



GW COLONIAL ONE



HPC Workshop 1

What we're covering:

- Who to contact for help
- The onboarding process
- Logging in
- Cluster architecture
- Navigating the shell
- Modules, environment variables and .profile
- Quotas
- Purges
- File transfer and management:
 - scp
 - Globus
 - Lustre vs NFS filesystems
- SLURM
 - Your first submit script



HPC - Getting Help

Email hpchelp@gwu.edu for any issues you have with Colonial One.

Individual schools have Local Support Partners who support users in their community. Here are the LSPs for various schools:

- CCAS: Glen Maclachlan, William Bonnett
- SEAS: Jason Hurlburt, Marco Suarez, Zacharie Day
- SMHS: Dacien Reese-Stremtan
- SPHS: Joseph Creech
- Computational Biology Institute: Adam Wong
- DIT: William Burke and Terrence Lewis

All of these people watch and respond to the hpchelp@gwu.edu mailing list.



HPC - Onboarding

I want to use Colonial One! What is the onboarding process?

1. Request a Colonial One account from the HPC Support Team:

<https://colonialone.gwu.edu/getting-access/>

2. Your account is created, storage is allocated, and you are assigned to a PI group, or the nopi group if you do not have a faculty sponsor. Users receive a welcome email once the account is created with links to the Colonial One web site.
3. The user's Local Support Partner can provide short tutorials on the cluster and how to use it for your research project if requested.
4. Continued support is provided via the hpchelp@gwu.edu support email list.

A small image in the top-left corner shows a campus scene with people walking on a path and a building in the background.

HPC - Exercise 1. Logging In

Requirements:

- SSH Client
- Colonial One account

Log into Colonial One:

ssh [username@login.colonialone.gwu.edu](https://login.colonialone.gwu.edu)

Use your NetID and password!

HPC Workshop - HPC @ GWU

Colonial One





HPC Workshop - HPC @ GWU

Current Specs:

- Dell C8220 cluster, 262 nodes
- 53x GPU nodes, 2x NVIDIA K20 GPUs, 12x GPU nodes, 1x 12GB P100
- 1x 2TB Node, Quad 12-Core 3.0GHz Xeon E7-8857v2 CPUs
- 196 x CPU nodes, 2x 2.6GHz 8-core Xeon CPUs, 64/128/256GB of RAM

Totals:

- 4k Intel Xeon CPU cores & >350,000 NVIDIA CUDA cores
- over 33 TB of RAM
- Mellanox FDR Infiniband fabric
- Two primary filesystems
- 262 TB NFS fileservers for /home and /groups
- 262 TB Lustre filesystem for high-speed scratch
- 268 TB Dell Compellent for archival



HPC Workshop - HPC @ GWU

Colonial One:

- Serves over 1000 users in nearly 150 research groups
- Runs 24/7, 365 days a year
- Processes > 2,000 jobs every day
- User demand is 91% of capacity
- Open to entire GW community
- 129 open proposals for funding reference Colonial One

Jobs Run on Colonial One:

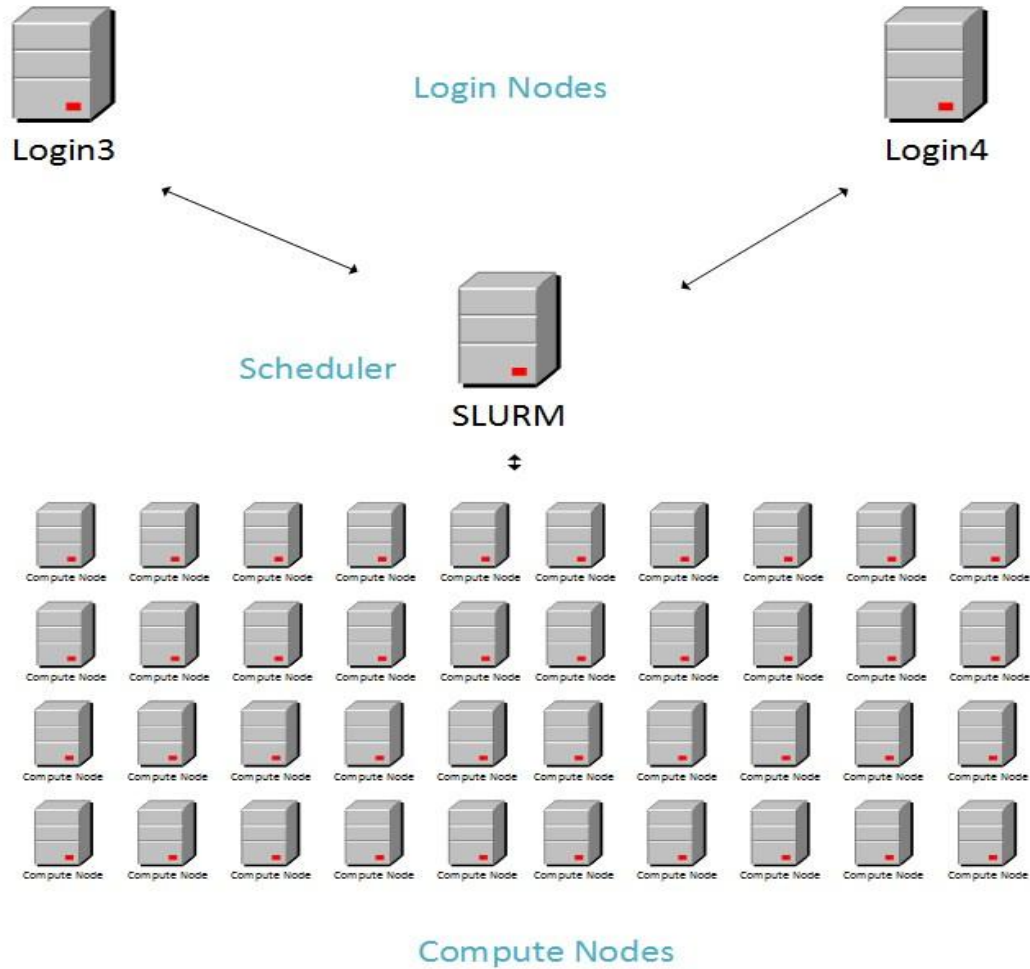
- Study structure of subatomic particles
- Large-scale molecular dynamics simulations
- Network analysis
- Drug design for cancer therapy
- Protein engineering for immune response against bacteria and viruses including HIV/AIDS
- fMRI analyses of injured brains
- Genomic sequencing
- Phylogenetic mapping of evolutionary traits
- Satellite imagery
- Population and census dynamics



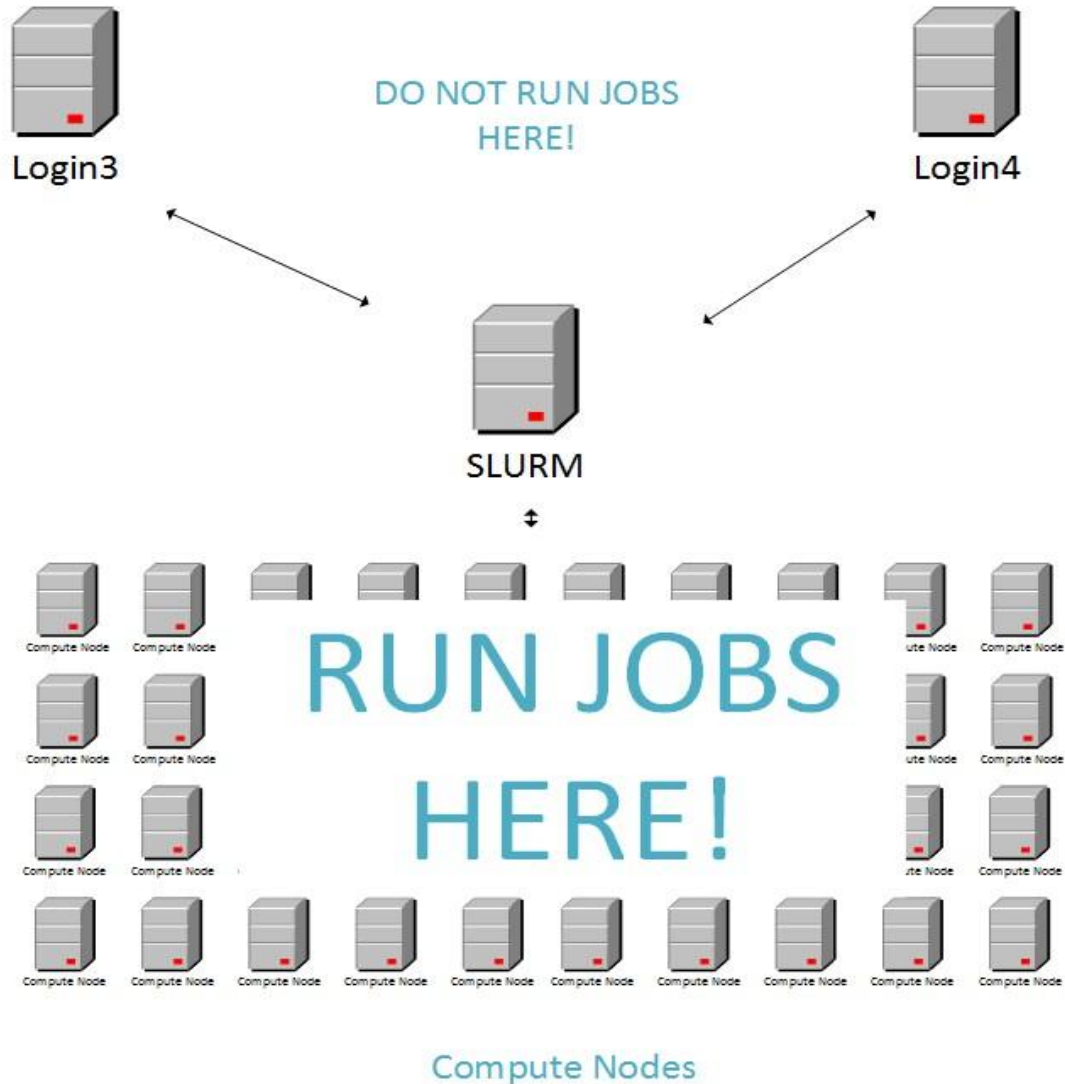
Cluster Architecture

- Login Node - Server that acts as your interface to the cluster
- Scheduler - Server that schedules jobs
- Compute Nodes - Servers that run jobs

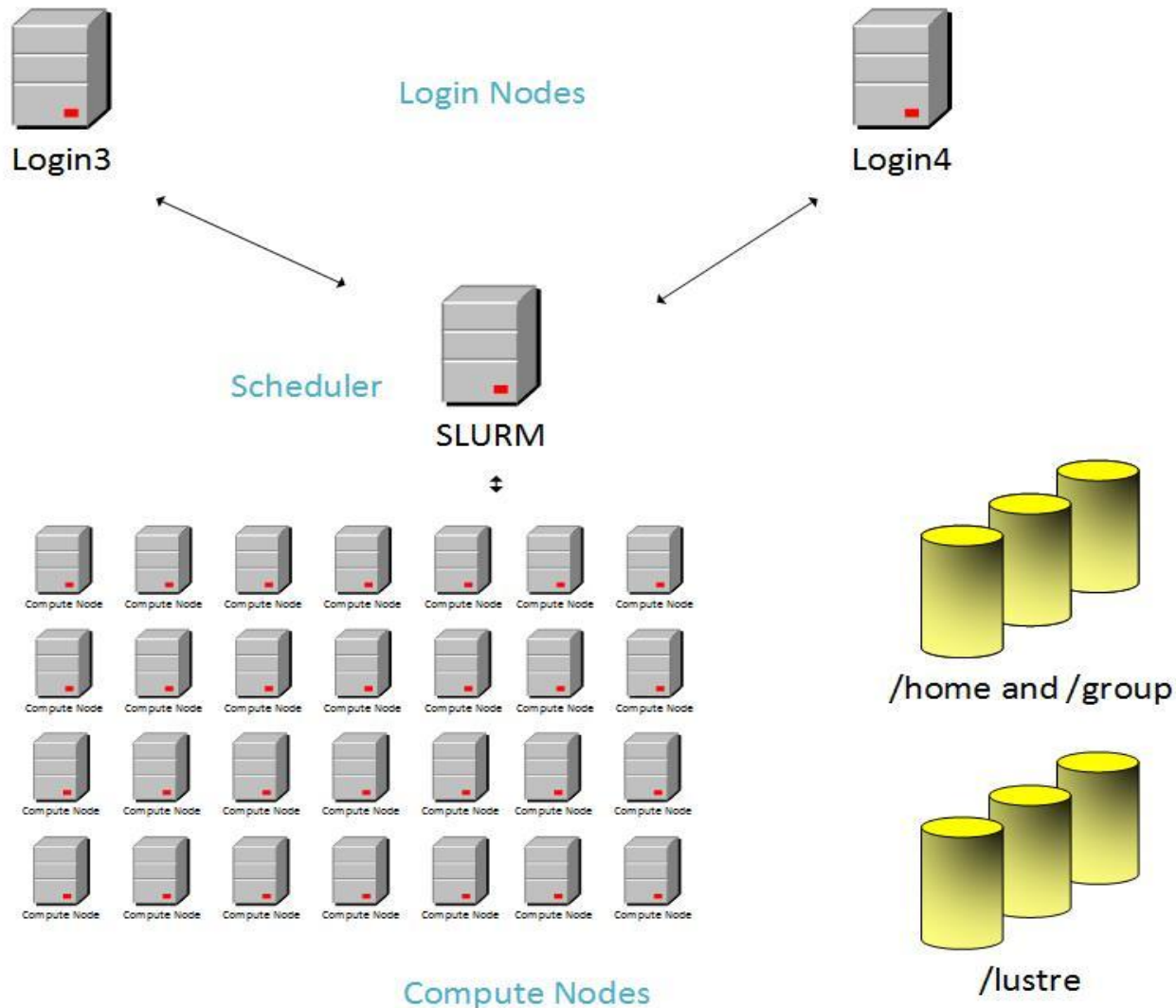
Cluster Architecture



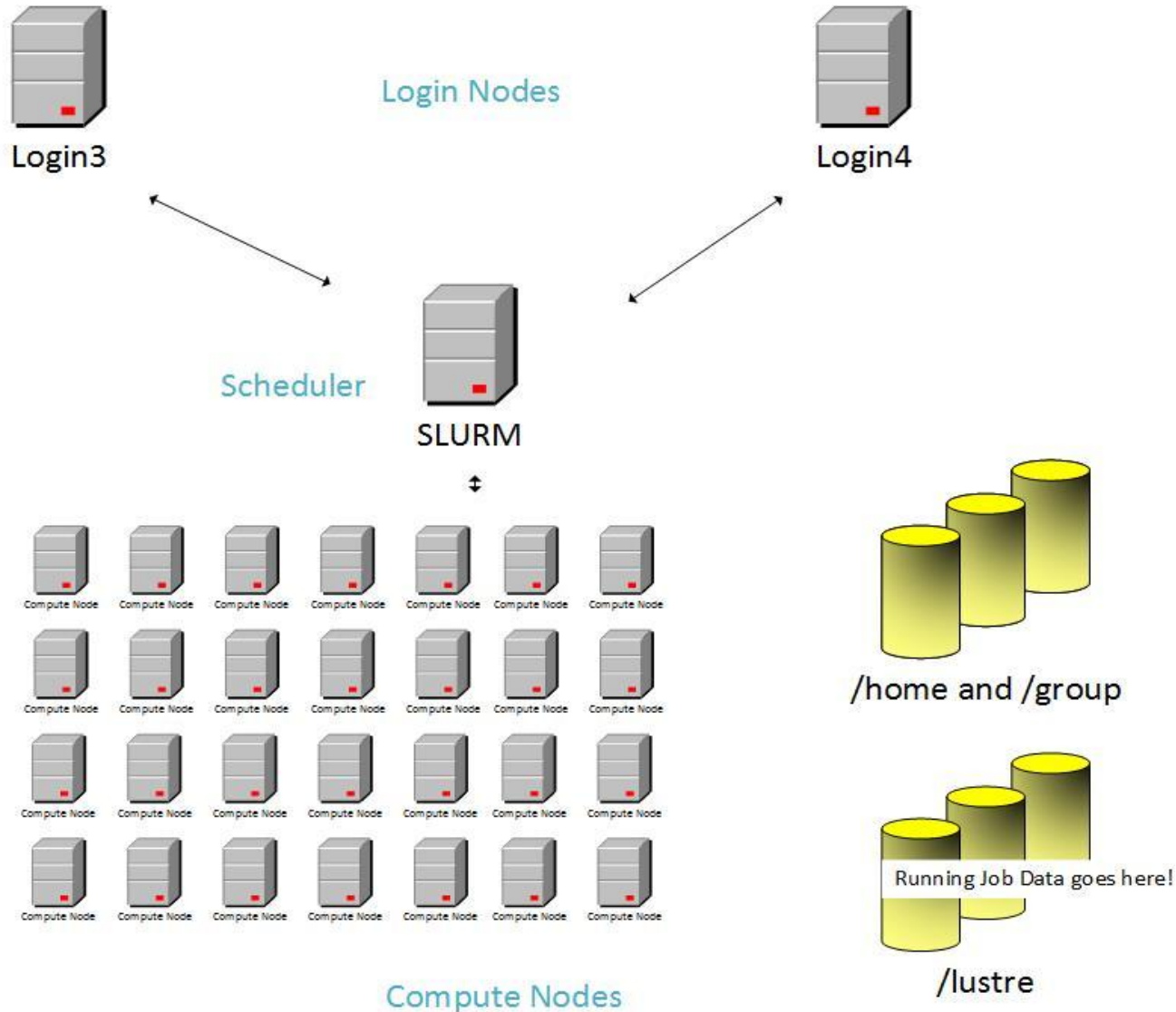
Cluster Architecture



Cluster Architecture (Storage)



Cluster Architecture (Storage)





HPC - Navigating the shell

Pathname

- A path through the directory system
- *pwd* – shows current path
- Absolute vs. Relative path

/ - the forward slash

- Represents the very bottom (root) of the file system
- acts as a divider in between directories on the file system



HPC - Navigating the Shell

- `pwd`: Print Working Directory, shows you where you are
- `.` versus `..` : Your current directory versus the directory one level above
- The `~` character: Shortcut your home directory
- `ls`: list current path contents
- `ls -la`: list all details of the current path in long form
- `cd`: change directory
 - `cd /absolute/path`
 - `cd path/relative/to/where/I/am`



HPC - Exercise 2. Navigating the Shell

1. Type: `cd ~`
2. Type: `nano testfile.txt`
3. Add some text, CTRL+O to save, CTRL+X to exit
4. Type: `mkdir testdir`
5. Type: `ls -la test*`
6. Type: `mv testfile.txt testdir/`
7. Type: `ls -la test*`
8. Type: `cd testdir`
9. Type: `ls -la test*`
10. Type: `cd ~`



HPC - Navigating the Shell

Questions & Discussion



HPC - Modules

Modules load an environment so a program can run correctly.

Module commands:

- module list
- module avail
- module load
- module unload
- module spider



HPC - Exercise 3. Modules

1. From the command prompt, type: *R*
2. The command is not found!
3. Type: *module load R/3.4.2*
4. Type: *R*
5. Notice R loads!
6. Type: *quit()* to exit R
7. Choose No when asked to save workspace



Environment variables

Environment variables are a set of dynamic named values that can affect the way running processes will behave on a computer. They are part of the environment in which a process runs.

Environment commands:

- *printenv*
- *printenv Variable_Name*
- *echo \$Variable_Name*
- *export Variable_Name=Value*



Exercise 4. Environment variables

1. Check path variable by typing: *echo \$PATH*
2. type: *module load python/3.4.2*
3. type: *echo \$PATH*
4. Notice */c1/apps/python/3.4.2/bin* has been added to your path.
5. type: *python*
6. Once python loads, type: *1+1*
7. Type: *quit()* to exit python



HPC - Shell Configuration Files

- `.bashrc`: Runs when logging into a BASH session. Local to the BASH shell
 - You can enter the same commands inside `.bashrc_profile` as you can inside `.profile`
 - Add a welcome message to your `.bashrc` file:
 - Type: `nano ~/.bashrc`
 - Type: `echo "HELLO!"`
 - `CTRL+O` to save file
 - `CTRL+X` to exit file
- Other shells have similar names: `.cshrc` (C shell), `.ksh` (Korn shell)



HPC - Quotas

Home and Group Quotas

- Soft quotas are in place now
- Home quota: /home/username - default 25GB
- Group quota: /groups/groupname - default 250GB
- Check quota: type *quotareport* at the shell

Colonial One is not meant for archival data. Please remove data from old jobs once you finish your project.



HPC - Purge

What data is purged?

- Home and Group shares are not purged
- We will ask users to delete data from Home and Group if they are using too much space
- The high speed lustre file system IS purged every month
- Lustre is to be used for scratch space while running jobs

When is data purged?

- At the beginning of every month



HPC - Purge Policy

Lustre Purge Policy Coming into Effect 3/1/2017

1. **Frequency:** A purge will be conducted on the first day of every month (starting on **3/1/2017**). In the past, purges have been scheduled based on how close lustre utilization was to capacity. In the updated procedure, a purge will be conducted irrespective of lustre utilization. Again, a purge will be conducted on the first of every month even if the 1st falls on a weekend or holiday.
2. **File Access Time:** All files whose access time is greater than 60 days will be subject to purging. NOTE: updating access times with the sole intent of circumventing purging of files may result in disciplinary action including account suspension.
3. **File Size:** Files will be subject to purging regardless of the size they occupy on disk.



HPC - File Transfer with SCP

SCP - Secure Copy - Command line copy tool

Use to copy files from one *nix machine to another

- Usage:
 - `scp from [...] to`
 - `scp <sourcefile> <destfile>`
 - `scp host:<sourcefile> <destfile>`
 - `scp user@host:<sourcefile> <destfile>`
- Syntax is like `cp`
 - `-r` flag to recursively copy directories
 - `man scp` for more options



HPC - File Transfer with Globus

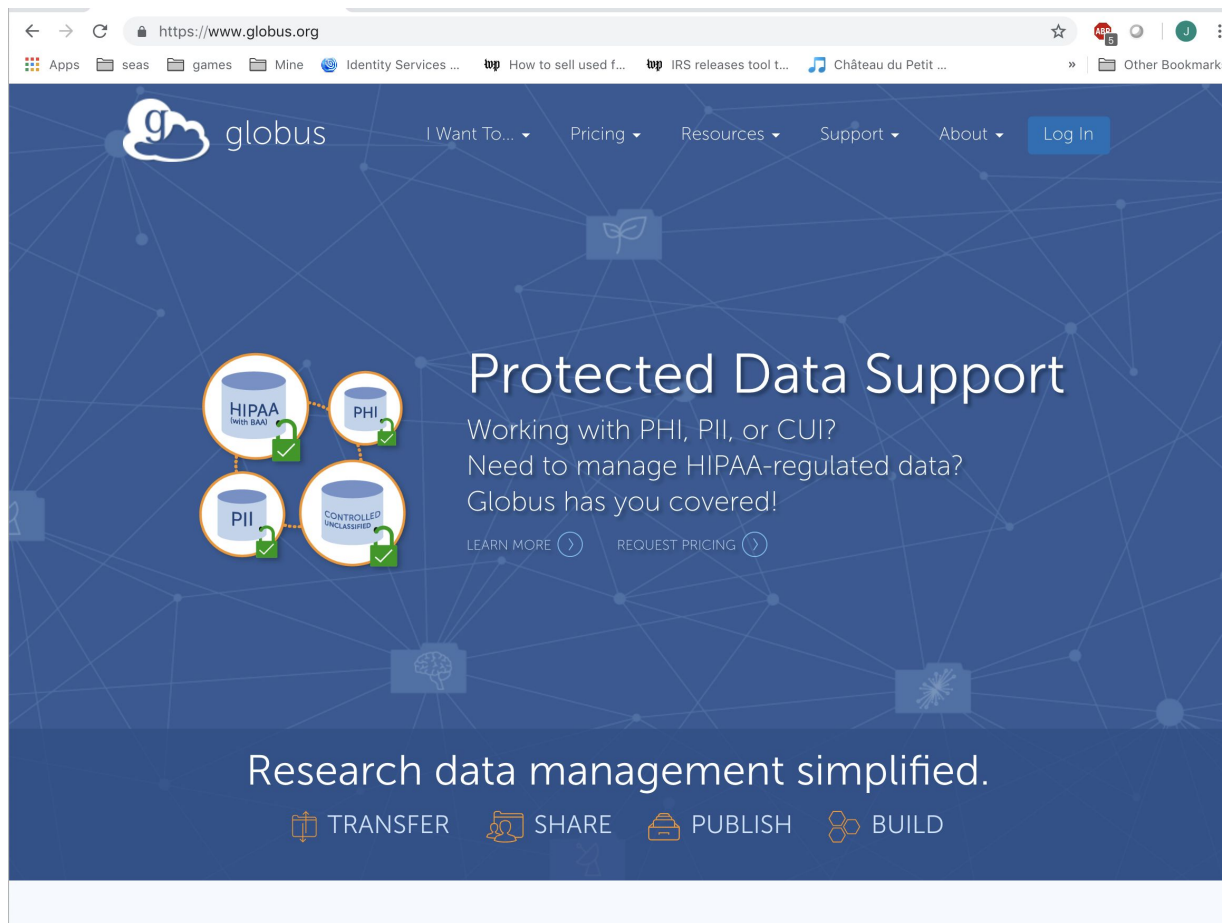
Globus is the industry standard for transferring large amounts of science and engineering research data between datacenters and endpoints.

Key points are:

- Built on GridFTP technology.
- Data Transfers can be encrypted in flight (not encrypted by default).
- Transfer run in the background and can be interrupted and restarted, even if a file is partially transmitted.
- Free for individuals (institutes must pay to use the service).
- Globus is used literally everywhere by everyone transferring data in the HPC world. AWS, National Labs, National Supercomputing Centers, Universities, and even GWU!

HPC - File Transfer with Globus

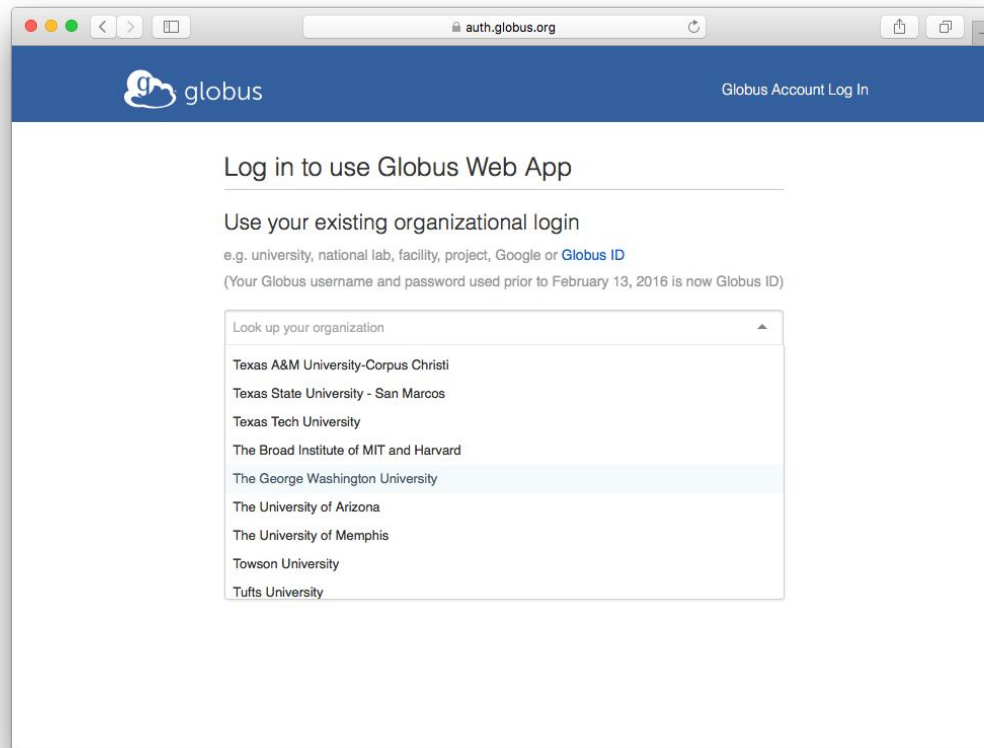
Open a web browser and go to: www.globus.org



The screenshot shows the Globus website homepage. The browser address bar displays <https://www.globus.org>. The navigation menu includes "I Want To...", "Pricing", "Resources", "Support", "About", and a "Log In" button. The main content area features a "Protected Data Support" section with the text: "Working with PHI, PII, or CUI? Need to manage HIPAA-regulated data? Globus has you covered!". Below this text are two buttons: "LEARN MORE" and "REQUEST PRICING". To the left of the text are four icons representing data types: "HIPAA (with BAA)", "PHI", "PII", and "CONTROLLED UNCLASSIFIED". At the bottom of the page, the text "Research data management simplified." is displayed above four icons labeled "TRANSFER", "SHARE", "PUBLISH", and "BUILD".

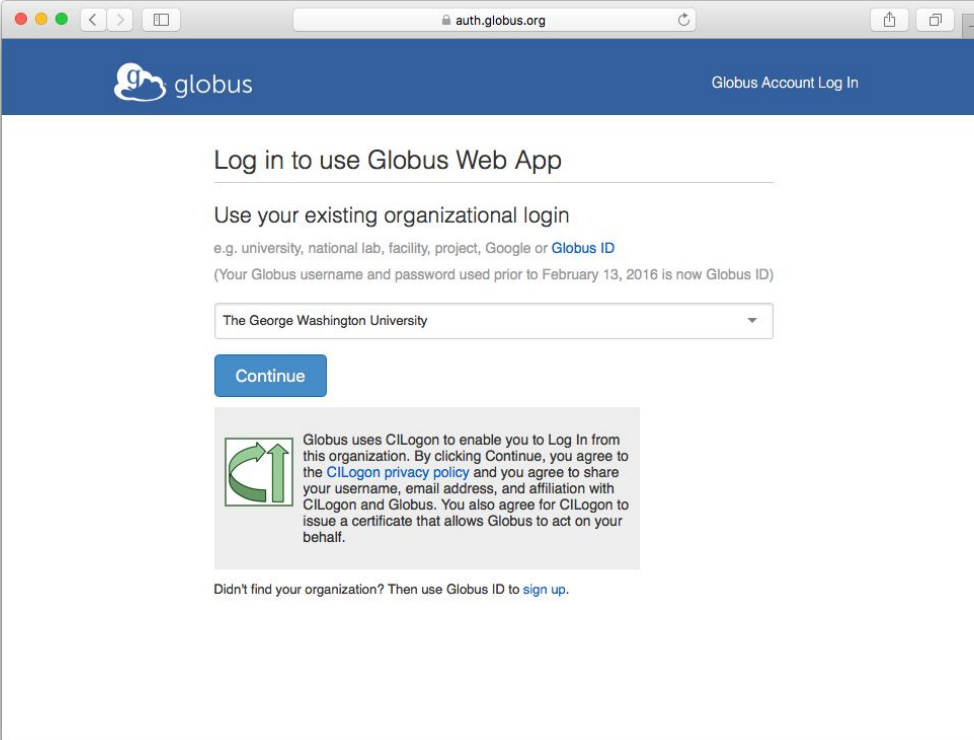
HPC - File Transfer with Globus

Search the list for The George Washington University



HPC - File Transfer with Globus

Click Continue



auth.globus.org

globus Globus Account Log In


Log in to use Globus Web App

Use your existing organizational login

e.g. university, national lab, facility, project, Google or [Globus ID](#)
(Your Globus username and password used prior to February 13, 2016 is now Globus ID)

The George Washington University

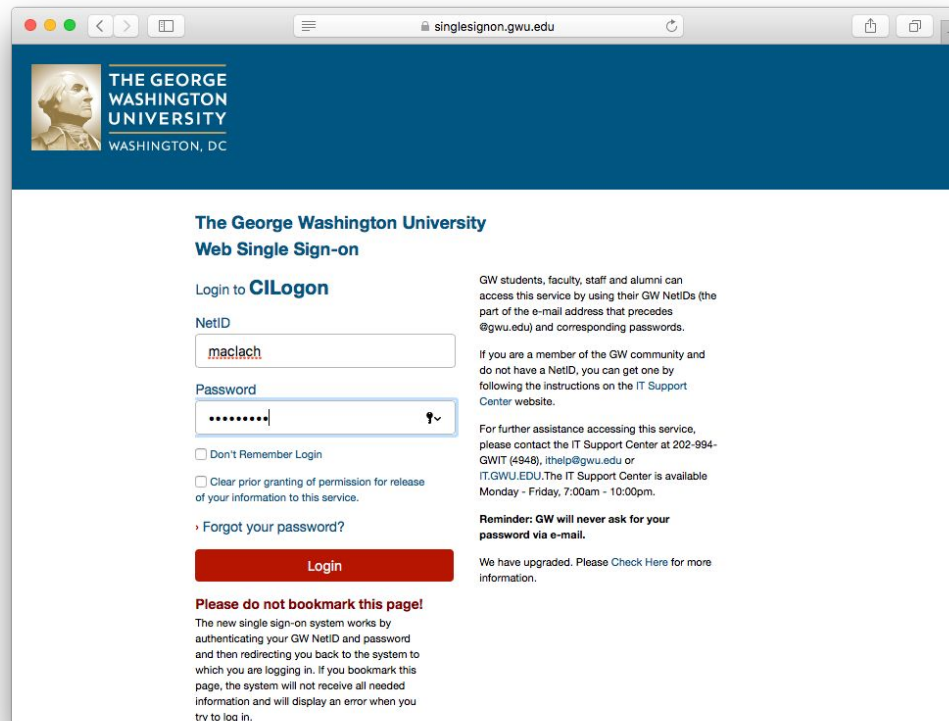
Continue

 Globus uses CILogon to enable you to Log In from this organization. By clicking Continue, you agree to the [CILogon privacy policy](#) and you agree to share your username, email address, and affiliation with CILogon and Globus. You also agree for CILogon to issue a certificate that allows Globus to act on your behalf.

Didn't find your organization? Then use Globus ID to [sign up](#).

HPC - File Transfer with Globus

Sign in with your NetID and password



THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

The George Washington University Web Single Sign-on

Login to **CILogon**

NetID

Password

Don't Remember Login

Clear prior granting of permission for release of your information to this service.

[Forgot your password?](#)

Login

Please do not bookmark this page!
The new single sign-on system works by authenticating your GW NetID and password and then redirecting you back to the system to which you are logging in. If you bookmark this page, the system will not receive all needed information and will display an error when you try to log in.

GW students, faculty, staff and alumni can access this service by using their GW NetIDs (the part of the e-mail address that precedes @gwu.edu) and corresponding passwords.

If you are a member of the GW community and do not have a NetID, you can get one by following the instructions on the IT Support Center website.

For further assistance accessing this service, please contact the IT Support Center at 202-994-GWIT (4948), ithelp@gwu.edu or IT.GWU.EDU. The IT Support Center is available Monday - Friday, 7:00am - 10:00pm.

Reminder: GW will never ask for your password via e-mail.

We have upgraded. Please [Check Here](#) for more information.

HPC - File Transfer with Globus

Click Continue



Welcome – You've Successfully Logged In

This is the first time you are accessing Globus with your **The George Washington University** login.

If you have previously used Globus with another login you can link it to your **The George Washington University** login. When linked, both logins will be able to access the same Globus account permissions and history.

Continue

Link to an existing account

[Why should I link accounts?](#)

HPC - File Transfer with Globus

Click on the Globus Icon to return to the home page



Account ▾

[< Return to Account Identities on the Globus Web App](#)

Identities

This is a list of identities linked to your Globus account. You can view details and unlink identities. Applications that do not require an identity from a particular provider will use the primary identity by default.

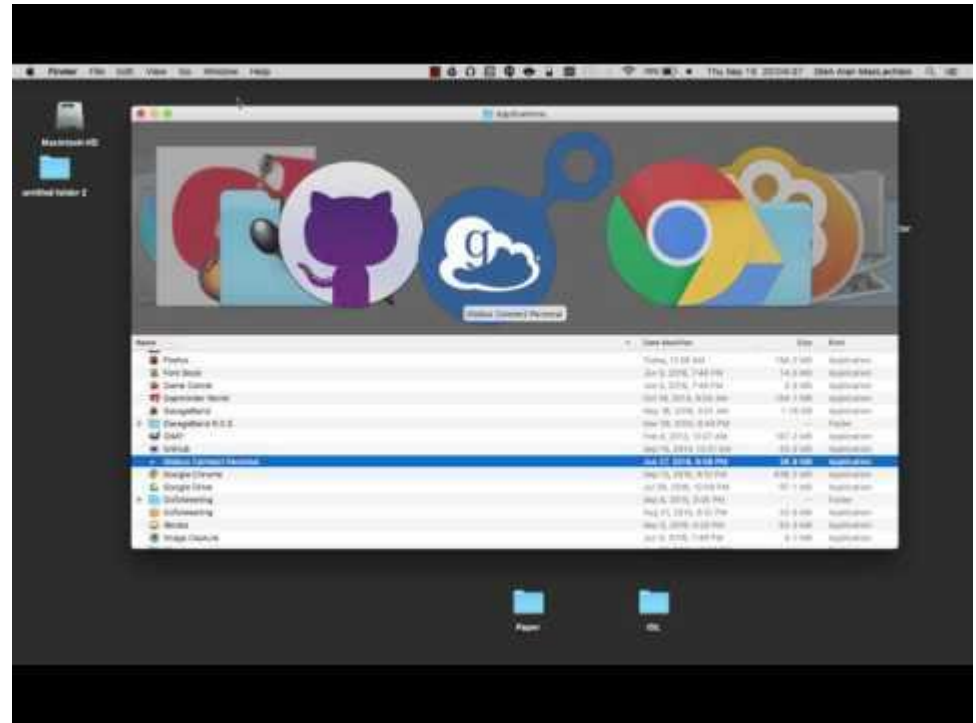
[+ link another identity](#)

The George Washington University (hurlburj@gwu.edu)	primary	⋮
---	---------	---

HPC - File Transfer with Globus

Setting up a personal endpoint...

1. Go to EndPoints
2. Add globus personal connect endpoint and name it
3. generate and copy set up key
4. Download installer and install.
5. Open app and paste setup key
6. Go back to webpage and find endpoint



HPC - File Transfer with Globus

The screenshot displays the Globus File Manager interface. On the left is a dark blue sidebar with navigation options: File Manager, RECENTLY USED (gw#colonialone, My Work PC), PINNED BOOKMARKS (You have no pinned bookmarks), Bookmark Manager, Activity, Endpoints, Publish, Groups, Console, Account (huribur@globusid.org), Help, and Globus Home. The main area is titled 'File Manager' and contains input fields for 'Collection' (Start here...), 'Path', and 'Transfer or sync to...'. A context menu is open over a file, listing actions: Share, Transfer or Sync to..., New Folder, Rename, Delete Selected, Preview (limited), Download (https), Open (https), Get Link, Show Hidden Items, and Deactivate. At the bottom, there are buttons for 'Start', 'Transfer & Sync Options', and 'Start'.

HPC - File Transfer with Globus

The screenshot displays the Globus File Manager interface. On the left is a dark blue sidebar with the Globus logo and a menu of navigation options: File Manager, RECENTLY USED (gw#colonialone, My Work PC), PINNED BOOKMARKS (You have no pinned bookmarks, Bookmark Manager), Activity, Endpoints, Publish, Groups, Console, Account (hurlburj@globusid.org), Help, and Globus Home. The main content area is titled 'File Manager' and features a search bar labeled 'Collection' containing the text 'gw#colonialone'. Below the search bar, a dropdown menu shows the search results for 'gw#colonialone', including the email 'gw@globusid.org' and the endpoints 'login3 and login4. https://colonialone.gwu.edu/'. The top right of the interface includes a 'Panels' toggle and a 'Bookmark Manager' link.

HPC - File Transfer with Globus

The screenshot displays the Globus File Manager interface. On the left is a dark blue sidebar with the Globus logo and navigation options: File Manager, RECENTLY USED (gw#colonialone, My Work PC), PINNED BOOKMARKS (You have no pinned bookmarks, Bookmark Manager), Activity, Endpoints, Publish, and Groups. The main area is titled 'File Manager' and shows a search bar for 'Collection' with 'JBHWorkPC' entered. Below the search bar is a navigation bar with tabs: Recent, Bookmarks, Your Collections (highlighted with a dashed box), Shared With You, and More Options. The main content area lists several endpoints with search and share icons:

Endpoint Name	Search Icon	Share Icon
apple iMac GWU	🔍	🔑
Jason_GWU_laptop	🔍	🔑
Jason TEST	🔍	🔑
Jason Test endpoint	🔍	🔑
JBHWorkPC	🔍	🔑
kraken03 at SEAS	🔍	🔑

HPC - File Transfer with Globus

The screenshot displays the Globus File Manager interface. On the left is a dark sidebar with navigation options: File Manager, RECENTLY USED (gw#colonialone, My Work PC), PINNED BOOKMARKS (You have no pinned bookmarks), Bookmark Manager, Activity, Endpoints, Publish, Groups, Console, Account (hurlburj@globusid.org), Help, and Globus Home.

The main area is titled "File Manager" and shows two panels. The left panel is for the collection "gw#colonialone" with the path "/~/". The right panel is for the collection "JBHWorkPC" with the path "/~/". Both panels have a toolbar with "select all", "up one folder", "refresh list", and "columns" options.

The left panel displays a list of files and folders:

Item	Timestamp
__pycache__	3/22/2017 9:59am
a.out	8/27/2015 8:46am
admin_stuff	3/27/2017 6:16pm
anacondainstall	3/30/2017 6:02pm
anotherstg	3/8/2018 3:17pm
bak_anaconda_4.2	3/30/2017 2:15pm
CDS.log	12/1/2017 3:40pm
centos7_clean.img	3/2/2018 10:08am
class_stuff	3/27/2017 6:24pm
CLOCK	11/3/2016 10:28am
composite	7/14/2017 12:32pm
cuda-8.0-linux-x64-v5.1.tgz	4/25/2017 11:04am
data	4/19/2016 3:13pm
Desktop	2/3/2017 2:49pm
desktop.ini	3/2/2017 7:19pm

The right panel displays a list of files and folders:

Item	Timestamp	Size
Actual_ODBC_Pack.dmg	10/25/2018 10:0...	50.56 MB
Applications	7/26/2018 2:16pm	160 B
balarasgrp_lustre_purge.dat	5/12/2017 2:28pm	15.62 MB
bdlabs	7/25/2018 1:02pm	192 B
bdlabs.tar.gz	7/24/2018 10:19...	349.09 MB
biofluid-lab	7/25/2018 1:03pm	480 B
biofluid-lab.zip	7/25/2018 12:57...	8.86 MB
Box	11/30/2018 11:47...	4.09 KB
Box Sync	10/12/2018 10:31...	1.40 KB
clang+llvm-4.0.1-x86_64-linux-gnu-...	8/10/2017 2:00pm	222.50 MB
composite_seas_gwu_edu_cert.cer	7/18/2018 1:13pm	2.49 KB
composite_seas_gwu_edu_interm.cer	7/18/2018 1:14pm	5.60 KB
config.tar.gz	8/9/2017 1:28pm	6.35 KB
Creative Cloud Files	11/30/2018 11:27...	96 B
Desktop	11/30/2018 12:0...	704 B

At the bottom, there is a "Start" button with a play icon, a "Transfer & Sync Options" dropdown menu, and another "Start" button with a play icon.



HPC - Lustre and NFS

Lustre:

- Lustre is a free and open standard for creating a parallel high-speed file system
- It works by “striping” data over several different storage volumes.
- Lustre is a high speed storage system
- Lustre should be used for running jobs
- Is purged monthly
- Hosts /lustre/groups

NFS:

- Network File System
- Hosts /home and /group directories
- NFS is slow compared to lustre
- Is not purged
- Not for job I/O



HPC - Lustre

How to use Lustre:

- Using Lustre is one of the simplest things you can do on Colonial One or any cluster. You simply need to read or write to a lustre directory. Nothing else is required!

On Colonial One the lustre file system is found here:

- `cd /lustre/groups`
- Find your group:
 - *Type: groups*
 - *The first group listed is your primary group*
 - *Your group directories are located here:*
 - `/groups/<primarygroup>`
 - `/lustre/groups/<primarygroup>`



HPC - Job Scheduler

Colonial One uses SLURM to schedule and prioritize jobs on the cluster.

SLURM (Simple Linux Utility for Resource Management) is a software package for submitting, scheduling, and monitoring jobs on large compute clusters.



HPC - First Submit Script

Copy a test submit script into your home directory:

```
cp /c1/workshops/workshop1/first_submit_script.sh ~
```

CD to your home directory and edit the file to add your email address and home directory:

```
cd ~  
nano first_submit_script.sh
```

Submit your script:

```
sbatch first_submit_script.sh
```

The background of the slide features a blue-tinted image of server racks on the left and a close-up of a computer keyboard on the right. The text is overlaid on a dark blue horizontal band.

For More Information

Colonial One overview:

<http://it.gwu.edu/colonialone-high-performance-computing>

User documentation:

<http://colonialone.gwu.edu>

Or send us email:

Colonial One support - hpchelp@gwu.edu