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 <u>https://filezilla-project.org/</u>, 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.

E FileZilla							- 0	×
<u>File Edit View Transfer Server B</u> ookmarks <u>H</u> elp								
The for Ten Lunder Seven Soomark Lieb	E 🔍 🤿 🧥							
Host: Username:	Password:	Port:	Quickconnec	t 🔻				
								~ ~ ~
Local site: C:\tmp\				\sim	Remote site:			~
⊕- Python27 → Recovery → System Volume Information ⊕- temp ⊕- temp ⊕- Users ↓ var				^				
in Windows					Filename	^ Filesize Filety	pe Last modified	P
⊕ D: ⊕ G: (\\cbsufsrv8\winjarekp)								
				~		Not connected to any server		
Filename	1,647 TXT F 323 File 277,522 Adob 4,428 TXT F 23,141 SH Fil 2,174 File 350,305 Adob 488,365 Adob	rosoft Exce 2/25/20 File 2/22/20 De Acroba 2/19/20 File 2/14/20 ille 2/7/201 2/6/201 be Acroba 2/6/201 be Acroba 2/6/201 rosoft Exce 2/1/201	19 3:24:57 PM 19 2:35:24 PM 19 1:51:55 PM 19 1:51:59 PM 19 1:15:19 PM 9 1:15:10 PM 9 4:45:16 PM 9 1:21:104 PM 9 11:23:14 AM 9 11:22:05 AM 9 5:39:47 PM 9 5:39:47 PM	*	<			>
42 files. Total size: 8,536,808,317 bytes					Not connected.			
re: Inex. Total size: 0,30,000,517 bytes		Dire	ction Remote		ros comecto.	Size Priority Status		
Queued files Failed transfers Successful transfers							Queue: empty	••

FileZilla windows 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".

🛃 sftp://jarekp@cbsulogin.biohpc.comell.edu - FileZilla									×
<u>File Edit View Transfer Server B</u> ookmarks <u>H</u> elp									
出・圖言言語 🖸 🏗 🛛 🏗 🕄 🏗	i 🔍 🤗 🦚								
Host: sftp://cbsulogin.bic Username: jarekp	Password:	Port:	Quickconnect	•					
Status: Connecting to chsulogin biohpc.comell.edu Status: Connected to chsulogin Status: Retrieving directory listing Status: Listing directory/home/jarekp Status: Directory listing of "/home/jarekp" successful									0
Local site: C:\tmp\				~	Remote site: /home/jarekp				~
Python27				^	□- <mark>?</mark> /				
← Recovery System Volume Information ⊕ L temp ← Users ← Var					P i home jarekp				
				7	Filename	Filesine	Filebase	Last modified	0
🕀 🔐 D:				11/		Filesize	Filetype	Last modified	
🗃 🛫 G: (\\cbsufsrv8\winjarekp)				1			File folder	8/28/2015 4:27:50	
ii - 👳 Z: (\\cbsusrv06\www.root)				~	.aspera .byobu		File folder	6/30/2015 3:48:25	
Filename	El	Classica.	Last modified	~	.cache		File folder	2/6/2019 4:18:50	
	Filesize	ritetype	Last modified		config		File folder	1/30/2018 11:38:1	
					dbus		File folder	6/30/2015 3:53:25	
labcomputers_new.xlsx		Microsoft Exce					File folder	6/30/2015 3:48:32	
ENTO.txt		TXT File	2/22/2019 2:35:24 PM		.eggcups .emacs.d		File folder	6/30/2015 3:56:52	
attachment1	323		2/22/2019 1:51:55 PM		filezilla		File folder	2/1/2017 4:53:05	
A newnet.pdf		Adobe Acroba	2/19/2019 3:59:47 PM		fontconfig		File folder	6/30/2015 3:48:30	
panandassembly.txt		TXT File	2/14/2019 1:15:19 PM		gconf		File folder	2/28/2019 2:42:09	
parallel_studio_xe_2015_update3_online.sh BLOSUM100	23,141		2/7/2019 12:11:04 PM		gconfd		File folder	11/25/2015 9:35:4	
	2,174		2/6/2019 4:45:16 PM 2/6/2019 4:45:01 PM		globus		File folder	6/30/2015 3:57:15	
aaa					gnome		File folder	6/30/2015 3:56:11	
UPS-seagate-20190206.pdf ups_label_20190206.pdf		Adobe Acroba	2/6/2019 11:23:14 AM		.gnome2		File folder	11/4/2015 11:05:1	
desktop_network_support_rates.xlsx			2/1/2019 5:39:47 PM		.gnome2_private		File folder	6/30/2015 3:48:24	
IIDS_ceanate_20100201 pdf		Adobe Acroba	2/1/2019 3:39:47 PM	~	anote		File folder	6/30/2015 3:56:44	
<			>		<				>
42 files. Total size: 8,536,808,317 bytes					163 files and 139 directories. Total size: 12,796,224,016 bytes				
Server/Local file			Direction Remote f	ile	Size Priority	Status			
Queued files Failed transfers Successful transfers							0.0		
							🧰 🕐 Qu	ieue: empty	

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.

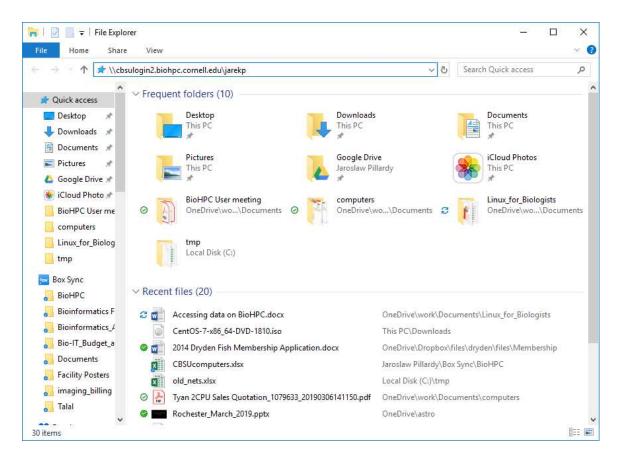
atus: Disconnected from server	Password: Port:	Quickconnect	•				
atus Disconnected rom server Connected to clsulogin atus Connected to clsulogin atus Retreining directory listing atus Directory listing of '/horme/jarelp'' successful atus Retreining directory listing of '/fs/cbsubscb09/st atus Listing directory listing of '/fs/cbsubscb09/storage' atus Directory listing of '/fs/cbsubscb09/storage' atus Directory listing of 'r/s/cbsubscb09/storage' atus Directory listing of 'r/s/cbsubscb09/storage' suc							
ocal site: C:\tmp\			~	Remote site: /fs/cbsubscb09/storage			
Python27 System Volume Information System Volume Information temp Users Var			^	G-?/ C-?/ F Coubscb09 B- storage G-?/ home B- jarekp			
Windows			-	Filename	Filesize Filetype	Last modified	-
⊕ D: ⊕ G: (\\cbsufsrv8\winjarekp)							
Z: (\\cbsusrv06\wwwroot)				amy amy	File folder	1/14/2019 3:19:32 PM	N
			~	archive .	File folder	5/22/2017 6:40:44 PM	M
ilename	Filesize Filetype	Last modified	^		File folder	2/25/2019 2:44:48 PM	N
				daniel	File folder	11/16/2018 3:57:07 P	
labcomputers_new.xlsx	10,318 Microsoft Exce	2/25/2019 3:24:57 PM		📴 data	File folder	8/30/2018 7:32:53 PM	
ENTO.txt	1,647 TXT File	2/22/2019 2:35:24 PM		ian	File folder	1/28/2019 3:35:03 PM	
] attachment1	323 File	2/22/2019 1:51:55 PM		jens	File folder	2/27/2019 3:36:46 PM	
newnet.pdf	277,522 Adobe Acroba	2/19/2019 3:59:47 PM		madison	File folder	1/5/2019 12:44:33 PM	
panandassembly.txt	4,428 TXT File	2/14/2019 1:15:19 PM		melissa	File folder	3/6/2018 10:59:24 AM	
parallel_studio_xe_2015_update3_online.sh	23,141 SH File	2/7/2019 12:11:04 PM		Minsoo	File folder	2/16/2018 4:51:07 PM	
BLOSUM100	2,174 File	2/6/2019 4:45:16 PM		monica	File folder	2/25/2019 6:03:14 PM	
388	24 File	2/6/2019 4:45:01 PM		olivia	File folder	12/1/2017 12:57:36 P	
UPS-seagate-20190206.pdf	350,305 Adobe Acroba			ramya	File folder File folder	5/8/2018 6:26:43 PM	
ups_label_20190206.pdf	488,365 Adobe Acroba			resources siddharth	File folder	11/28/2018 1:35:32 P 2/20/2019 1:26:21 PM	
desktop_network_support_rates.xlsx	10,978 Microsoft Exce 246.675 Adobe Acroba		~	storage	File folder	5/14/2016 1:47:17 AM	
	JANKA MARE ACTORS	>		<	The folder	3, 14, 2010 114111 14	>
				18 directories			
2 files. Total size: 8,536,808,317 bytes		Direction Remote fi			Size Priority Status		

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.



Windows Security	×
Enter network credentia	als
Enter your credentials to conne cbsulogin2.biohpc.cornell.edu	ect to:
BioHPC\jarekp	
•••••	୕
Domain: BioHPC	
Remember my credentials	
Access is denied.	
More choices	
OK	Cancel

1 2 + Jarekp Home Share	View				
	vork > cbsulogin2.biohpc.cornell.edu	> jarekp	v Ö	Search jarekp	
Share ^	Name A	Date modified	Туре	Size	
Upload	subversion	6/30/2015 3:56 PM	File folder		
Var	thumbnails	6/30/2015 3:49 PM	File folder		
websites	.Trash	6/30/2015 3:48 PM	File folder		
work	.vim	6/30/2015 3:48 PM	File folder		
StarlightLive.win	.VirtualBox	6/30/2015 3:48 PM	File folder		
	.vnc	10/4/2018 5:23 PM	File folder		
This PC	.vnc.template	6/30/2015 3:48 PM	File folder		
🔰 3D Objects		6/30/2015 3:48 PM	File folder		
Desktop	arabidopsis	6/30/2015 3:50 PM	File folder		
Documents	atest	7/8/2015 5:02 PM	File folder		
Downloads	🔜 billie	6/30/2015 3:48 PM	File folder		
Music	📊 birdsuite	6/30/2015 3:55 PM	File folder		
Pictures	📙 blast2go	6/30/2015 3:56 PM	File folder		
	hlastdb	6/30/2015 3:48 PM	File folder		
Videos	📙 blastdb20130521	6/30/2015 3:48 PM	File folder		
🔓 Local Disk (C:)	📙 blastdb20130521.fasta	6/30/2015 3:48 PM	File folder		
🛫 winjarekp (\\cbs	o buckler	6/30/2015 3:56 PM	File folder		
🛫 hcbsu (\\cbsufsi	Buckler-RNASeq	6/30/2015 3:48 PM	File folder		
👷 DRV (\\cbsugen	butterfly	10/7/2015 4:37 PM	File folder		
www.root (\\cbs	cluster_usage	1/23/2017 6:16 PM	File folder		
	📊 crimap	6/30/2015 3:48 PM	File folder		
Network	Desktop	2/8/2016 4:42 PM	File folder		
~	Documents	6/30/2015 3:48 PM	File folder		

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).

~		>
😪 Map N	letwork Drive	
What ne	etwork folder would you like to map?	
Specify the	e drive letter for the connection and the folder that you want to connect to:	
<u>D</u> rive:	Y: ~	
F <u>o</u> lder:	\\cbsulogin2.biohpc.cornell.edu\jarekp v <u>B</u> rowse	
	Example: \\server\share	
	Reconnect at sign-in	
	Connect using different credentials	
	Connect to a Web site that you can use to store your documents and pictures.	
	Finish	Cancel
	Lunsu	Cancer

3. Globus

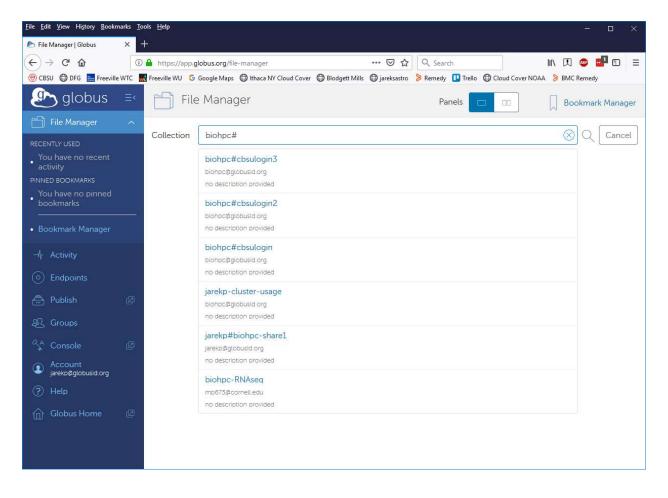
Globus Online is an online system based on GridFTP engine that supports fast and reliable data transfer and can be scheduled and controlled online in a web browser.

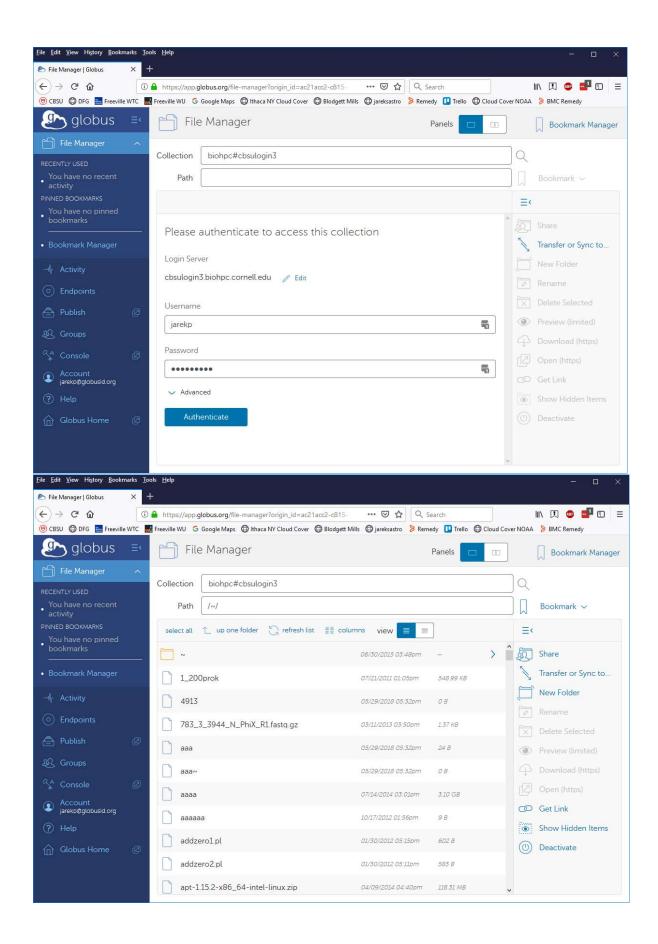
Data can be transferred between endpoints, which can be data servers or Globus Connect clients.

BioHPC Cloud Globus endpoint servers are is 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 .