IOWA

High Performance Computing Systems

PSQF 7375 — Research in Psychometrics

March 29, 2021

Overview

- What is a Compute Cluster?
- High Performance Computing
- High Throughput Computing
- The Argon Cluster
- Mapping Your Problem to a Cluster

- Note: These slides were adopted from Ben Rogers' HPC Workshop slides

Introductions & Resources

- Resources
  - Wiki Documentation - https://wiki.uiowa.edu/display/hpcdocs/Argon+Cluster
  - Cluster help: research-computing@uiowa.edu

What is a Compute Cluster?

- Large number of computers
- Software that allows them to work together
- A tool for solving computational problems that require more memory or CPU than is available on a single system

High Performance Computing

- Using multiple computers in a coordinated way to solve a single problem
- Provides the ability to:
  - Use 10s-1000s of cores to solve a single problem
  - Allows access to 10s-1000s of GB of Ram
- Likely to require substantial code modification to use a library such as MPI
- Common Examples:
  - Computational Fluid Dynamics
  - Molecular Dynamics
  - Um…Psychometrics

High Throughput Computing

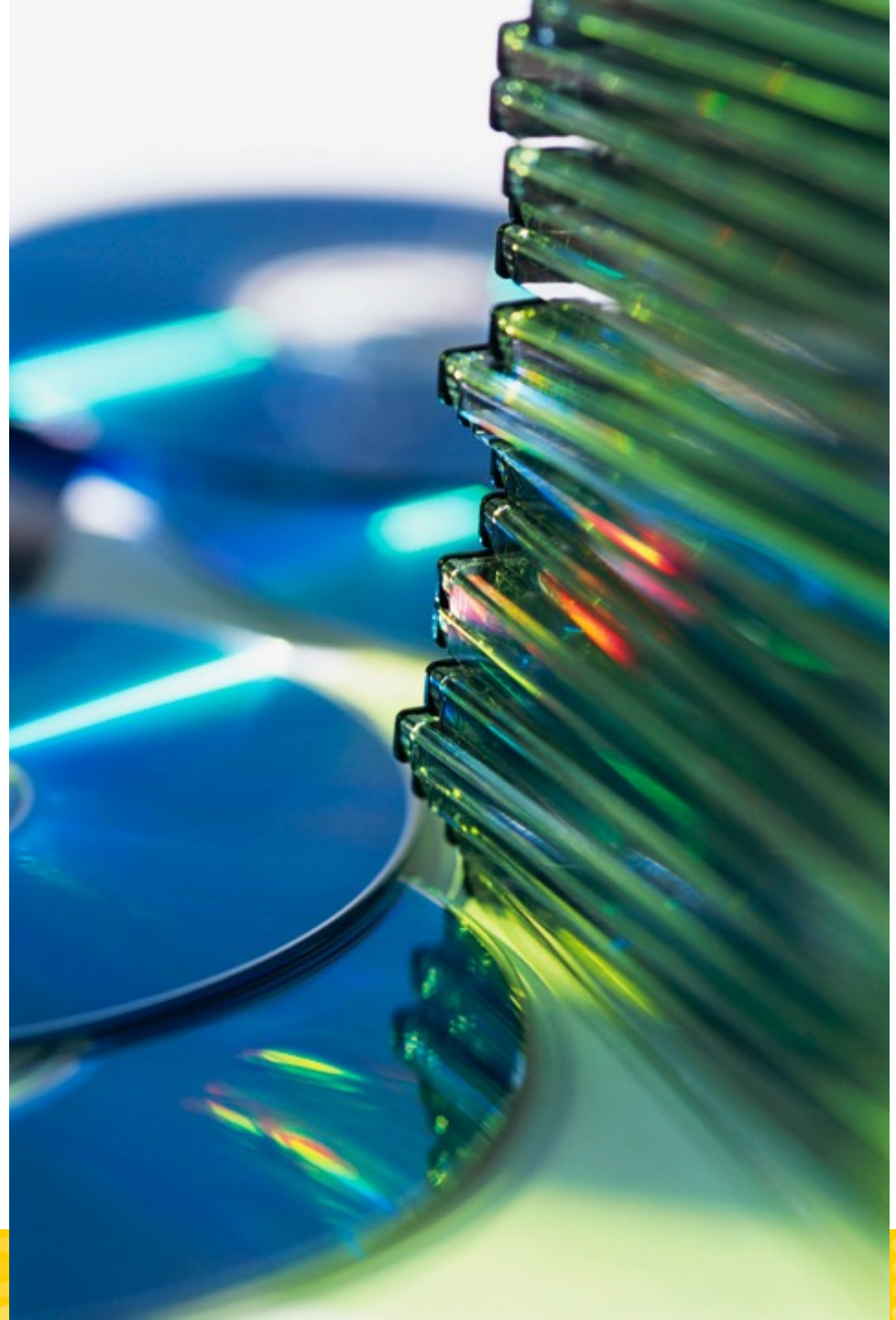
- Using multiple computers in a coordinated way to solve many individual problems
- Provides the ability to:
  - Analyze many data sets simultaneously
  - Efficiently perform a parameter sweep
- Requires minimal code modifications
- Common Examples:
  - Image Analysis
  - Genomics

The Argon Cluster

- Collaborative Cluster
- CentOS 7 (Linux)
- As of 2019:
  - 600+ compute nodes
  - ~14,000 processor cores
  - ~28,000 threads (slots)
  - 64-512GB of Ram/node
  - ~120 GPU Accelerators
  - 100Gb Omnipath and 40Gb Infiniband Network

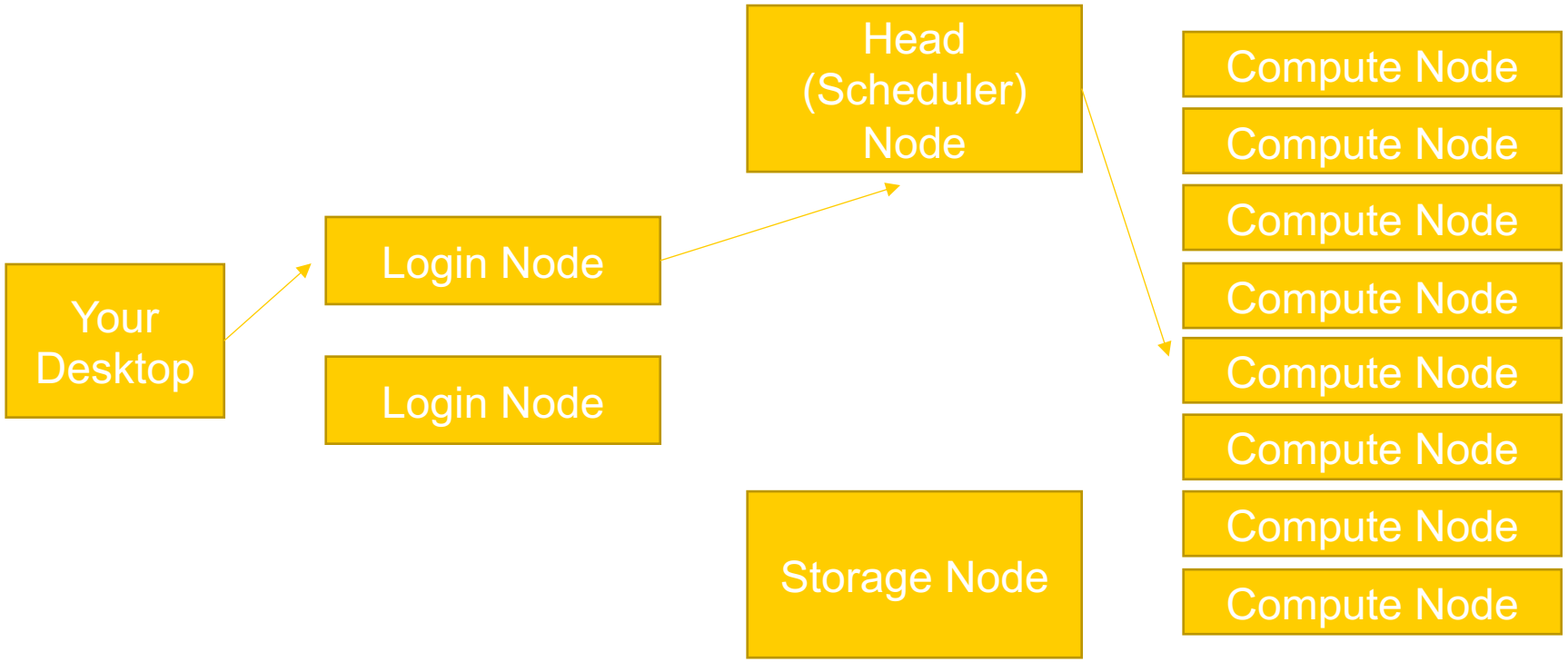
Argon Storage

- Home Account Storage
  - NFS
  - 1TB per User
  - /home/$hawkid
- Shared Scratch Storage
  - NFS: 330TB - /nfsscratch
  - Deleted after 60 days
- Local SSD or HD Scratch
  - 900GB+/Compute Node
  - /localscratch
- No Backups!



IOWA

Cluster Structure Overview



Storage Continued

- Paid Storage Service (Large Scale Storage or LSS)
  - $40/TB/Year/Copy of Data
  - $80/TB/Year with Backups
  - Available outside cluster
  - Found at /Shared/$LAB & /Dedicated/$LAB

Transferring Data to Argon

- If off campus, must connect via VPN first
- CIFS (Windows) Network File Share
  - Available for home accounts & paid storage
  - https://wiki.uiowa.edu/display/hpcdocs/Home+Accounts
  - Argon Home Server Information
    - \\argon-home.hpc.uiowa.edu\myHawkID
- sftp (secure file transfer protocol)
  - Available for all file shares
  - https://wiki.uiowa.edu/display/hpcdocs/Using+SCP+or+SFTP
- Also, smb://data.hpc.uiowa.edu/argon_home
  - On Mac, go to Finder…Go…Connect to Server and paste link above
- New to sftp? https://filezilla-project.org/download.php?type=client#close
  - Follow along in class

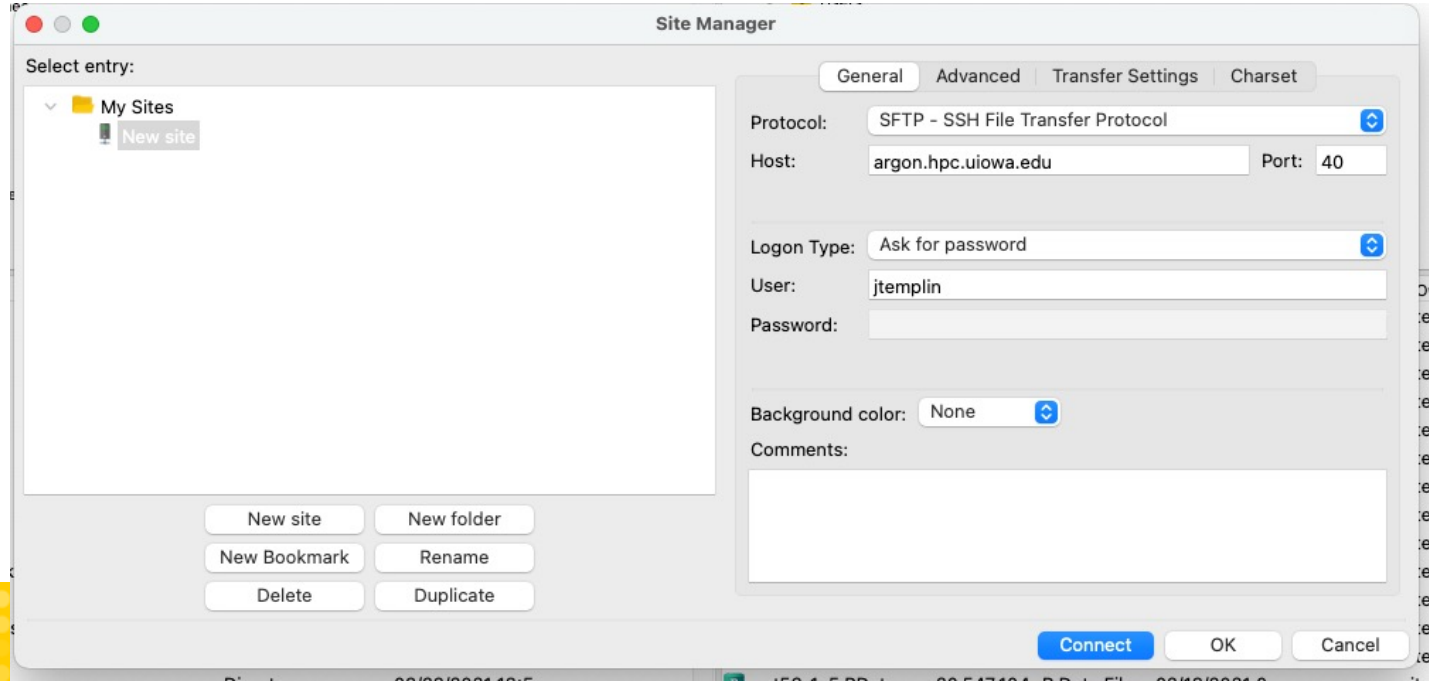
Transferring Data to Argon: Command Line

- To demonstrate, we will transfer the example from today from the notebooks site to the argon site
- Go to https://notebooks.hpc.uiowa.edu
- Start our course RStudio server
- Go to the terminal window (by console) or open a new terminal in View…Move Focus to Terminal
- Use SCP to move the file over (replace HawkID with yours):
  - scp -p 40 ~/classdata/HPC/ArgonDemo.tar.gz HawkID@argon.hpc.uiowa.edu:ArgonDemo.tar.gz
- Enter password
- Use Duo to confirm
- Done!

IOWA

Transferring Data to Argon: FileZilla and sftp

- FileZilla is an open source file transfer app:https://filezilla-project.org
- Protocol: SFTP
- Host: argon.hpc.uiowa.edu
- Port: 40
- User: HawkID
- Once connected you can manually move files



So It Just Runs Faster, Right?

- Not quite!
- Just running on Argon won't necessarily make your program faster

http://basementgeographer.blogspot.com/2012/03/international-racing-colours.html

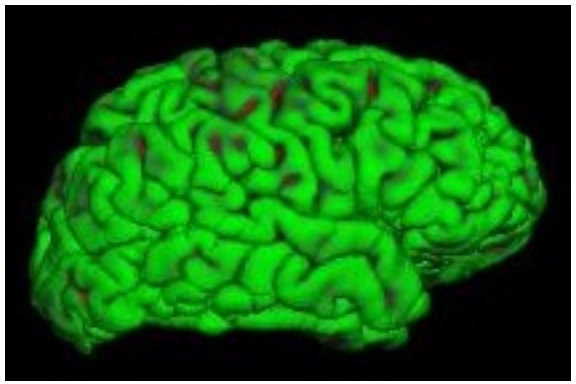
IOWA

Mapping Your Problem to a Cluster

- Questions
  - Does your problem require more resources (CPU, GPU, memory) than available on your desktop system?
  - Does your job run on Linux?
  - Can your job run in batch mode?
  - Is your job HPC or HTC?

- Next Steps
  - Develop Strategy for Running Jobs
  - Install Software
    - Home Account Install
    - Singularity
    - Central Install (Existing via modules)
  - Develop Job Submission Scripts
  - Run Your Job

The Challenge: Analyze 1000 MRIs



- Run Freesurfer on 1000 MRIs
  - Takes 20 Hours per MRI
  - Requires 2GB of Memory/analysis
- Desktop Analysis Time
  - 20 Hours x 1000 MRIs = 20,000 Hours
    - 2.3 Years!
  - But I have a Quad Core Desktop with 8GB
    - That's still over six months!

Analyze 1000 MRIs: Using the Argon Cluster

- Good fit for cluster? – Yes
- Type of problem – HTC
- Software – Runs on Linux in batch mode
- Time to Analyze
- On Argon – As little as 20 hours
  - Time dependent on cores available, likely complete within a week.
  - Possible to run all analyses simultaneously
    - 1000 processor cores – Total on Argon > 14,000
    - 2TB of memory – Total on Argon > 100TB

Getting the Resources You Need

- Argon uses the Son of Grid Engine (SGE) scheduler to allocate resources.
- To request specific compute resources you need to know about:
  - Queues
  - Slots
  - Resources

Queues

- Investor queues
  - Investor owned, guaranteed access
  - Access restricted to specified users
- UI queue
  - Treated like an investor queue but everyone has access
- all.q
  - No job limits
  - Subject to eviction

Centrally funded queues	Node Description	Wall clock limit	Running jobs per user
UI	(20) 56-core 256GB (66) 32-core 64GB	None	5
UI-HM	(5) 56-core 512GB (3) 24-core 512GB	None	1
UI-MPI **(56 slot minimum)**	(19) 56-core 256GB	48 hours	1
UI-GPU	(10) 32-core 64GB with (1) K20 accelerator (5) 56-core 256GB with P100 accelerator (2) 40-core 192GB with (4) 1080Ti accelerators (4) 40-core 192GB with (4) Titan V accelerators (1) 40-core 192GB with (2) Titan V accelerators	None	1
UI-DEVELOP	(1) 56-core 256GB (1) 56-core 256GB with P100 accelerator	24 hours	1

IOWA

Slots

- Slots are equivalent to a share of a machines resources (1 slot on a 56 core machine = 1/56th of the system resources)
  - Processor Cores + Hyperthreads
  - Memory
- The number of slots you specify can determine how soon your job will be able to run
- When using mpi (HPC jobs) we recommend using full machines (eg. Request a number of slots evenly divisible by the number of slots on a machine.)

Argon is Heterogenous Hardware

- This is beneficial because you get access to more hardware from a single system
- Downsides
  - More resource requests may be needed for your job
  - Greater performance variance due to different ages of hardware (up to seven years in our system)
  - Some applications do not function on older architectures.
    - In most cases this would be because the applications were compiled on a newer process architecture with hardware optimizations that don't exist on older hardware

Resources

- There are many resources that can be requested.
  - https://wiki.uiowa.edu/display/hpcdocs/Advanced+Job+Submission

- Most common to be aware of are:
  - Memory – Example: -l mem_256G=true
  - GPU – Example: -l gpu=true
  - # of GPUs – Example: -l ngpus=2
  - CPU Architecture – -l cpu_arch=broadwell

Hands On - Logging In

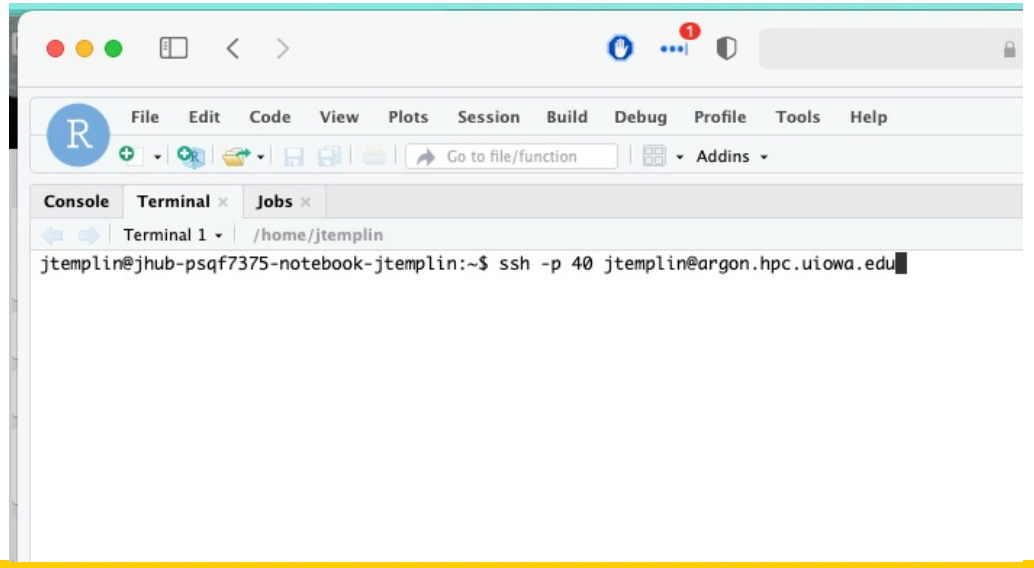
- Access via terminal window and ssh
- Getting logged in to Argon
  - Need to use your HawkID (Iowa domain) password
  - Windows will use Putty
  - Mac will use ssh from Terminal
  - Make sure you do *not* check “save password”
  - You will be prompted to accept a key.
    - Say yes to accepting the key
  - ssh –p 40 argon.hpc.uiowa.edu

Log into Argon: Windows

- No default ssh client on windows, so two options:
  - Download open source/free client putty: https://www.microsoft.com/en-us/p/putty-unofficial/9n8pdn6ks0f8?activetab=pivot:overviewtab
  - Use RStudio via Notebooks: https://notebooks.hpc.uiowa.edu

- For RStudio:
  - Go to Terminal
  - Connect with command:

  ssh –p 40 HawkID@argon.hpc.uiowa.edu
  - Enter your university password
  - Confirm with Duo

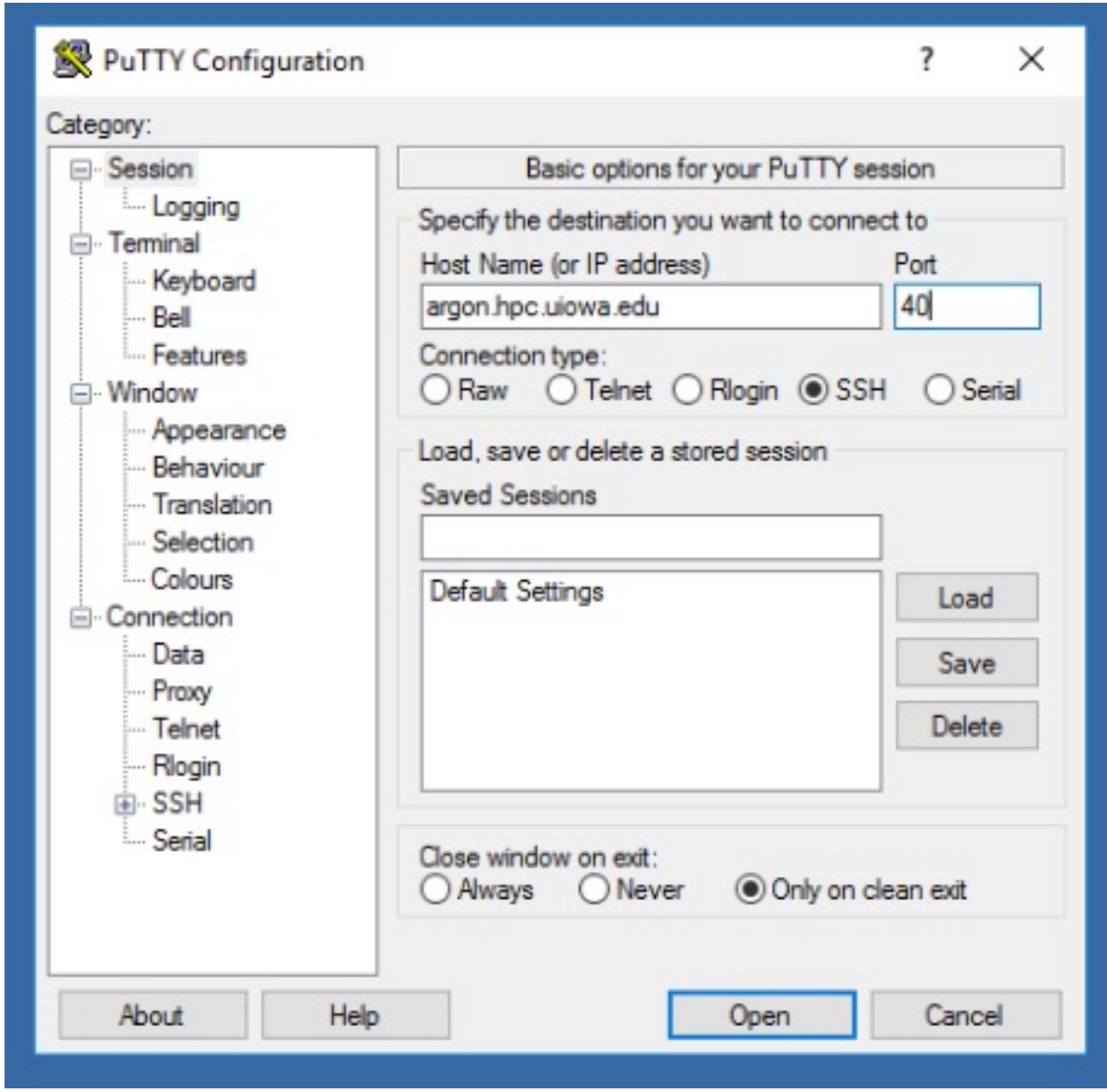


IOWA

Log into Argon: Windows with PuTTy

- For PuTTy:
  - Open program
  - Host Name: argon.hpc.uiowa.edu
  - Port: 40
  - Connection type: SSH

- Press Open



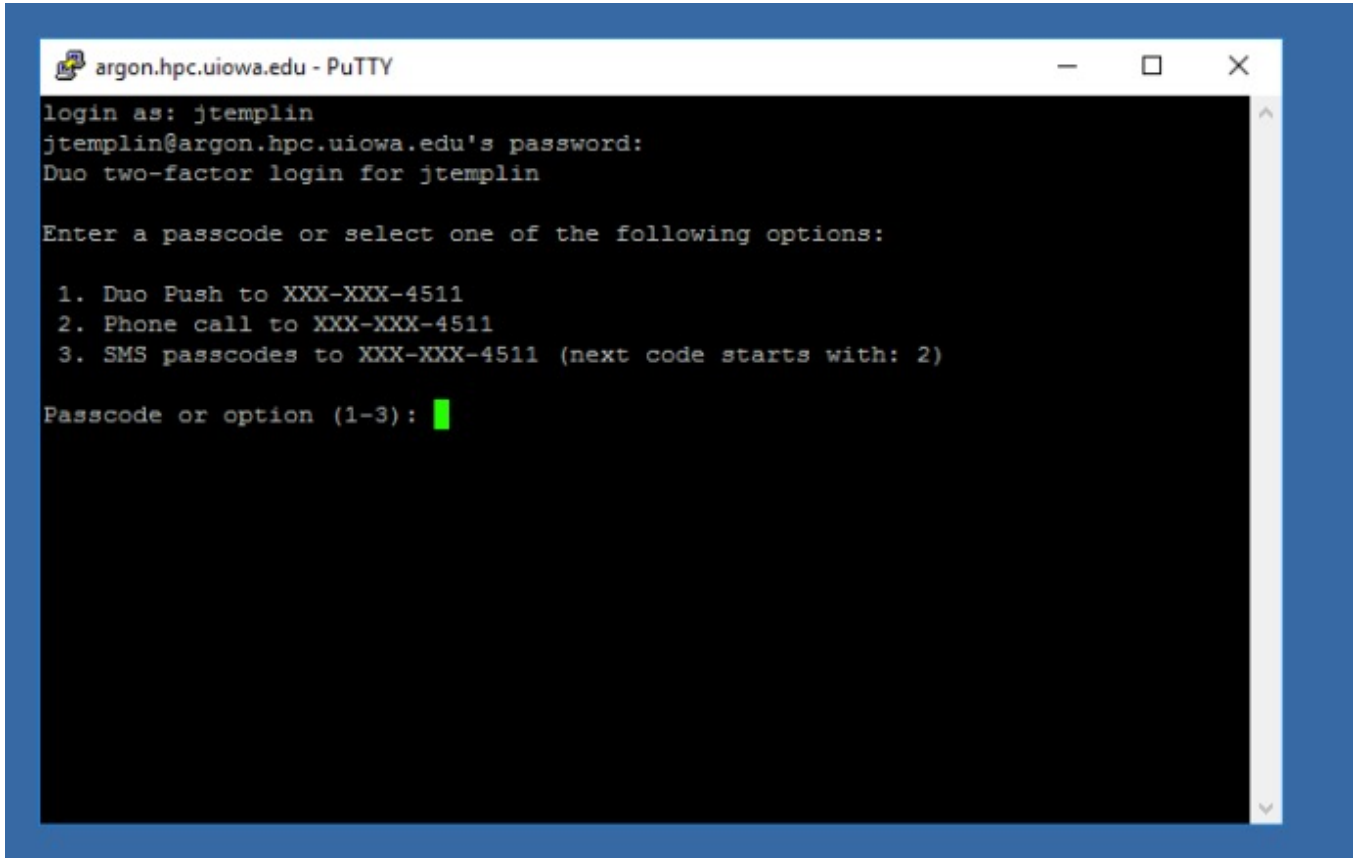
Log into Argon: Mac

- Press Command-Space to bring up Spotlight search
- Type Terminal
- Open Terminal program
  - Connect with command:

  ssh –p 40 HawkID@argon.hpc.uiowa.edu
  - Enter your university password
  - Confirm with Duo

Windows with PuTTY

- Next screen: terminal window
  - Login as: HawkID
  - Password: UI password
  - Confirm with Duo
- Once in, all commands are linux based



IOWA

I’m logged in; now what?

- I’m connected; how do I start working?
  - Argon is a batch, queued system
  - Jobs must be submitted to a queue, wait their turn; then they are processed
    - When running in a batch system you do *not* interact with your program in *real time*.
    - Have to specify options in your job script

Job Scripts

- Similar to regular shell scripts
- First line must contain path to a valid shell
  - Ex. #!/bin/bash
- Comments start #
- Directives start #$
  - Specify options to Grid Engine
  - These are *not* comments
- Explanation of sleeper script in editor

Submit a job

- Copy job file from

  /RS/hpc_training/sleeper.sh

  cp /RS/hpc_training/sleeper.sh .
- Edit sleeper.sh
  - Vi or emacs if you have a favorite editor
  - If new, try nano
- Change to your email address
- Launch job
  - qsub sleeper.sh
    - Your job 5588776 ("sleeper") has been submitted

What's my job doing?

- Check the status of your job
  - qstat –u [username]
  - Shows all *your* jobs
- Check for output files
  - [Scriptname].o[jobnumber]
  - Example: Sleeper.o3101075

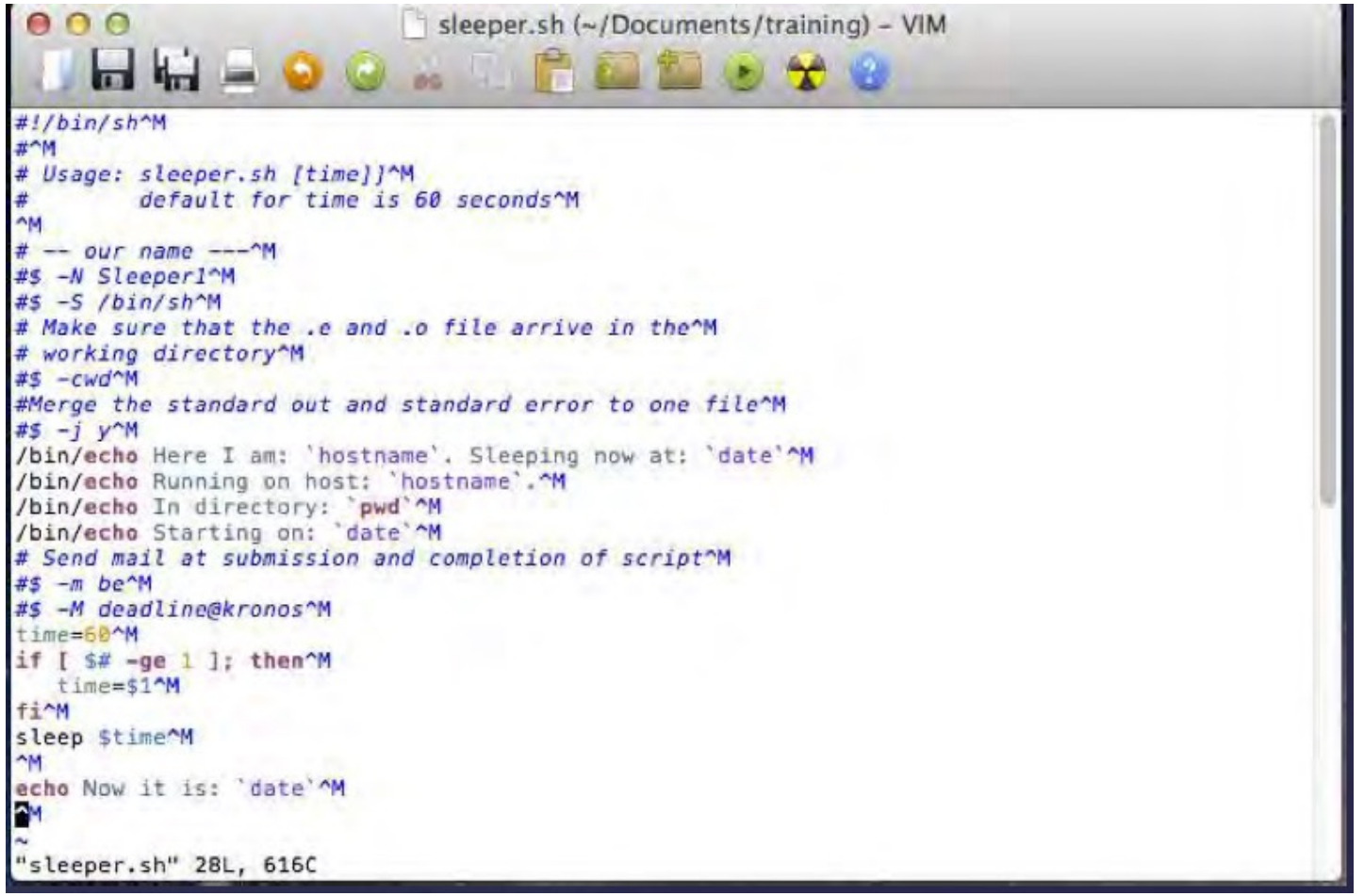
Modules

- What are environment modules?
  - System for changing your shell environment
  - Can be loaded and unloaded
  - Find modules via “module avail”
  - Load via “module load”
  - Unload via “module unload”
- Try to run R
- Load the R module
  - “module load R”
- Now try to running R again
- If a program is not working, you may need to load a module

Common issue

- Newline characters
- “I’m getting weird errors and I don’t see anything wrong with my file.”

What are those ^M doing there?



IOWA

Options for Windows Users

- Use a windows editor that has a UNIX line ending setting such as notepad++ (free software)
- Edit your text files exclusively on Argon
- Use dos2unix on Argon to convert the files

Additional Help

- Individual Consulting
  - By appointment
- Contact research computing:
  - research-computing@uiowa.edu
- For additional details visit
  - http://www.hpc.uiowa.edu

Up Next: HPC Demo

- Next, we will follow the demonstration on my website
- https://jonathantemplin.com/introduction-to-the-university-of-iowa-high-performance-computing-system-argon-and-iowa-interactive-data-analytics-service-idas/

IOWA