

BIO599 - Monsoon Intro

Presented By: Joseph Guzman, Jason Buechler
Date: 2023-09-18

Workshop Roadmap

1. Introduction
2. Managing files
3. Software packages

This slide deck is located here:

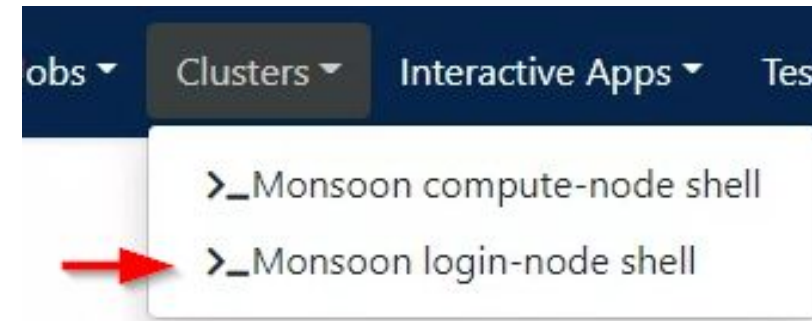
<https://rcdata.nau.edu/hpcpub/workshops/bio599-intro.pdf>

See this link for the workshop schedule: <https://in.nau.edu/arc/workshops>

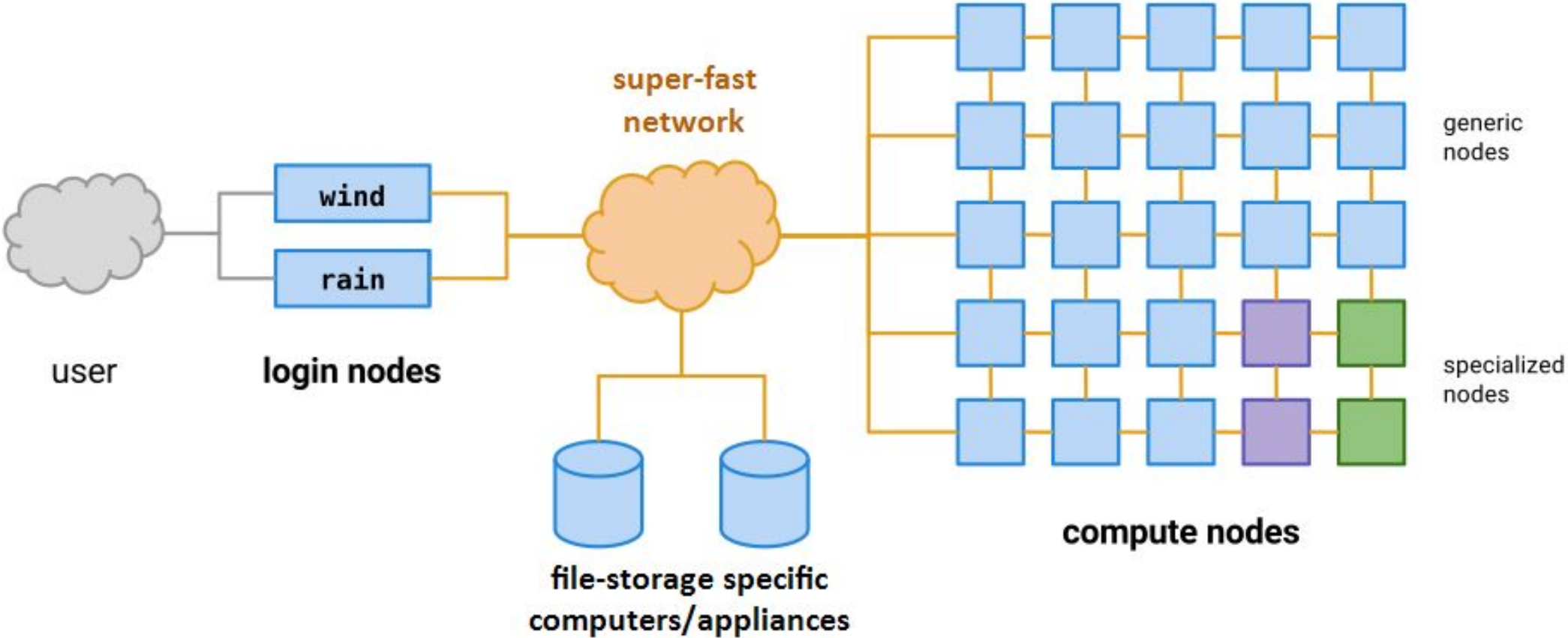
Section 1: Introduction

Get logged in!

- **Slides here:**
 - <https://rcdata.nau.edu/hpcpub/workshops/odintro.pdf>
- **From a Computer:**
 - Log into NAU VPN *if off-campus!*
 - Instructions here: <https://in.nau.edu/its/remote-services/>
 - VPN requires Two Factor Authentication
 - https://nau.service-now.com/kb_view.do?sysparm_article=KB0013321
 - Open a web browser
 - May need to search in start menu for it
 - Browse to <http://ondemand.hpc.nau.edu>
 - [Log in with your louie id](#)
 - Click on Clusters tab, and select Monsoon login-node shell



Cluster Topology



Section 2: Managing files

Storage

- **/home – 10GB quota**
 - Keep your scripts and executables here
 - Snapshotted twice a day: /home/.snapshot
 - Please do not write job output (logs, results) here!!
 - Run the command “quota” now
- **/scratch – 1PB total space, 30 day retention**
 - Very fast storage, capable of 20GB/sec
 - Quota: 15TB, 2M files
 - Checkpoints, logs
 - Keep all temp/intermediate data here
 - Should be your default location to perform input/output

Storage

- **/projects – 615TB**
 - Long-term storage project shares
 - 5TB is assigned to faculty member for group to share
 - \$24/TB/year above 5TB
 - Snapshots available
 - Backups available - \$.10/GB/month
- **/common**
 - Cluster support share
 - Contrib: place to put software/libs/confs/db's for others use

How do I move data to or from Monsoon?

Service	Tools	To Monsoon	From Monsoon
SSH	rclone, scp, rsync	YES	YES
SMB (a.k.a. samba, CIFS, “network drive”)	File Explorer, Other desktop client	YES	YES
FTP/FTPS	rclone, ftp	YES	NO
HTTP/HTTPS	browser, rclone, curl, etc	NO	YES (Data Portal)
Cloud (i.e. google drive)	browser, rclone, API, globus (NAU gdrive only)	YES	YES
Globus	browser (asynchronous)	YES	YES

- Other?
 - You can use Monsoon as a client for other services. But above are the service types that we support. Monsoon acts as a server for all of the services listed, except for “Cloud” and FTP/FTPS.

Pitfalls when moving or copying files

- Permission denied
 - read - needed for copying a file
 - execute - needed for traversing a directory
- File clobbering -- oops! 😞
 - snapshots
- Recursive copy
- File globbing
- Moving vs renaming a file
- Trailing slash

Section 3: Software Packages

Lmod Tutorial

1. Check what Lmod modules you have loaded:

```
module list
```

2. Unload a module:

```
module unload <pkg>
```

3. Unload all modules:

```
module purge
```

- OR login again

4. List all modules:

```
module av
```

5. Search for a module:

```
module av <pkg>
```

6. Load default module version:

```
module load <pkg>
```

7. Load specific module version (preferred):

```
module load <pkg>/<version>
```

- Why is it preferred to specify the version? Because we update Lmod packages, especially important for researchers who use conda

8. View module details

```
module display <pkg>/<version>
```

Conda/Mamba Tutorial

- Example:
 - `module load mambaforge`
 - `mamba create -n myenv`
 - `mamba activate myenv`
 - `mamba install -c conda-forge numpy`
- Now try out the web search ability
 - Visit: <https://anaconda.org/search>
 - Search for the “hexyl” package (it’s a hex viewer, similar use-case as the `xxd` and `od` commands)
 - Make a new conda environment and try out the installation command it gives you (replace “mamba” with “conda” to use the fast installer)
 - Try out the hexyl command:
 - i.e.
 - read your `.bashrc` file
 - `hexyl ~/.bashrc | less`
 - read random binary data
 - `hexyl <(head -c 1K /dev/urandom) | less`

Spack Tutorial

- Search for a package:
 - `spack list mypkg`
 - `spack info mypkg`
- Two workflows for installing packages:
 - direct installs
 - Unlike conda, spack allows for installing and loading packages without an environment. Spack is acutely aware of the dependencies specific to each package.
 - Switch to using environments if you have a complicated set of packages to avoid repetitive loads commands.
 - Example:
 - `module load spack`
`spack install gzip`
`spack load gzip`
 - install into an environment
 - Example:
 - `module load spack`
`spack env create -d ./myenv`
`spack env activate ./myenv`
`spack add gzip`
`spack install`

END

Need to contact us?

- Email:
 - ask-arc@nau.edu
- Zoom Office Hours
 - Every Wednesday
 - Alternates Between 1:00PM-2:00PM and 2:00PM-3:00PM
 - <https://in.nau.edu/arc/office-hours/>

Questions?

Packaging Tools Overview

Lmod

- Loadable Module system, includes packages that were created by us, the ARC team
 - Many of these were requested by researchers i.e. davinci
- view available packages by running: `module av`

Anaconda

- General-purpose package manager, with a focus on Python and R packages
- view available packages here: <https://anaconda.org/>

pip

- De-facto standard package manager for python packages
- view available packages here: <https://pypi.org/>

CRAN

- Repository of packages for R
- view packages here: <https://cran.r-project.org/index.html>

Spack (NEWLY REVAMPED)

- HPC-focused package manager, created by LLNL

configure+make

- Standard tools for source compilation in Linux

Singularity

- Docker-compatible containerization tool

.rpm files

- package recipe file for RHEL-based distros
- The equivalent in debian-based are .deb files

If you are considering one of these tools then you should contact us:

ask-arc@nau.edu

OpenSSH and Creating SSH Keys

- Highly recommended to use an OpenSSH Client
 - Will provide the OpenSSH implementation of ssh, scp, etc.
 - On Mac, Linux, and *BSD this should be installed by default
 - On Windows most people would recommend PuTTY, but you may also be able to get the Unix-centric OpenSSH client on new versions of windows with this command (see the MS docs [here](#)):
 - `Add-WindowsCapability -Online -Name OpenSSH.Client~~~~0.0.1.0`
- To enhance security and convenience you can create an SSH key on your local machine.
 - `ssh-keygen -t ed25519`
- Then to use this key to log into monsoon you'd use this command:
 - `ssh-copy-id abc123@monsoon.hpc.nau.edu`
 - Now if you were to try to ssh into monsoon normally it will default to trying your key first, instead of prompting you.
 - `ssh abc123@monsoon.hpc.nau.edu`

SSH-Based Tools and SMB

SSH

- scp
 - most simple, convenient for small downloads
- rsync
 - used for “syncing” a source and destination folder
 - performs better when you need to do partial downloads
- rclone
 - generalist tool, has support for many different service types
 - created primarily as a one-stop shop for interacting with cloud storage services
 - requires light configuration before using

SMB:

- Ideal for desktop usage
- Mount a network drive on your file explorer app
- Windows: \\shares.hpc.nau.edu\cirrus
- Mac: smb://shares.hpc.nau.edu/cirrus

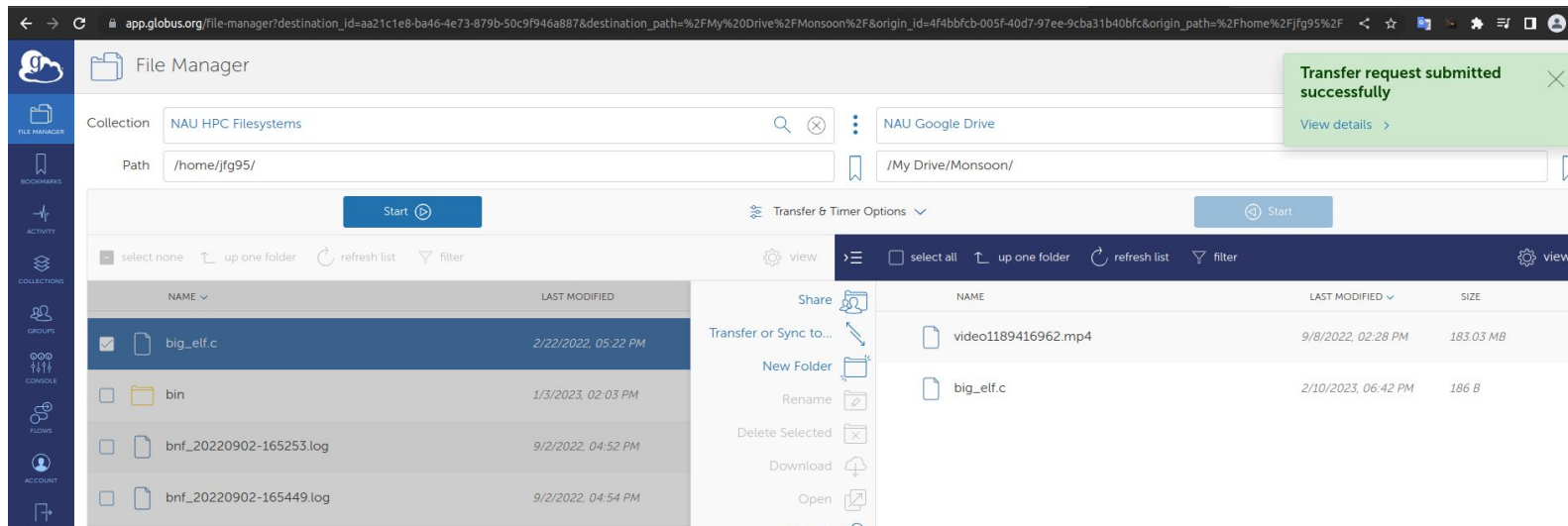
Globus (Part 1)

- Specialized HPC webapp accessible at <https://app.globus.org>
 - Primarily used by research institutions
 - Use-case: moving large amounts of data across institutions
 - Need the globus personal app to move data to or from your desktop
 - see: <https://www.globus.org/globus-connect-personal>
- To start, visit the webapp and lookup “Northern Arizona University” at the prompt

- The first time it will ask you:
 - if you want to link your account
 - say no if you've never used globus before
 - to confirm your email
 - to accept terms and conditions
 - to allow globus to use your id for file transfers
- Uses NAU SSO

Globus (Part 2)

- Globus calls access points “collections”
- Search for “NAU” under “collections” to access Monsoon via globus
 - The one named “NAU HPC Filesystems” has access to /home /scratch and /projects
 - The one named “NAU Google Drive” has access to your NAU-affiliated google drive
 - You can read/write from your desktop too if you have globus personal running



- In the browser app
 - search up a collection
 - right click on a file or directory
 - select “transfer or sync to...”
 - now search a second collection
 - select a destination folder for the second collection
 - click on start to move data to the second collection