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
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.
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.
HPC - Exercise 1. Logging In

Requirements:

- SSH Client
- Colonial One account

Log into Colonial One:

```
ssh username@login.colonialone.gwu.edu
```

Use your NetID and password!
Colonial One
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 fileserver for /home and /groups
- 262 TB Lustre filesystem for high-speed scratch
- 268 TB Dell Compellent for archival
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

- **Login Nodes**
  - Login3
  - Login4

- **Scheduler**
  - SLURM

- **Compute Nodes**
Cluster Architecture

DO NOT RUN JOBS HERE!

Login3

SLURM

Login4

RUN JOBS HERE!

Compute Nodes
Cluster Architecture (Storage)

Login Nodes

Scheduler

SLURM

Compute Nodes

/home and /group

/lustre
Cluster Architecture (Storage)

Login Nodes

Scheduler

SLURM

Compute Nodes

/Login and /group

Running Job Data goes here!

/lustre
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**
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 ~`
Questions & Discussion
Modules load an environment so a program can run correctly.

Module commands:
- module list
- module avail
- module load
- module unload
- module spider
1. From the command prompt, type: \texttt{R}
2. The command is not found!
3. Type: \texttt{module load R/3.4.2}
4. Type: \texttt{R}
5. Notice R loads!
6. Type: \texttt{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`
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)
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.
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
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.
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
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!
Open a web browser and go to: www.globus.org
Search the list for The George Washington University
Click Continue
HPC - File Transfer with Globus

Sign in with your NetID and password
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

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.

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

Click in the Collection field above to begin

Watch a two minute tour of what’s new
HPC - File Transfer with Globus

File Manager

Collection: gw#colonialone

gw#colonialone

gw@globusid.org

Endpoints to login3 and login4. https://colonialone.gwu.edu/
HPC - File Transfer with Globus
**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>
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.
Copy a test submit script into your home directory:

```
cp /cl/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
```
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