

# Accessing data on BioHPC

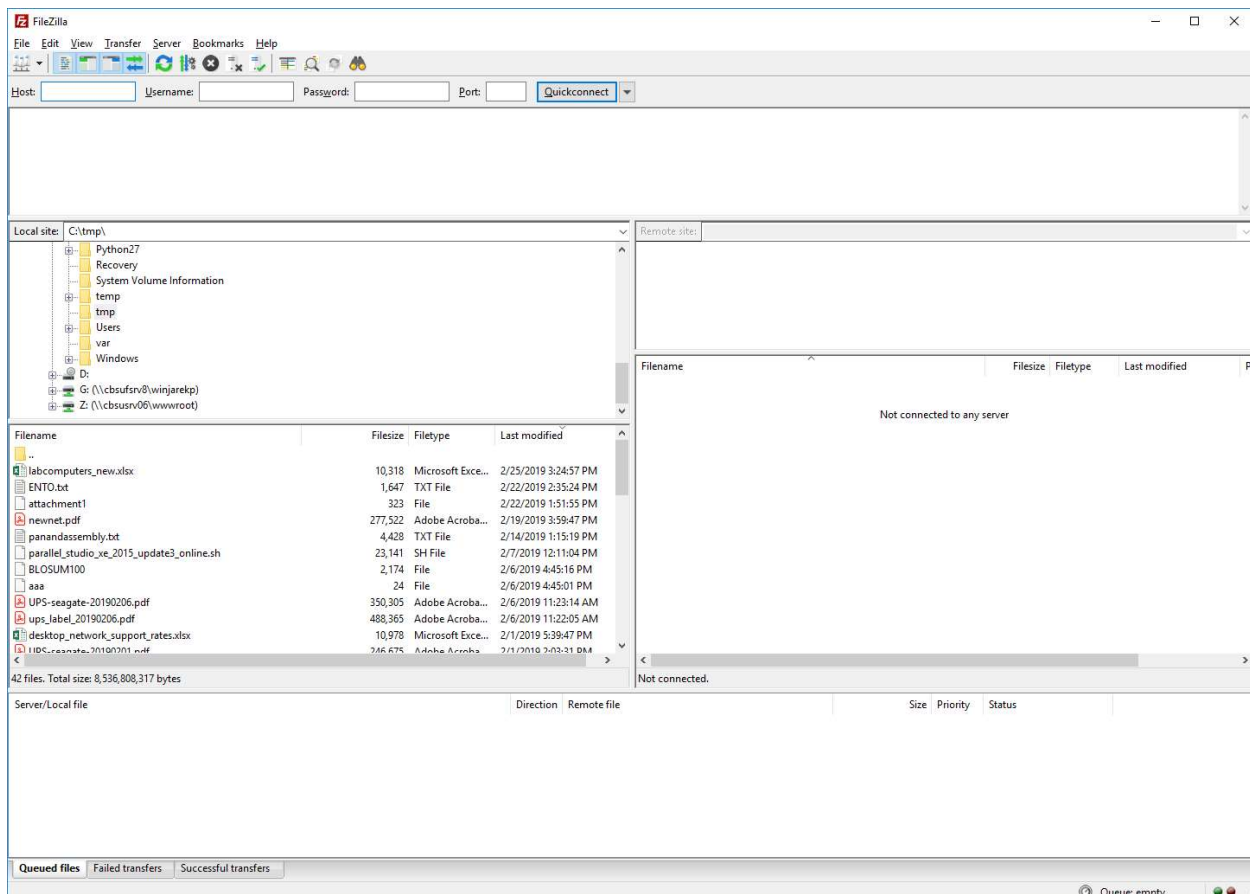
Data in BioHPC can be stored on home directories (path `/home/userid`, where `userid` is your user id), storage groups (`/home/groupname`, where `groupname` is storage group name) or on local storage attached to hosted servers. Home directories and storage groups are available on ALL BioHPC servers, local storage can be accessed on the hosted server or it can be cross-mounted on login servers or other servers (see chapter 4).

Below we discuss step-by-step how to access storage using login servers, if the discussed method applies to other servers it is noted in the explanations. There are three login servers available:

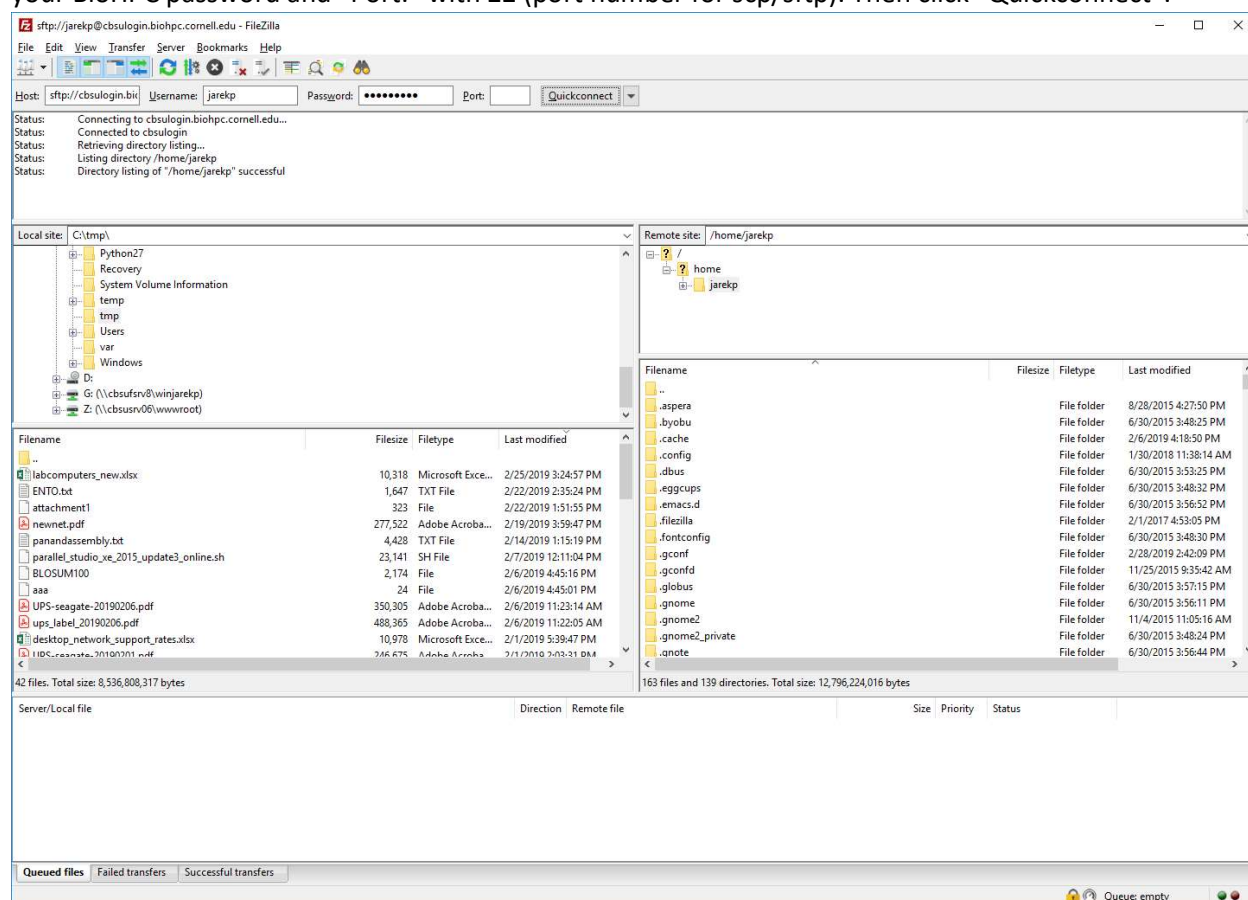
`cbsulogin.biohpc.cornell.edu`, `cbsulogin2.biohpc.cornell.edu` and `cbsulogin3.biohpc.cornell.edu`.

## 1. FileZilla

FileZilla uses common `scp` and `sftp` protocols supported by all Linux servers. It is a free, open source program available for download at <https://filezilla-project.org/>, it supports Windows, Max OS and Linux. FileZilla can be used to access data on login servers from any network (outside or inside campus), or to access data on any BioHPC server from inside Cornell campus network.



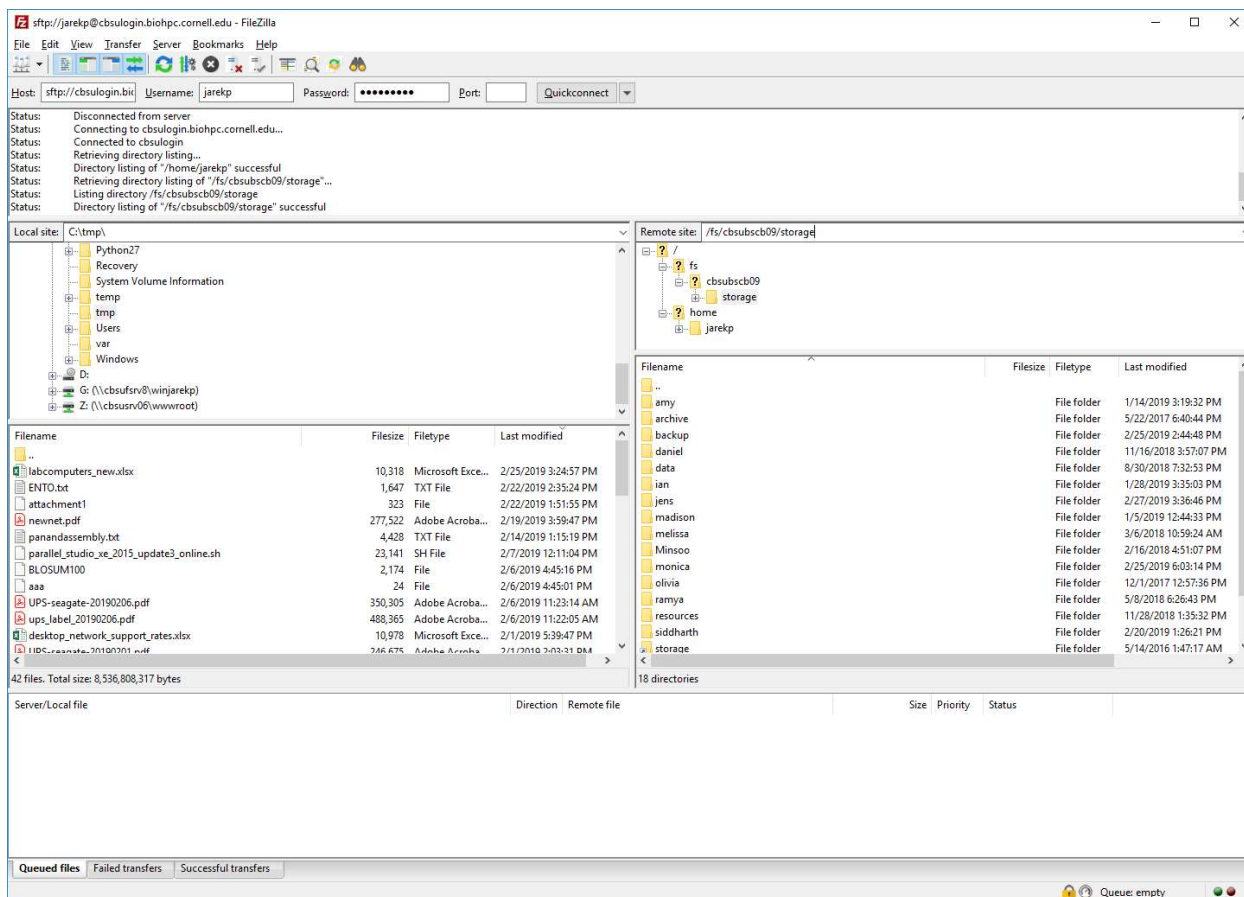
FileZilla window is divided into two panels: local (left) and remote (right). First you can browse to your local directory with files to transfer from or to. Then fill the “Host:” with your login server name (e.g. cbsulogin.biohpc.cornell.edu), “Username:” with your BioHPC username (e.g. jarekp), “Password:” with your BioHPC password and “Port:” with 22 (port number for scp/sftp). Then click “Quickconnect”.



FileZilla will connect to cbsulogin and list files/directories from your home directory in the right panel. You can now scroll, select and drag files between panels, dragging from right to left will download files to your local computer, from left to right will upload files to the login server. You can download or upload entire directories with all its content by dragging the directory itself.

If you want to access a directory that is not inside your home directory just type the path in the “Remote site:” box and press enter. My example below is /fs/cbsubscb09/storage, you need to have permission to access the directory, if not FileZilla will display red message “Permission denied” in the command status panel (top).

FileZilla will work with login servers regardless of the location of the client computer, no VPN is needed to connect to them from outside of campus. FileZilla will also work with all BioHPC servers when used from clients inside Cornell campus network.

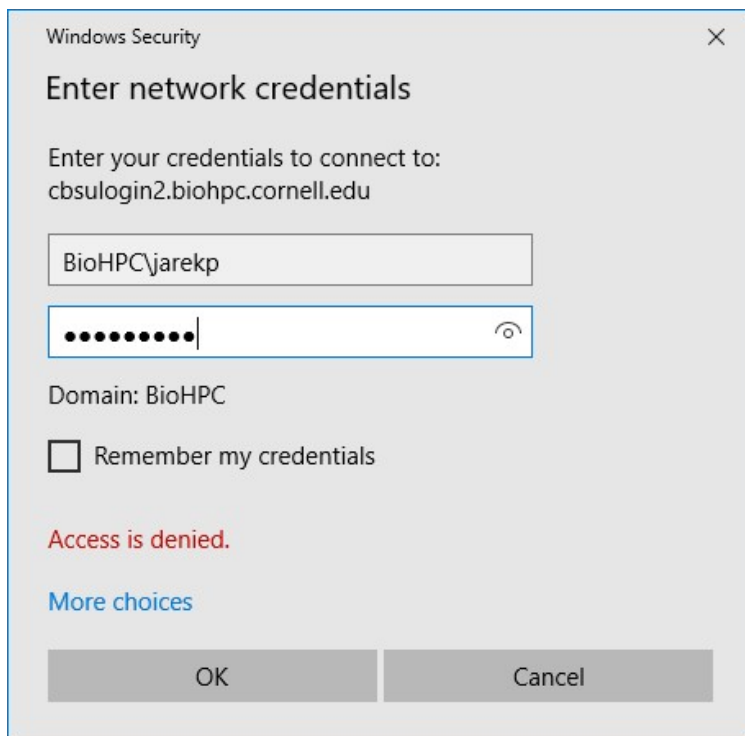
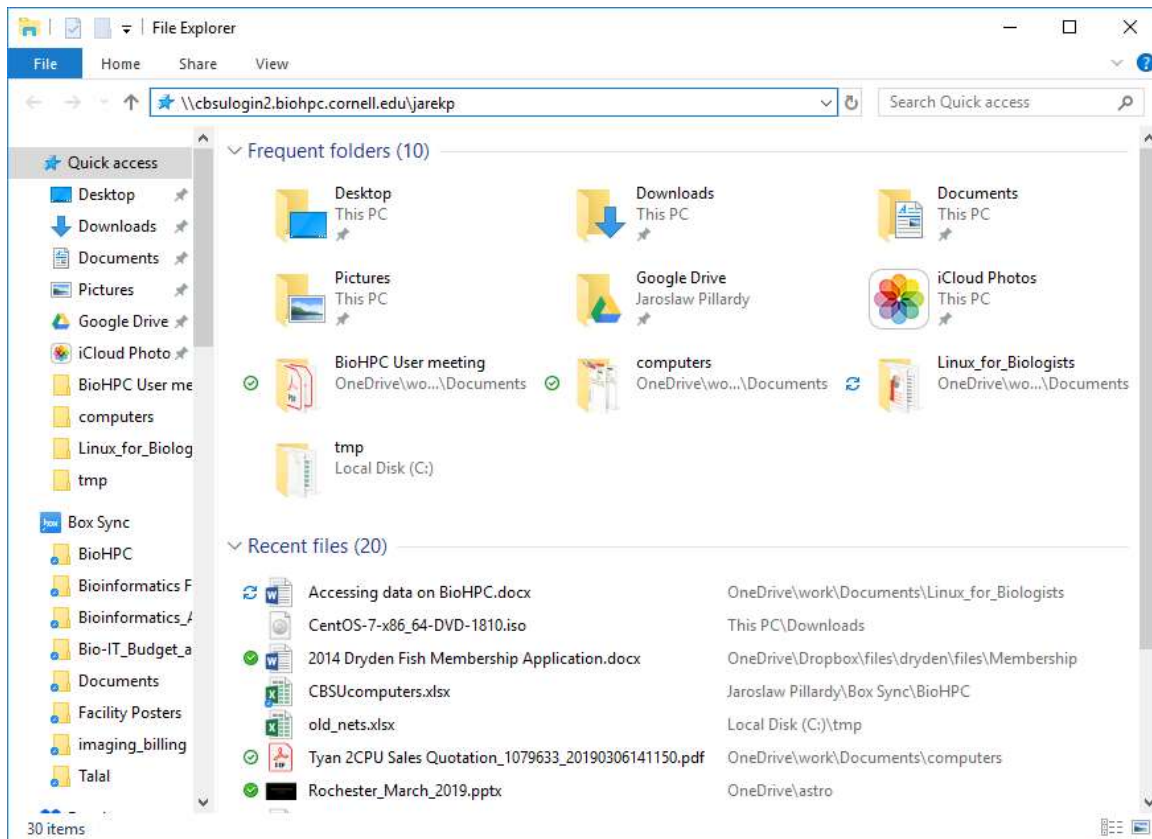


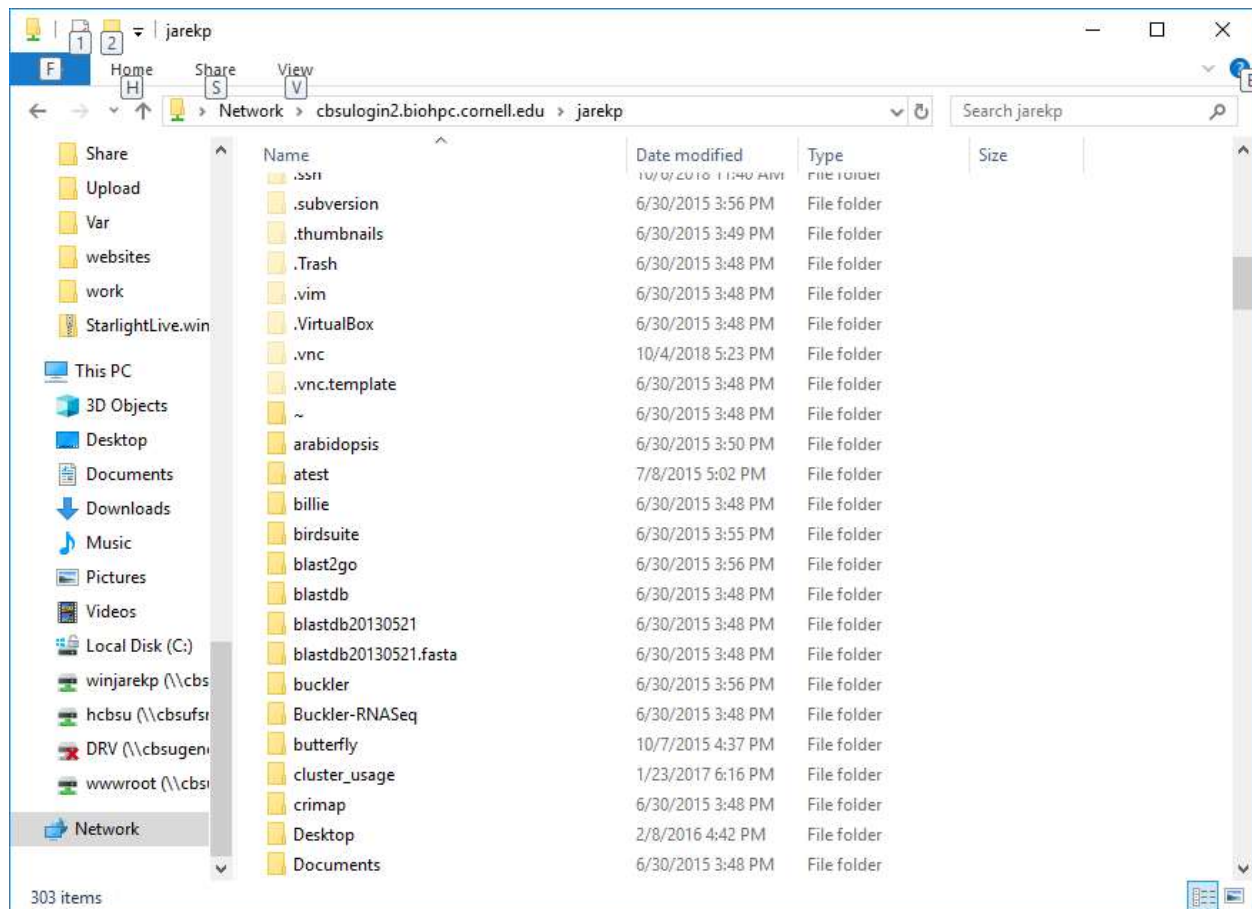
## 2. Samba (Windows share)

Samba is a Linux software enabling Linux servers to export storage as Windows shares (network drives). Samba works only from inside Cornell campus network, if connecting from outside you need to use Cornell VPN. Storage can be only accessed via BioHPC login servers, you can access your home directory or any storage group that has been added to Samba and you are a member of. Storage groups are added to Samba when requested by group owner.

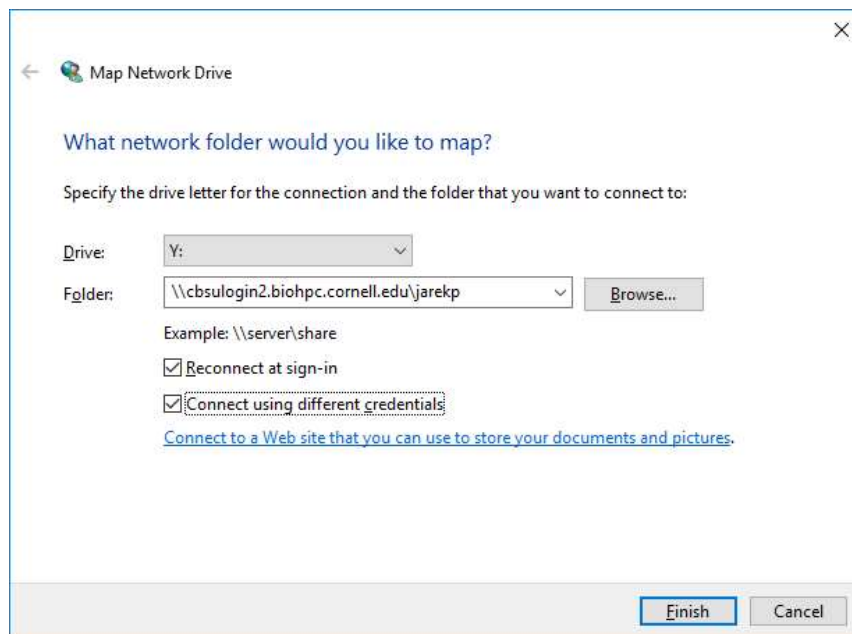
You can access the storage directly from File Explorer or you can map the share to a drive. The UNC address of the storage is `\\cbsulogin2.biohpc.cornell.edu\storagename`, where *cbsulogin2* can be replaced with *cbsulogin* or *cbsulogin3*, and *stoaregname* must be replaced with your *storage group name* or your *BioHPC user id* in case you want to access your home directory. Windows domain for the share is BioHPC, so your login id to use in Windows authentication prompt will be `BioHPC\BioHPCuserID`.

Here is an example to access home directory of user jarekp using cbsulogin2. UNC address is `\\cbsulogin2.biohpc.cornell.edu\jarekp` and the Windows authentication login id is `BioHPC\jarekp`.





You can also map the share to a Windows drive, you will need to check “Connect using different credentials” and use the same authentication (BioHPC\jarek).



### 3. 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.

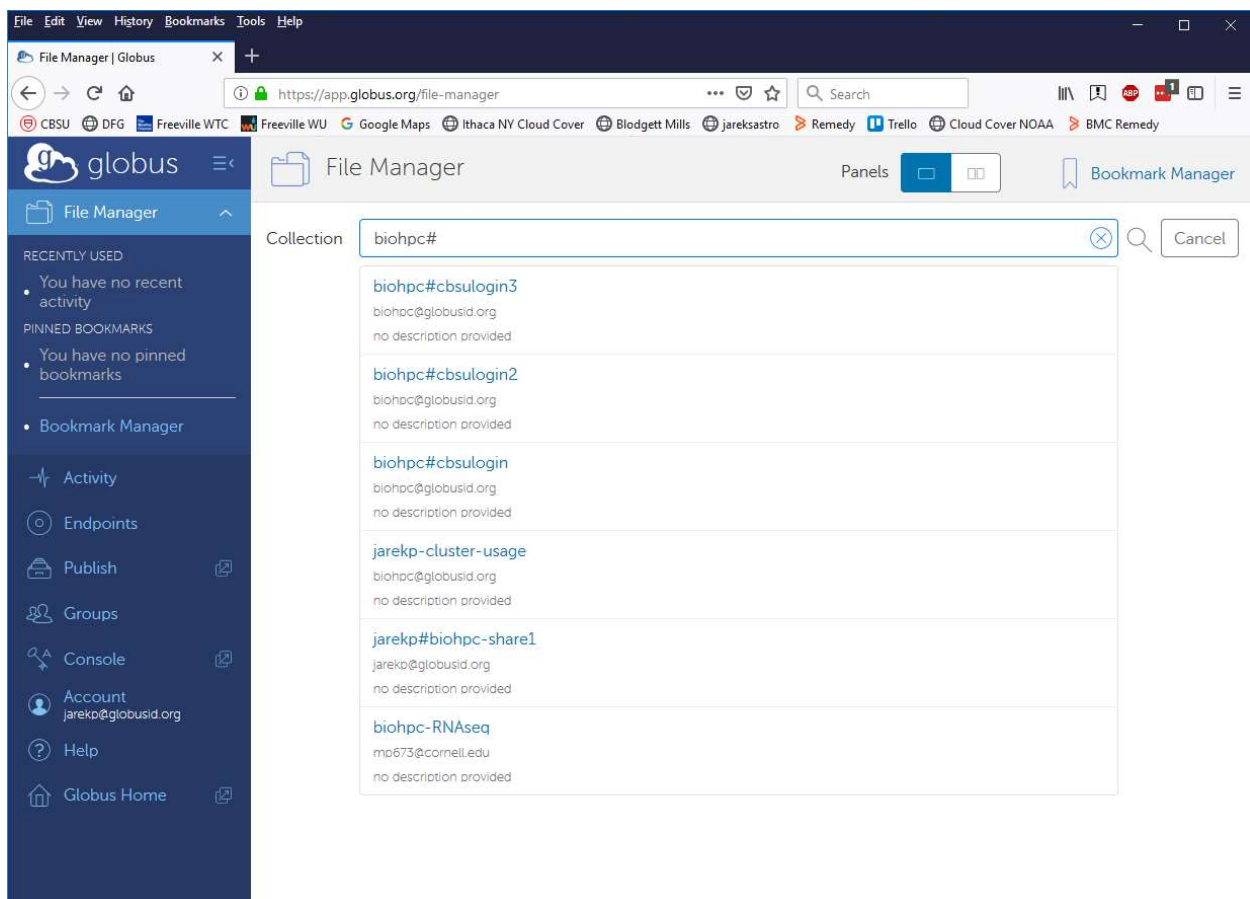
Data can be transferred between endpoints, which can be data servers or Globus Connect clients.

BioHPC Cloud Globus endpoint servers are [biohpc#cbsulogin](#), [biohpc#cbsulogin2](#) and [biohpc#cbsulogin3](#)

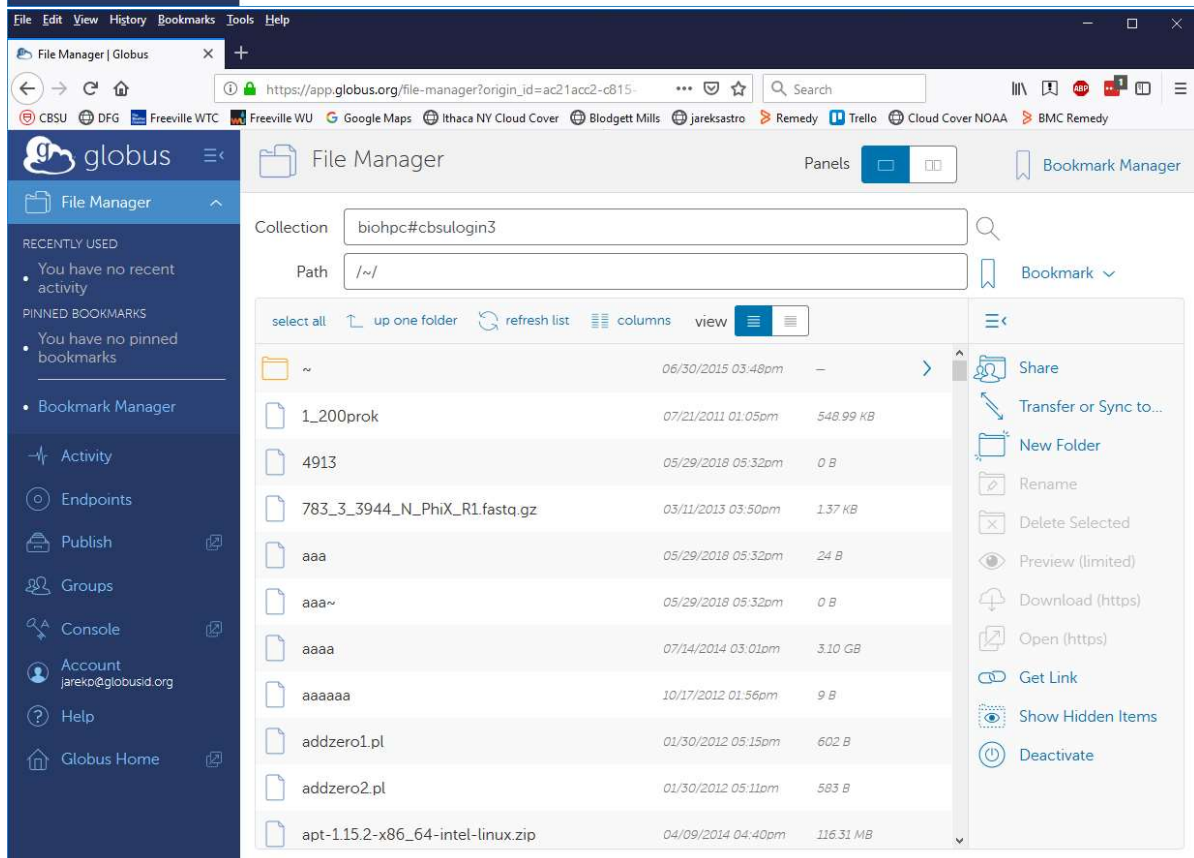
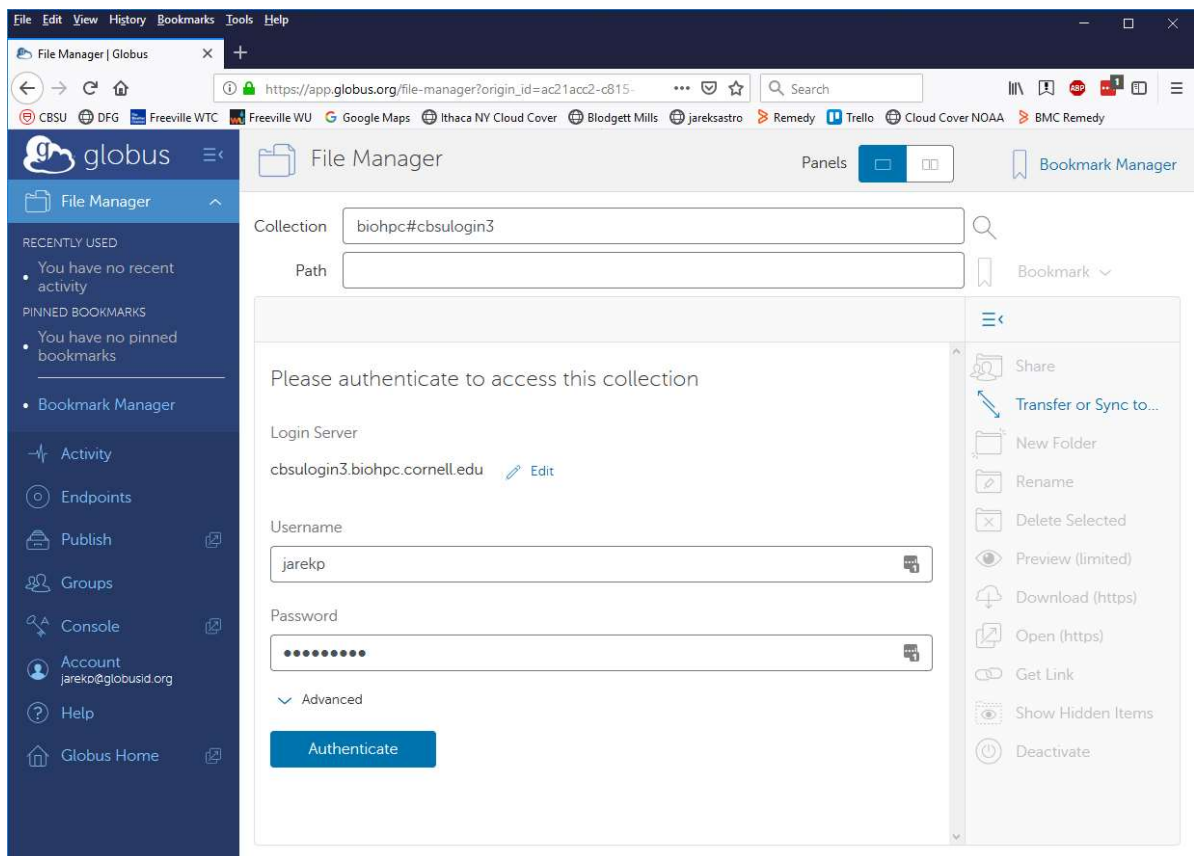
You can transfer data to another server or to your computer (using Globus Connect client).

Excellent tutorial on how to start using Globus is online on their website

<https://docs.globus.org/how-to/get-started/>







## 4. Hosted servers

All BioHPC servers have sizeable local storage which is available to all current users of the machine as /workdir. Hosted servers, i.e. servers purchased by research groups and available only to group members have also /local/storage directory available only to the group. This directory is exported and can be mounted on any other BioHPC server, including login servers, inside Cornell campus network. This directory can be only mounted on other [BioHPC servers](#), NOT any other Linux server on campus.

The command to mount the storage is

```
/programs/bin/labutils/mount_server servername /storage
```

This command must be executed on the target server where the storage is supposed to be accessed. The storage will be mounted on the target server as [/fs/servername/storage](#) .