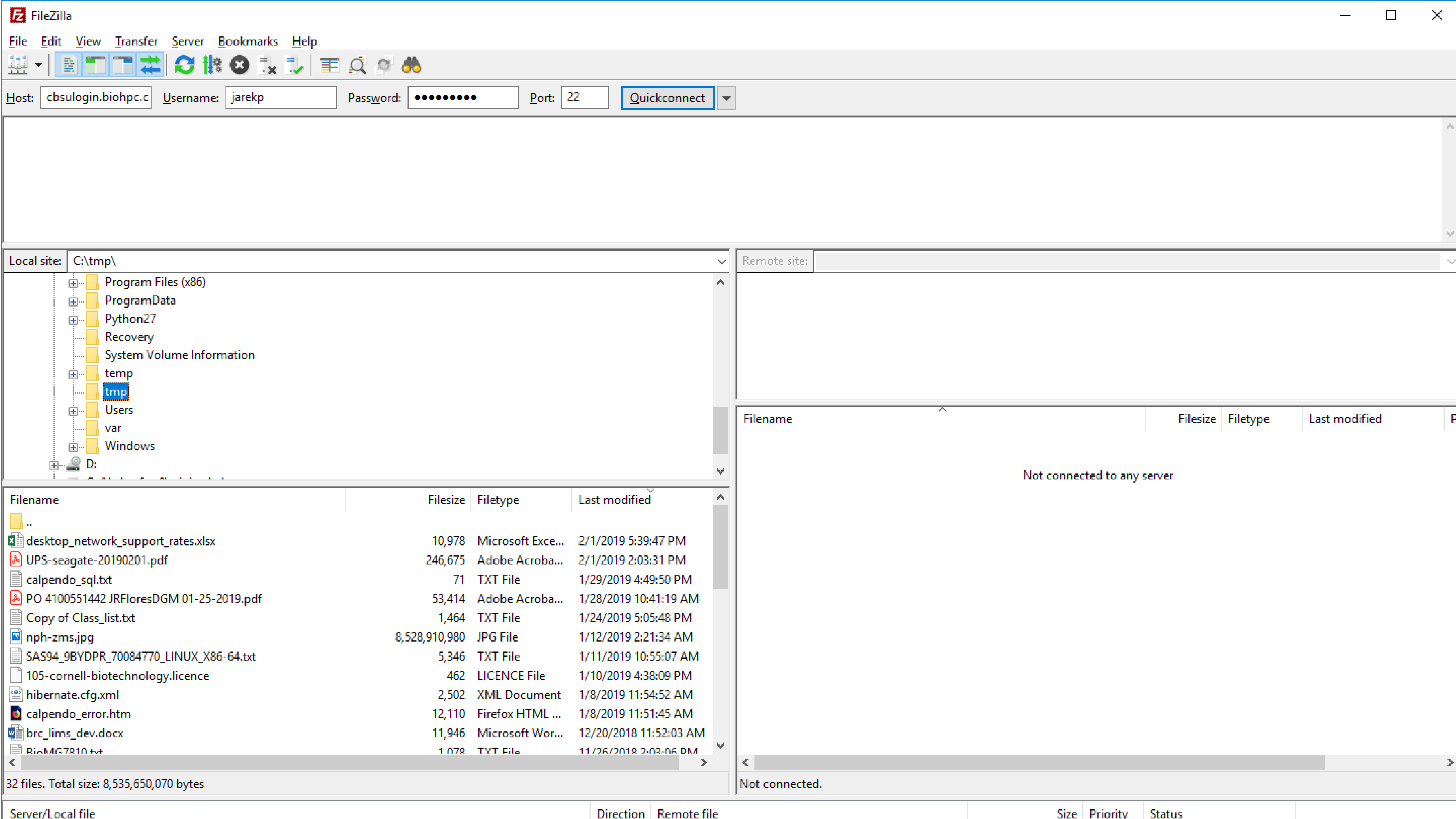
- **winscp**: http://winscp.net/eng/index.php
- **CoreFTP LE**: http://www.coreftp.com/
- **FileZilla** (client): http://filezilla-project.org/
- … others…
- When connecting to BioHPC servers from a client, use the **sftp** protocol (port 22). You will be asked for your user name and password (the same you use to log in to the BioHPC servers).
- Transfer text file in text mode, binary files in binary mode (the “default” is 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

Recommended!

On a Mac: file transfer program is **fetch** (recommended by Cornell CIT)

- https://it.cornell.edu/managed-servers/transfer-files-using-fetch
- graphical user interface
- Drag-and-drop functionality

But **FileZilla** has now also Mac version which I like best …



Host: cbsulogin.biohpc.c
Username: jarekp
Password: ●●●●●●●●●
Port: 22

Local site: C:\tmp\

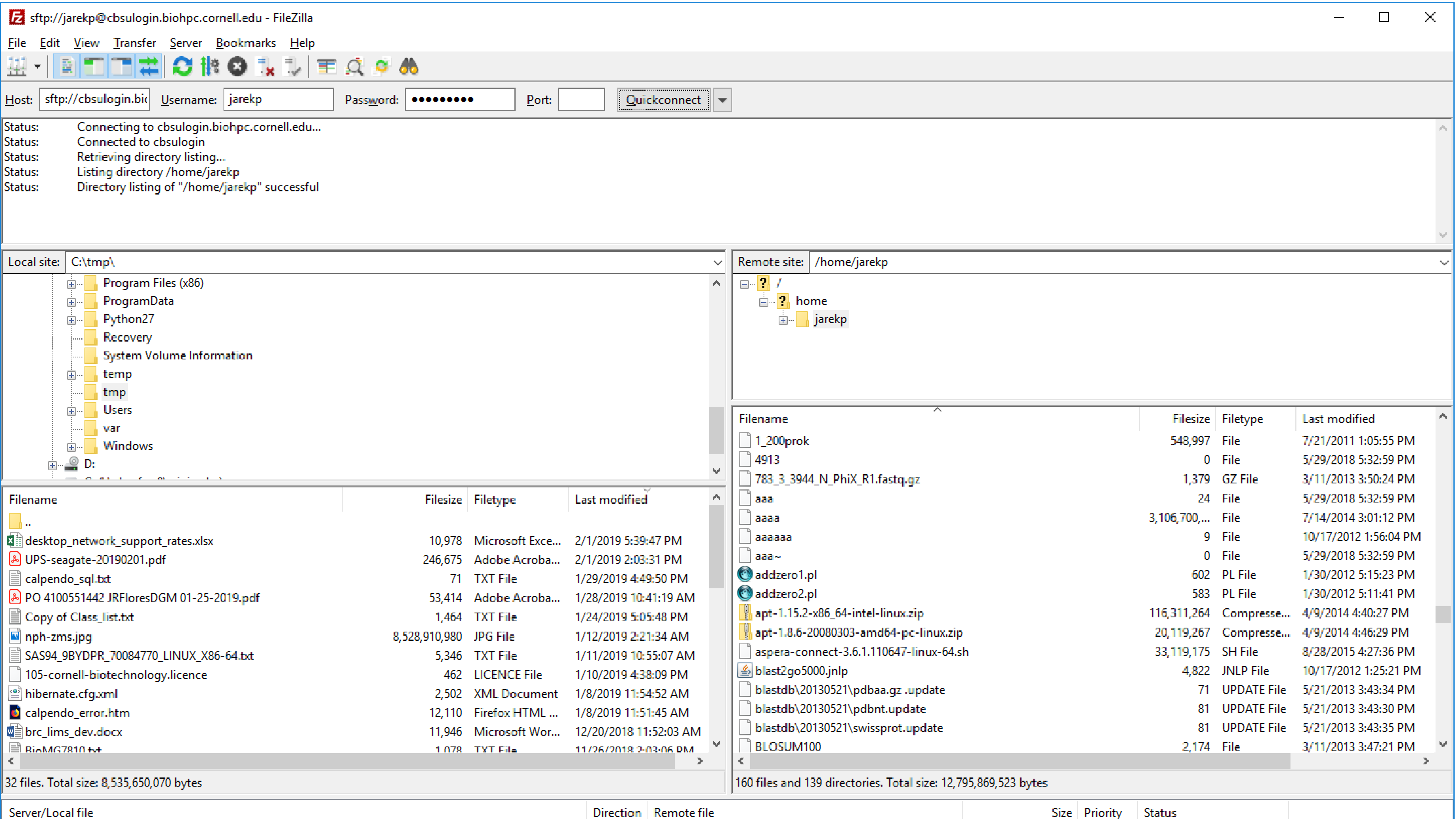
- Program Files (x86)
- ProgramData
- Python27
- Recovery
- System Volume Information
- temp
- tmp
- Users
- var
- Windows
- D:

Filename	Filesize	Filetype	Last modified
..			
desktop_network_support_rates.xlsx	10,978	Microsoft Exce...	2/1/2019 5:39:47 PM
UPS-seagate-20190201.pdf	246,675	Adobe Acroba...	2/1/2019 2:03:31 PM
calpendo_sql.txt	71	TXT File	1/29/2019 4:49:50 PM
PO 4100551442 JRFloresDGM 01-25-2019.pdf	53,414	Adobe Acroba...	1/28/2019 10:41:19 AM
Copy of Class_list.txt	1,464	TXT File	1/24/2019 5:05:48 PM
nph-zms.jpg	8,528,910,980	JPG File	1/12/2019 2:21:34 AM
SAS94_9BYDPR_70084770_LINUX_X86-64.txt	5,346	TXT File	1/11/2019 10:55:07 AM
105-cornell-biotechnology.licence	462	LICENCE File	1/10/2019 4:38:09 PM
hibernate.cfg.xml	2,502	XML Document	1/8/2019 11:54:52 AM
calpendo_error.htm	12,110	Firefox HTML ...	1/8/2019 11:51:45 AM
brc_lims_dev.docx	11,946	Microsoft Wor...	12/20/2018 11:52:03 AM
BioMG7810.txt	1,078	TXT File	11/26/2018 2:03:06 PM

32 files. Total size: 8,535,650,070 bytes

Not connected to any server

Not connected.



Host: sftp://cbsulogin.bi Username: jarekp Password: ●●●●●●●●● Port:

```
Status:     Connecting to cbsulogin.biohpc.cornell.edu...
Status:     Connected to cbsulogin
Status:     Retrieving directory listing...
Status:     Listing directory /home/jarekp
Status:     Directory listing of "/home/jarekp" successful
```

Local site: C:\tmp\

Filename	Filesize	Filetype	Last modified
..			
desktop_network_support_rates.xlsx	10,978	Microsoft Exce...	2/1/2019 5:39:47 PM
UPS-seagate-20190201.pdf	246,675	Adobe Acroba...	2/1/2019 2:03:31 PM
calpendo_sql.txt	71	TXT File	1/29/2019 4:49:50 PM
PO 4100551442 JRFloresDGM 01-25-2019.pdf	53,414	Adobe Acroba...	1/28/2019 10:41:19 AM
Copy of Class_list.txt	1,464	TXT File	1/24/2019 5:05:48 PM
nph-zms.jpg	8,528,910,980	JPG File	1/12/2019 2:21:34 AM
SAS94_9BYDPR_70084770_LINUX_X86-64.txt	5,346	TXT File	1/11/2019 10:55:07 AM
105-cornell-biotechnology.licence	462	LICENCE File	1/10/2019 4:38:09 PM
hibernate.cfg.xml	2,502	XML Document	1/8/2019 11:54:52 AM
calpendo_error.htm	12,110	Firefox HTML ...	1/8/2019 11:51:45 AM
brc_lims_dev.docx	11,946	Microsoft Wor...	12/20/2018 11:52:03 AM
BioMG7810.txt	1,078	TXT File	11/26/2018 2:03:06 PM

32 files. Total size: 8,535,650,070 bytes

Remote site: /home/jarekp

Filename	Filesize	Filetype	Last modified
1_200prok	548,997	File	7/21/2011 1:05:55 PM
4913	0	File	5/29/2018 5:32:59 PM
783_3_3944_N_PhiX_R1.fastq.gz	1,379	GZ File	3/11/2013 3:50:24 PM
aaa	24	File	5/29/2018 5:32:59 PM
aaaa	3,106,700,...	File	7/14/2014 3:01:12 PM
aaaaaa	9	File	10/17/2012 1:56:04 PM
aaa~	0	File	5/29/2018 5:32:59 PM
addzero1.pl	602	PL File	1/30/2012 5:15:23 PM
addzero2.pl	583	PL File	1/30/2012 5:11:41 PM
apt-1.15.2-x86_64-intel-linux.zip	116,311,264	Compresse...	4/9/2014 4:40:27 PM
apt-1.8.6-20080303-amd64-pc-linux.zip	20,119,267	Compresse...	4/9/2014 4:46:29 PM
aspera-connect-3.6.1.110647-linux-64.sh	33,119,175	SH File	8/28/2015 4:27:36 PM
blast2go5000.jnlp	4,822	JNLP File	10/17/2012 1:25:21 PM
blastdb\20130521\pdbaa.gz .update	71	UPDATE File	5/21/2013 3:43:34 PM
blastdb\20130521\pdbnt.update	81	UPDATE File	5/21/2013 3:43:30 PM
blastdb\20130521\swissprot.update	81	UPDATE File	5/21/2013 3:43:35 PM
BLOSUM100	2,174	File	3/11/2013 3:47:21 PM

160 files and 139 directories. Total size: 12,795,869,523 bytes

Exercise: download BLOSUM100 file from Lab machine to your local computer

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

Transferring data: Globus

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

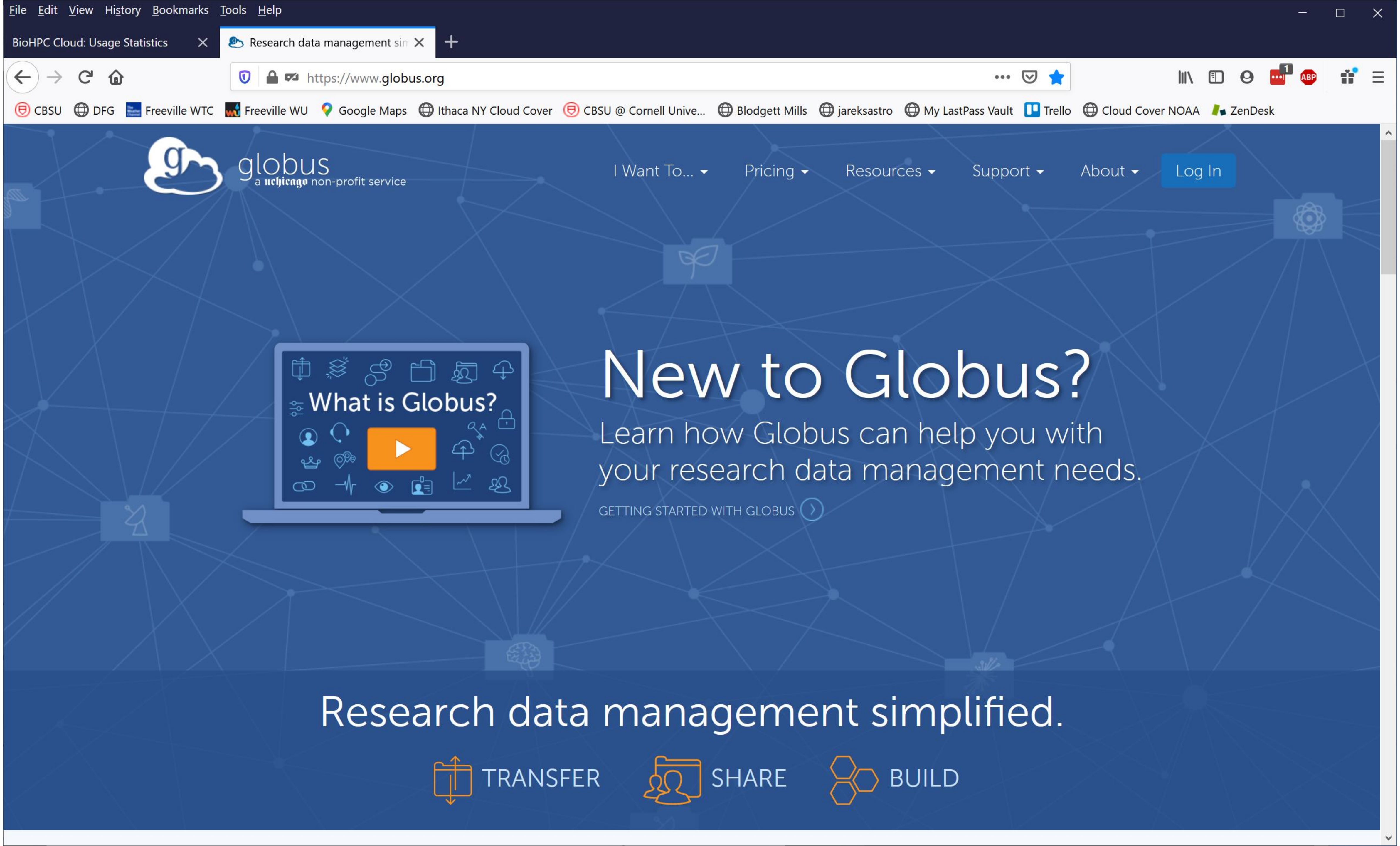
Data can be transferred between *endpoints*, which can be data servers or Globus Connect clients.

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

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

http://biohpc.cornell.edu/lab/doc/Globus_at_BioHPC_Lab.pdf

https://www.globusonline.org/quickstart/

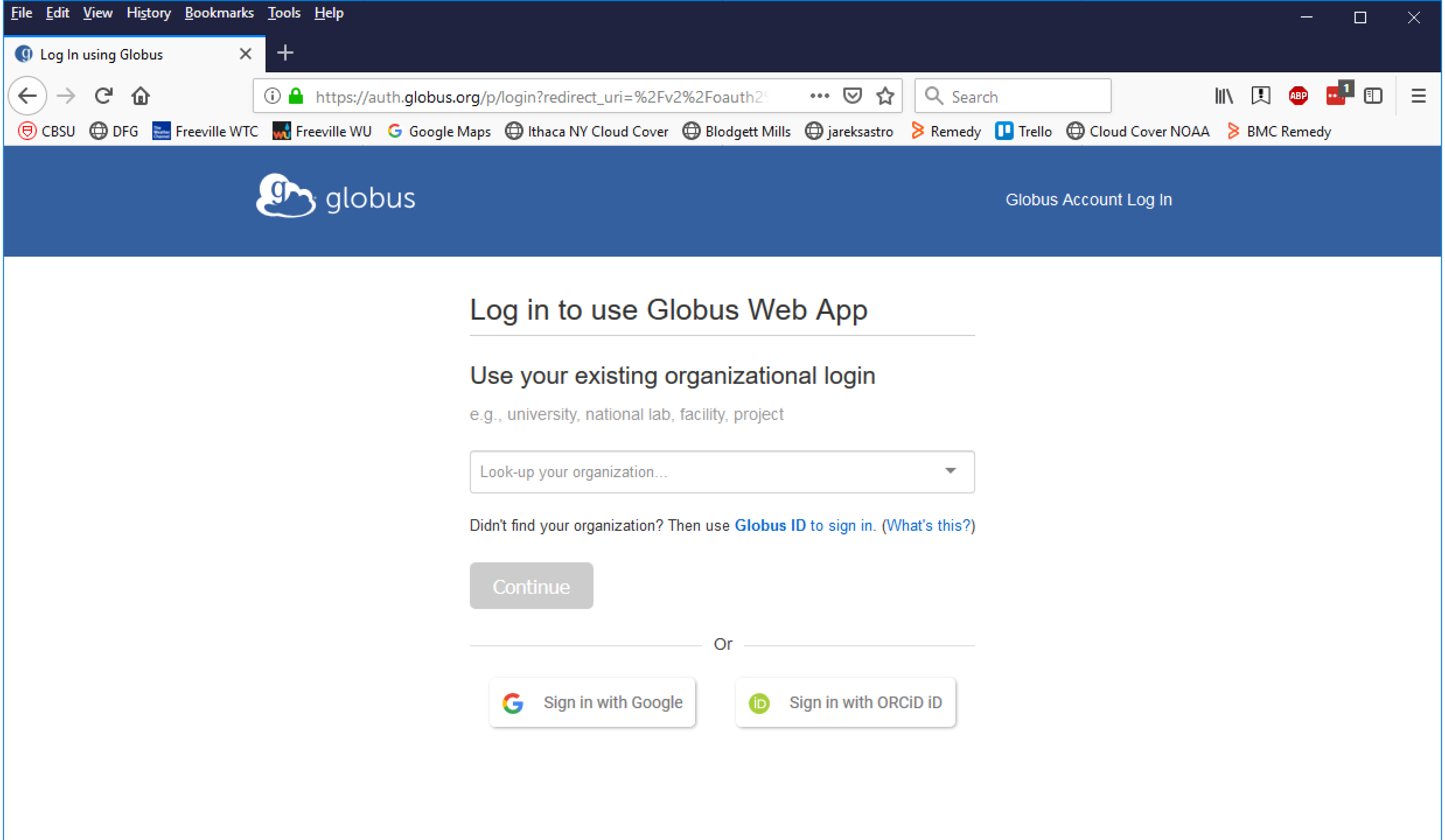


New to Globus?

Learn how Globus can help you with your research data management needs.

GETTING STARTED WITH GLOBUS

Research data management simplified.



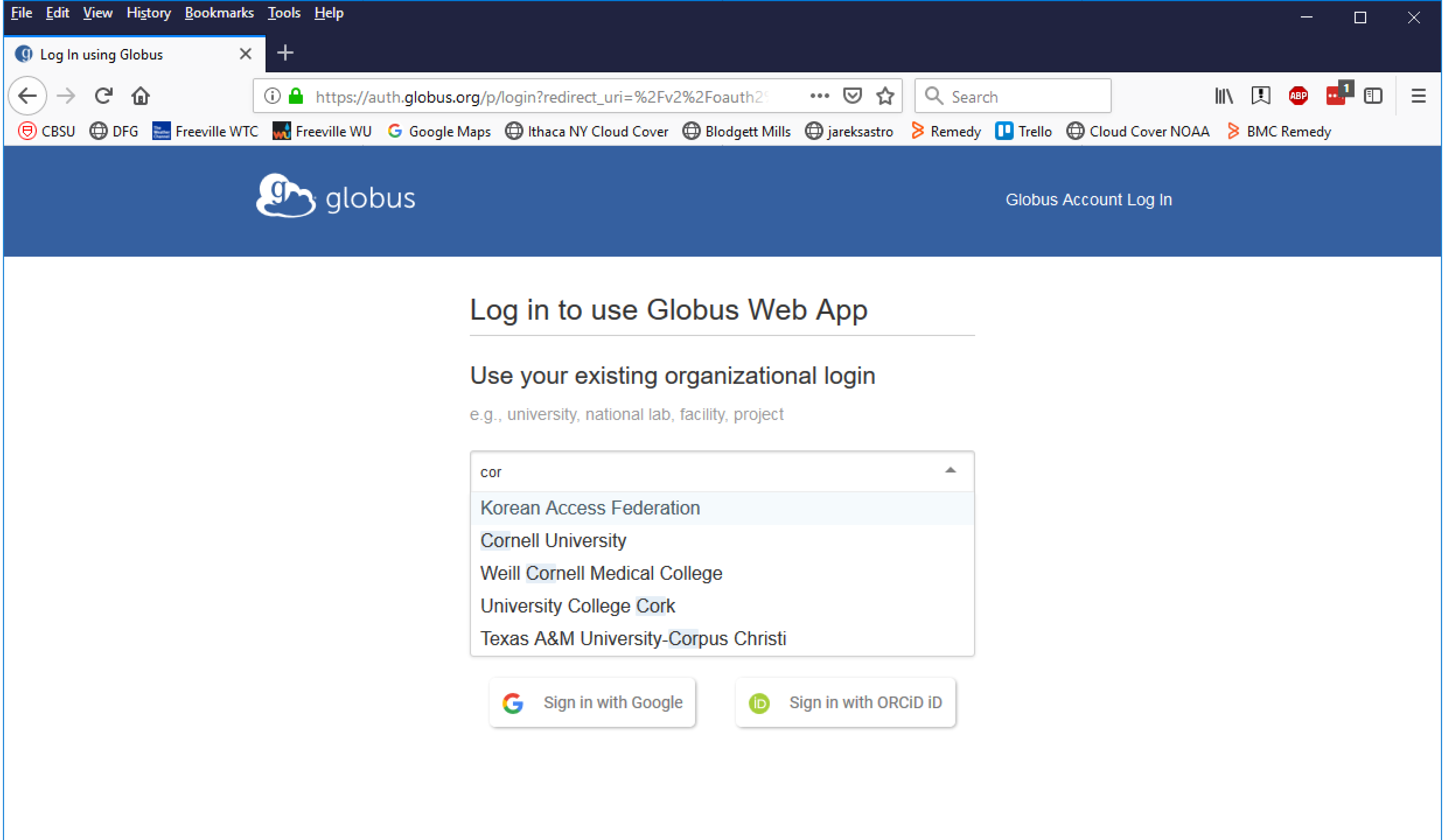
Log in to use Globus Web App

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?)



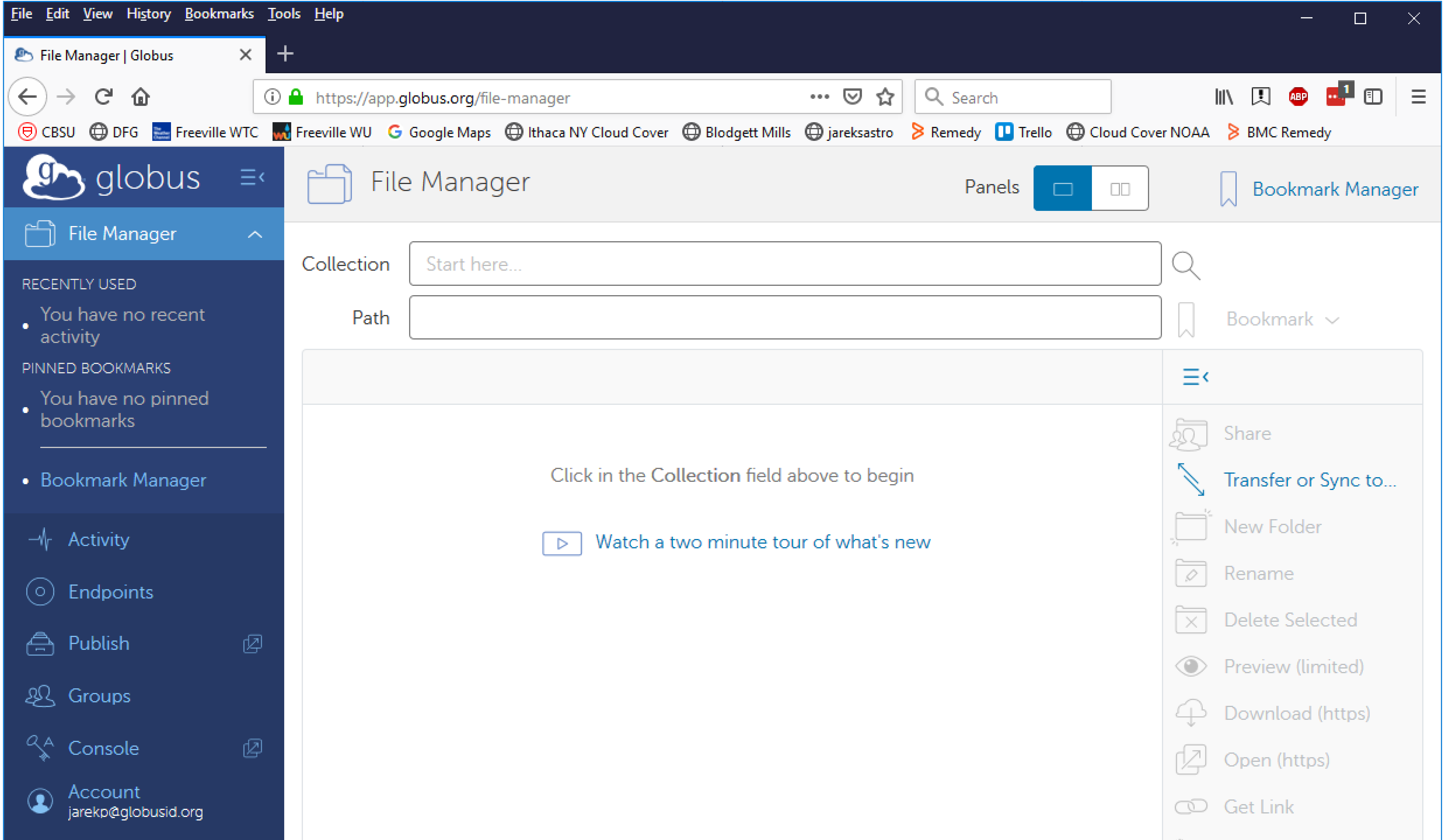
Log in to use Globus Web App

Use your existing organizational login

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

cor

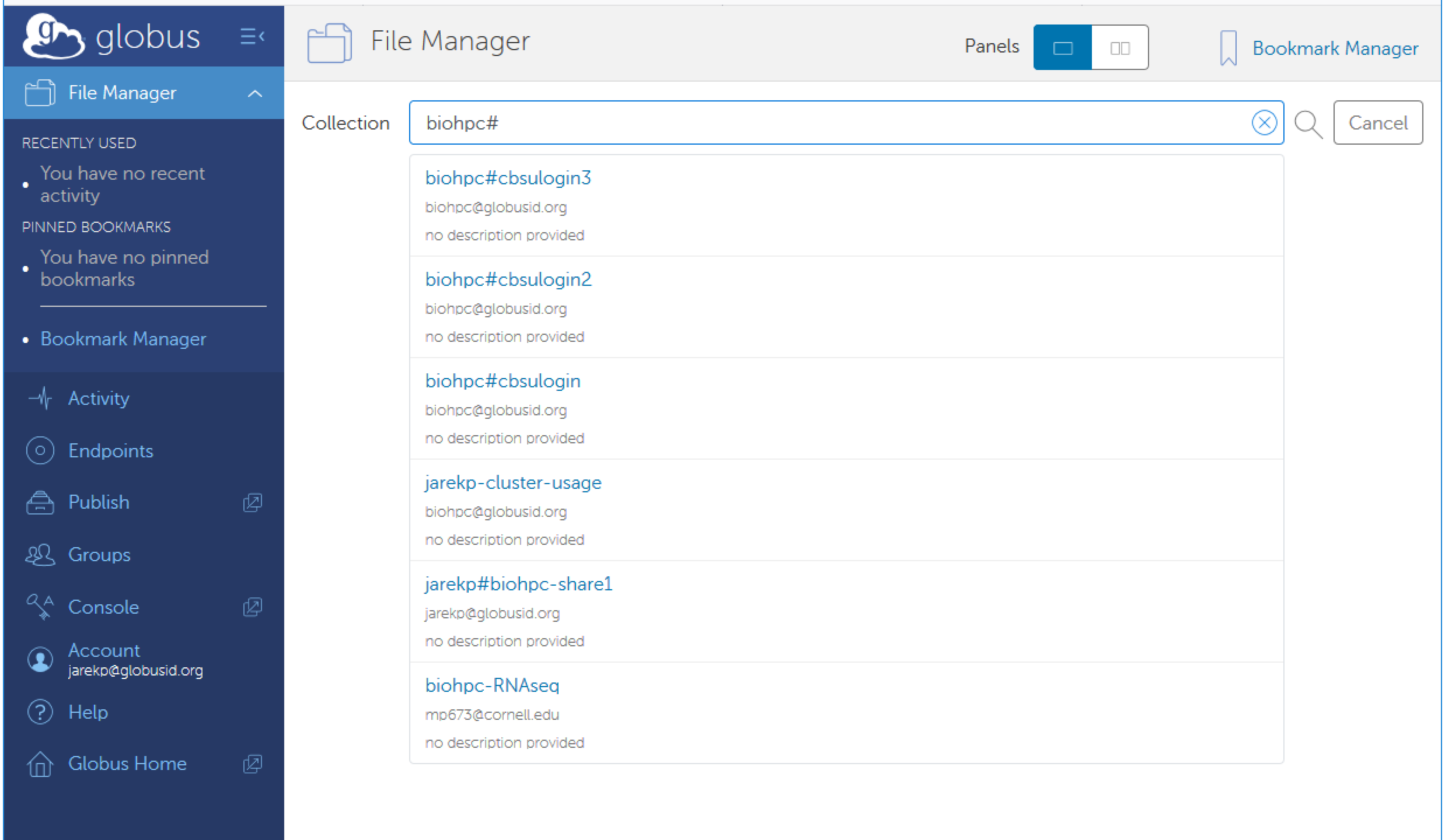
- Korean Access Federation
- Cornell University
- Weill Cornell Medical College
- University College Cork
- Texas A&M University-Corpus Christi



- You have no recent activity
- You have no pinned bookmarks

Click in the Collection field above to begin

Watch a two minute tour of what's new



Collection: biohpc#

biohpc#cbsulogin3
biohpc@globusid.org
no description provided

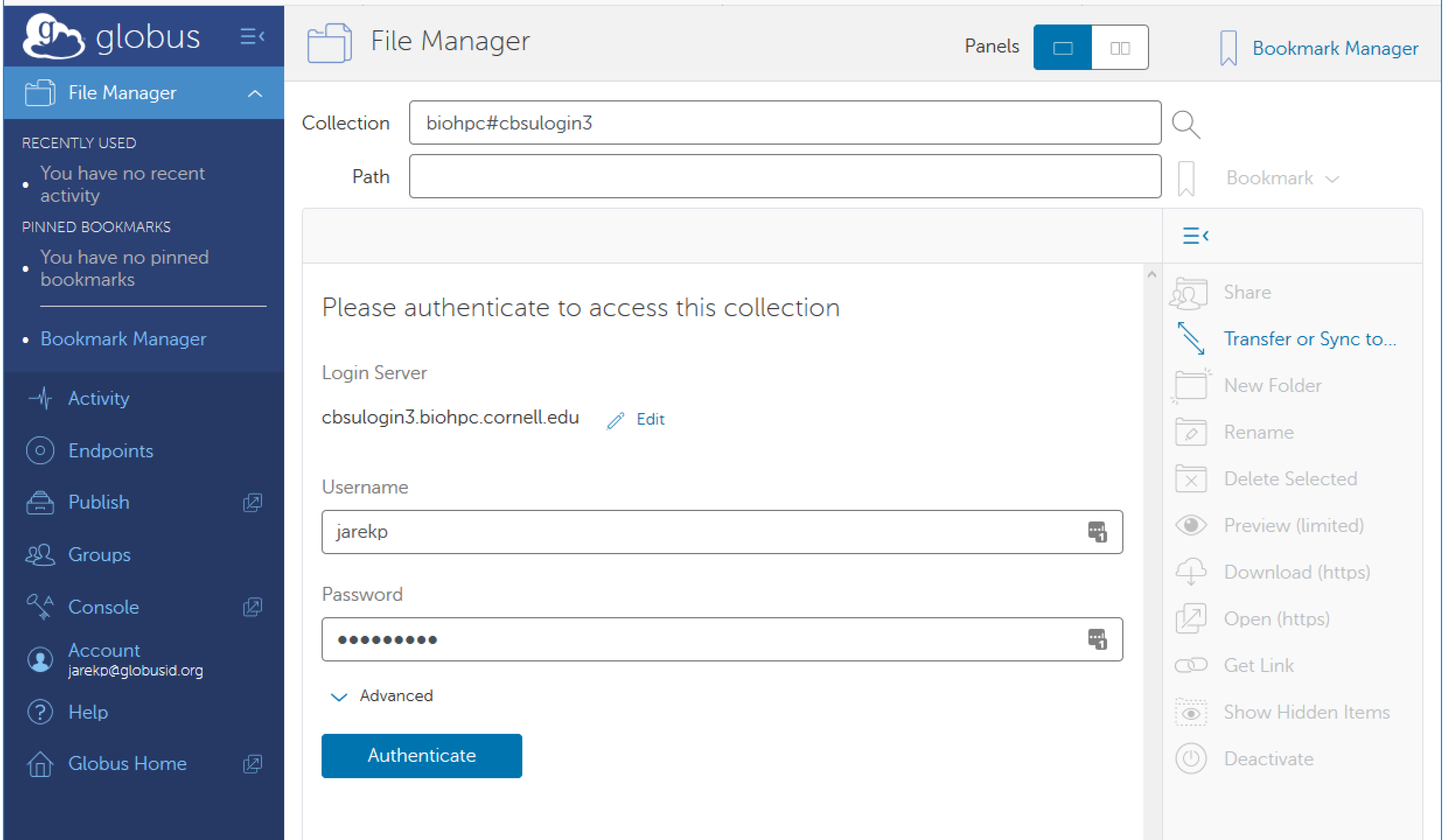
biohpc#cbsulogin2
biohpc@globusid.org
no description provided

biohpc#cbsulogin
biohpc@globusid.org
no description provided

jarekp-cluster-usage
biohpc@globusid.org
no description provided

jarekp#biohpc-share1
jarekp@globusid.org
no description provided

biohpc-RNAseq
mp673@cornell.edu
no description provided



Collection: biohpc#cbsulogin3

Path:

Please authenticate to access this collection

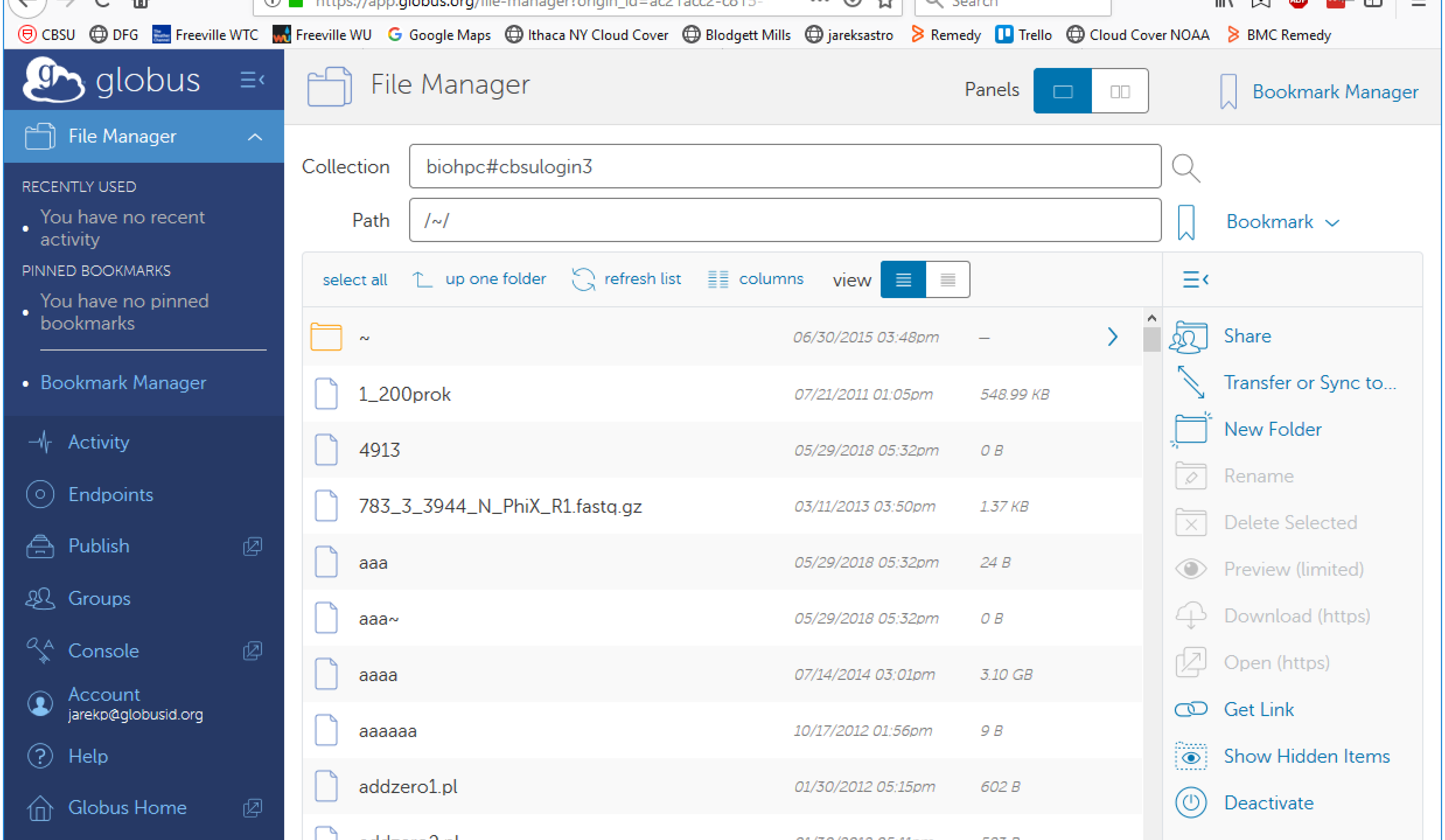
Login Server: cbsulogin3.biohpc.cornell.edu

Username: jarekp

Password: •••••••••

RECENTLY USED
- You have no recent activity

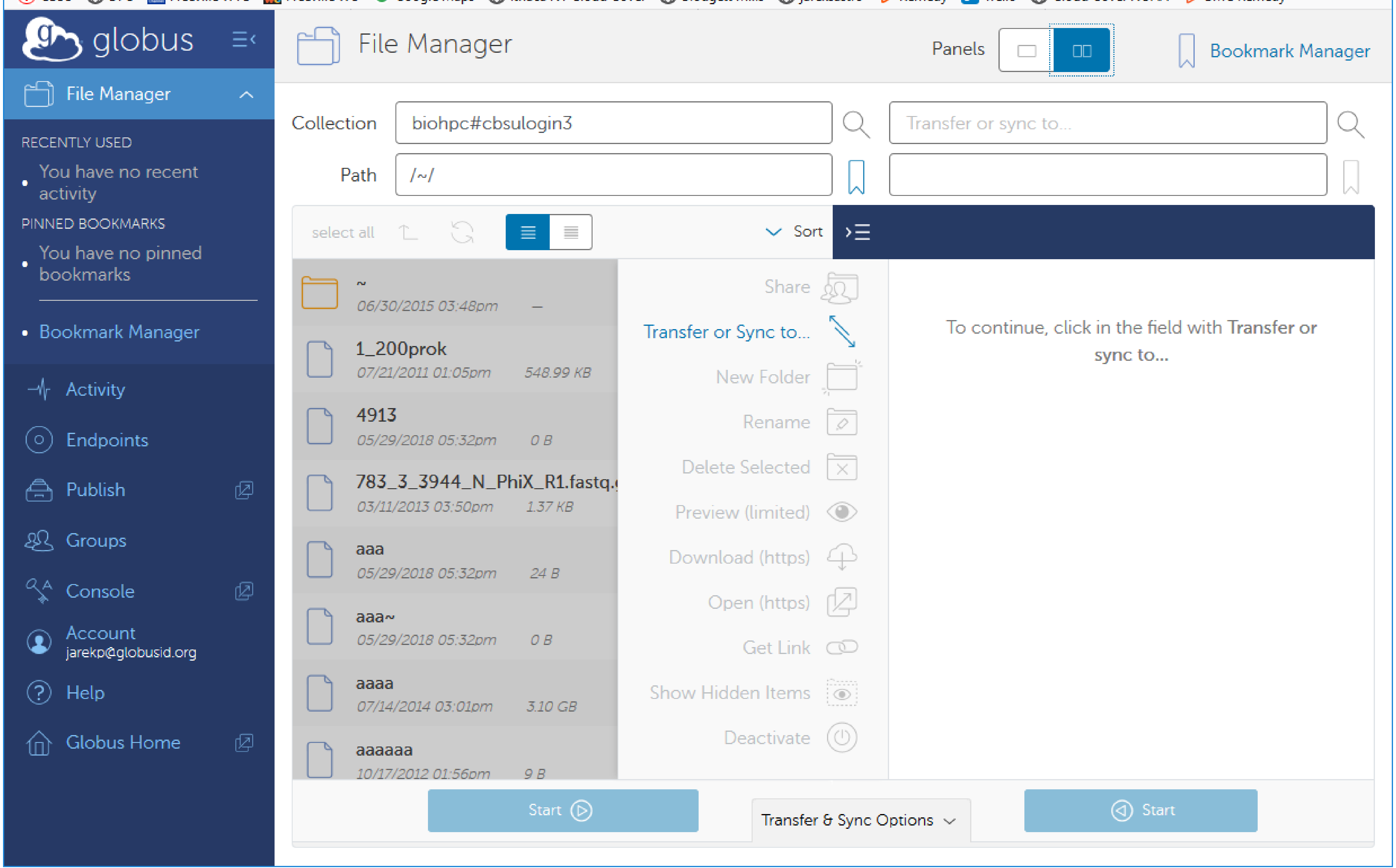
PINNED BOOKMARKS
- You have no pinned bookmarks



Collection: biohpc#cbsulogin3

Path: /~/

~	06/30/2015 03:48pm	–
1_200prok	07/21/2011 01:05pm	548.99 KB
4913	05/29/2018 05:32pm	0 B
783_3_3944_N_PhiX_R1.fastq.gz	03/11/2013 03:50pm	1.37 KB
aaa	05/29/2018 05:32pm	24 B
aaa~	05/29/2018 05:32pm	0 B
aaaa	07/14/2014 03:01pm	3.10 GB
aaaaaa	10/17/2012 01:56pm	9 B
addzero1.pl	01/30/2012 05:15pm	602 B



jarekp@globusid.org

Collection: biohpc#cbsulogin3

Path: /~/

Transfer or sync to...

~	06/30/2015 03:48pm	—
1_200prok	07/21/2011 01:05pm	548.99 KB
4913	05/29/2018 05:32pm	0 B
783_3_3944_N_PhiX_R1.fastq.g	03/11/2013 03:50pm	1.37 KB
aaa	05/29/2018 05:32pm	24 B
aaa~	05/29/2018 05:32pm	0 B
aaaa	07/14/2014 03:01pm	3.10 GB
aaaaaa	10/17/2012 01:56pm	9 B

To continue, click in the field with Transfer or sync to...

Globus data sharing

In addition to transferring data Globus can be used to share data.

For this you will need to create your own private shared endpoint (e.g. jarekp#biohpc-share1) residing on one of the existing endpoints (e.g. biohpc#cbsulogin).

You can share data this way with anyone on the Globus network.

For more details see

https://biohpc.cornell.edu/lab/doc/Using%20Globus%20to%20Share%20Data.pdf
https://docs.globus.org/how-to/share-files/

Exercise: Login to Globus using your BioHPC account

Getting started with a new account

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

Exercise 1. Connecting with ssh

Exercise 2. Connecting with VNC

Exercise 3. Connecting with X11 (X-Windows)

Exercise 4. Download data with wget

Exercise 5. Download data with FileZilla (sftp)

Exercise 6. Connecting with Globus