New Kid on the Block: Getting Started with Alpine

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

New Kid on the Block: Getting started with Alpine

- Website: www.rc.colorado.edu
- Helpdesk: rc-help@colorado.edu
- Slides: https://github.com/ResearchComputing/New_User_Seminar

- Survey: http://tinyurl.com/curc-survey18

Presentation prepared by Trevor Hall, CURC

Learning Goals

1. CU Research Computing (CURC) and Alpine cluster basics
2. Using CURC and Alpine resources
3. Navigating Research Computing
4. Alpine Software
5. Moving your data
6. Running a job (time allowing)
7. Help!

Things to take note of:

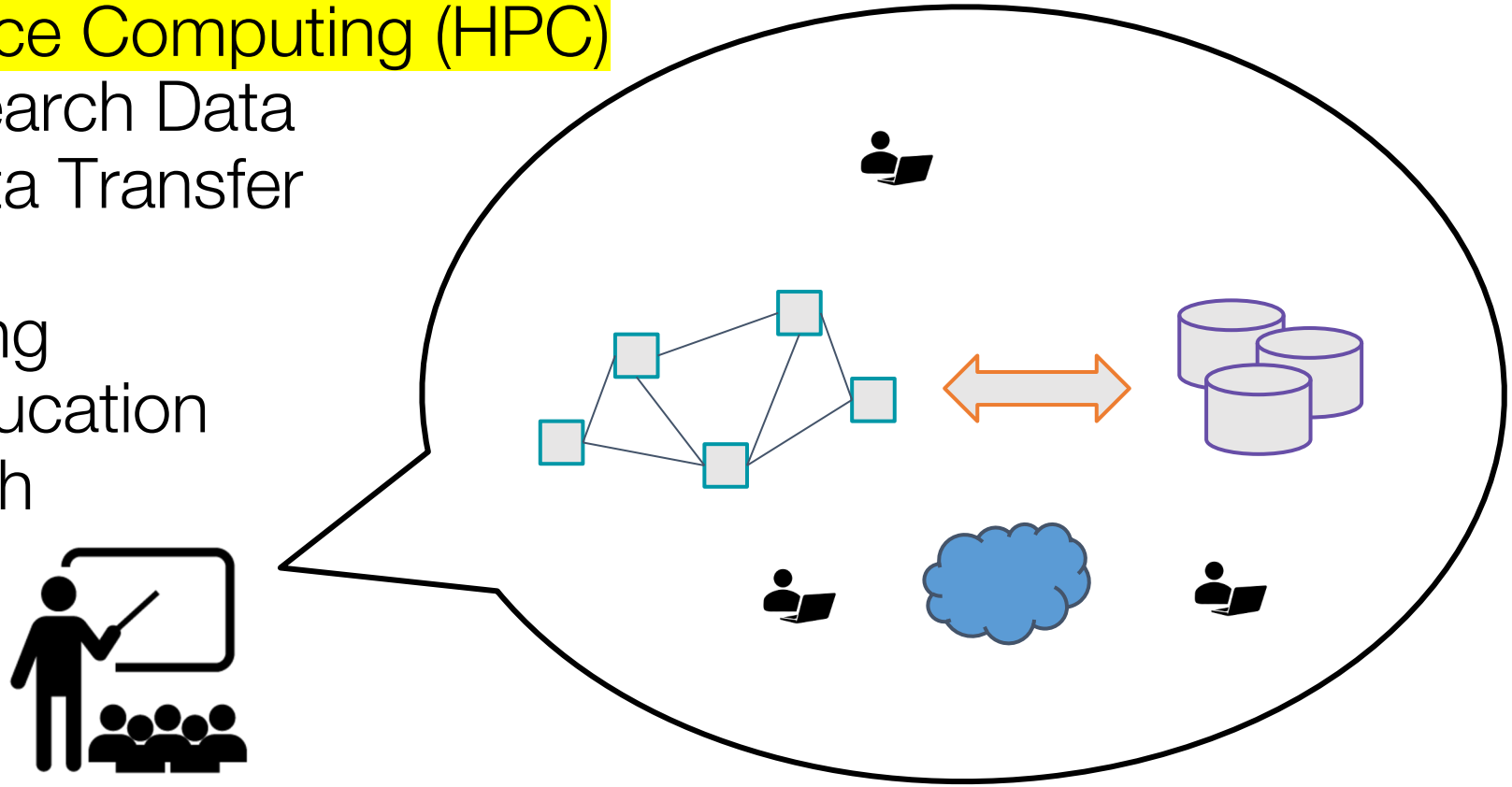
- Confusing, ambiguous, highly nuanced concepts
- Our goal is to help you avoid common mistakes, pitfalls, and frustrations

Ask Questions!

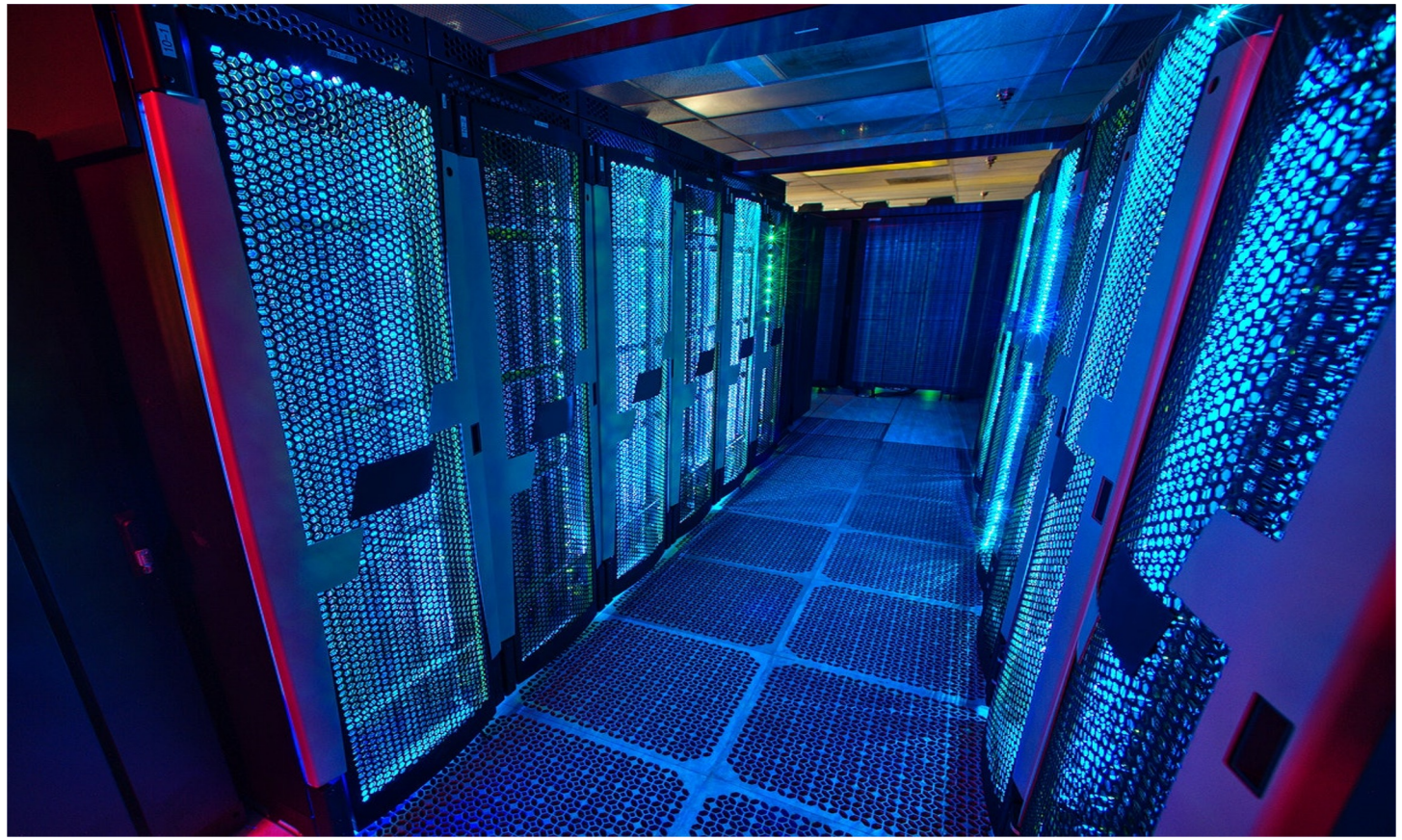
CU Research Computing and Alpine basics

CURC Resources Include:

- High Performance Computing (HPC)
- Storage of Research Data
- High-Speed Data Transfer
- Data Sharing
- Cloud Computing
- Training and Education
- Secure Research

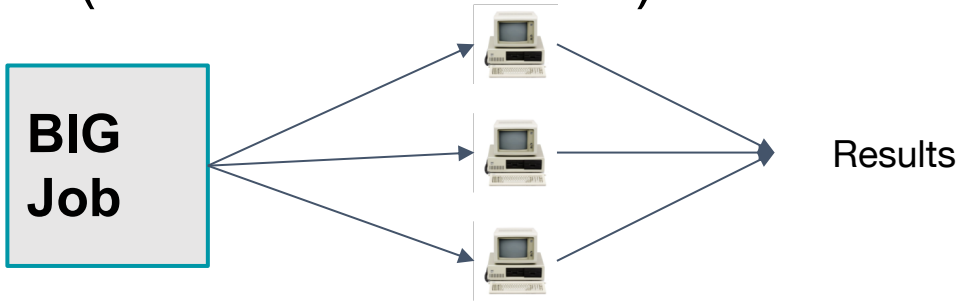


CURC is primarily known for: High Performance Computing (HPC)

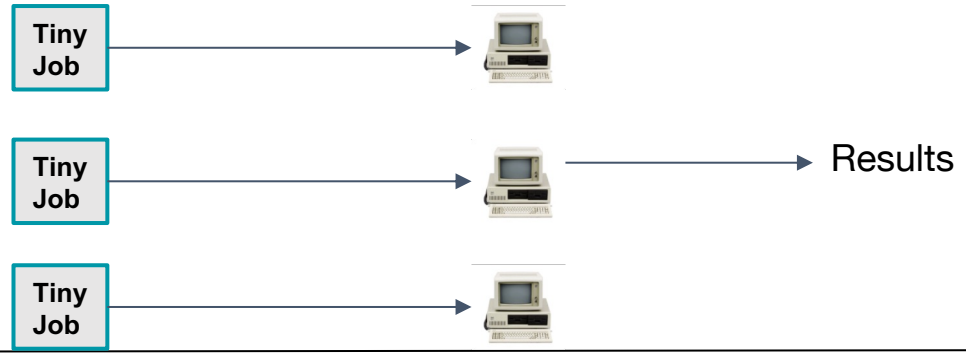


What can I use HPC for?

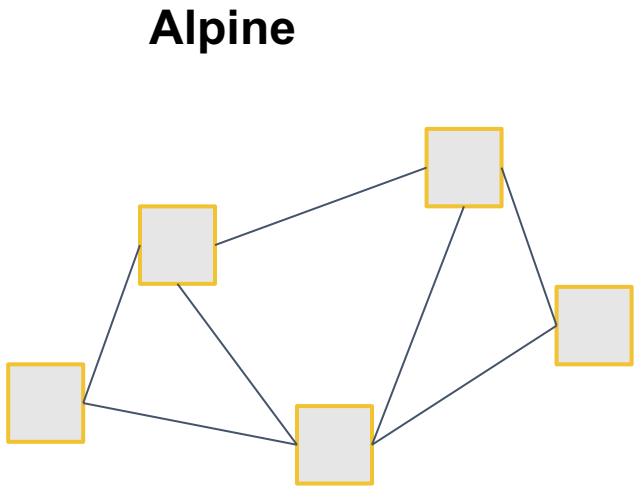
- Jobs that would take a long time on local machines can instead be distributed over hardware:
  - Parallelized to split up then joined (if software enabled)



  - Broken up into many serial jobs



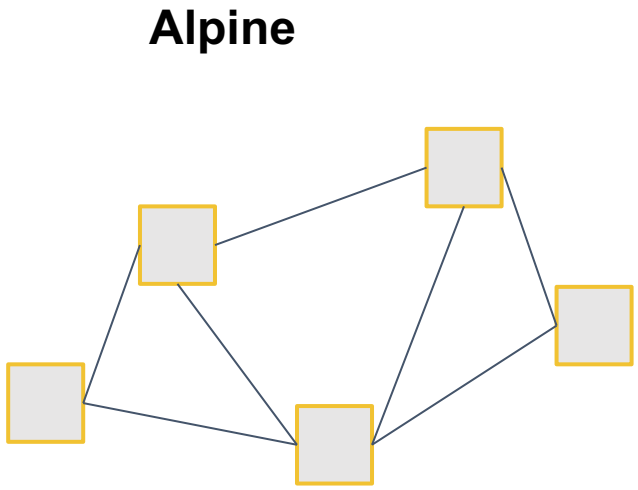
CURC HPC Cluster: Alpine



- Alpine is the 3rd-generation HPC cluster at CURC, following:
  - Janus
  - RMACC Summit

- Alpine is a heterogeneous cluster with hardware currently provided by CU Boulder, CSU, and Anschutz
- Access available to CU Boulder, CSU, AMC and RMACC users

HPC Cluster: Alpine

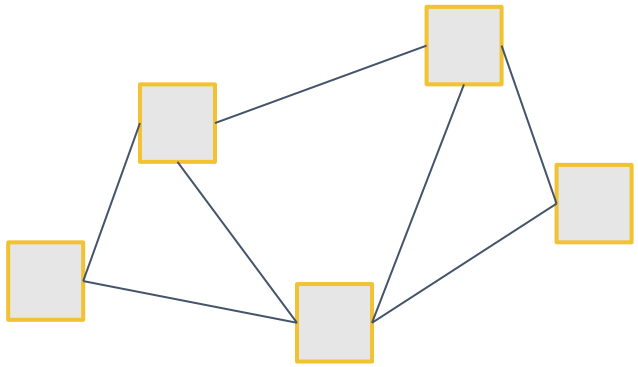


- Hardware on Alpine will continue to be purchased and released in stages:

- Alpine (stage 3):
  - 347 General CPU Nodes
    - *AMD Milan, 32,48,64 Core, 3.74G RAM/Core*
  - 12 NVIDIA GPU Nodes
    - *3x NVIDIA A100 (atop General CPU node)*
  - 8 AMD GPU Nodes
    - *3x AMD MI100 (atop General CPU node)*
  - 22 AMD High-Memory Nodes
    - *AMD Milan, 48,64 Core, 21.5G RAM/Core*
  - Additional Hardware contributed by CSU, AMC
    - *Nodes which boost priority for CSU/AMC users*

HPC Cluster: Alpine

Alpine



- Interconnect
  - **CPU nodes**: HDR-100 InfiniBand (200Gb inter-node fabric)
  - **GPU nodes**: 2x25 Gb Ethernet +RoCE
  - **Scratch Storage**: 25Gb Ethernet +RoCE
- Operating System
  - RedHat Enterprise Linux version 8 operating system

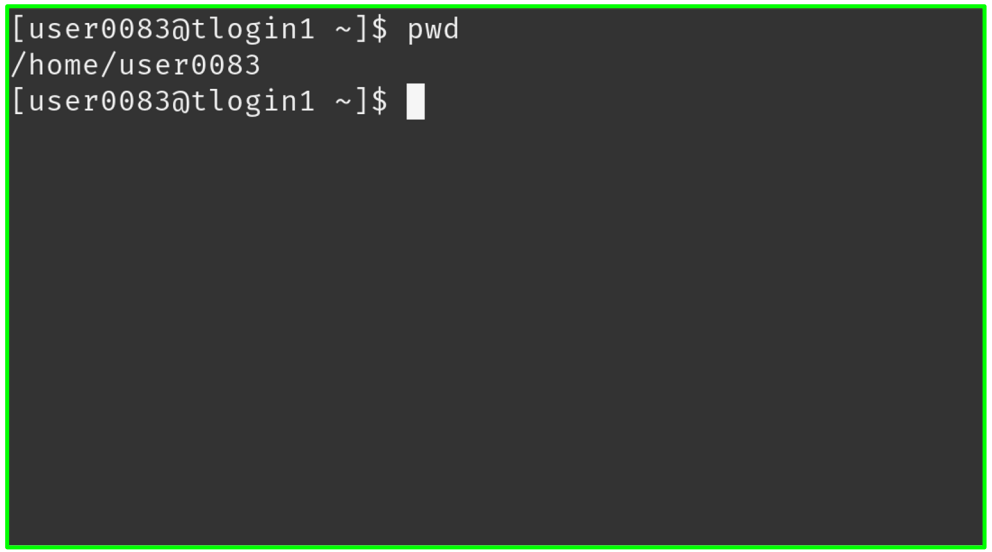
Using CURC and Alpine resources

How to Access RC Resources?

1. Get an RC account
2. Set up two-factor authentication with Duo
3. (Inform us of any specific needs)
4. Log in
5. Create greatness! (responsibly)

Accessing CURC

- Mac or Linux
  - Terminal application
- Windows
  - PuTTY
  - Powershell



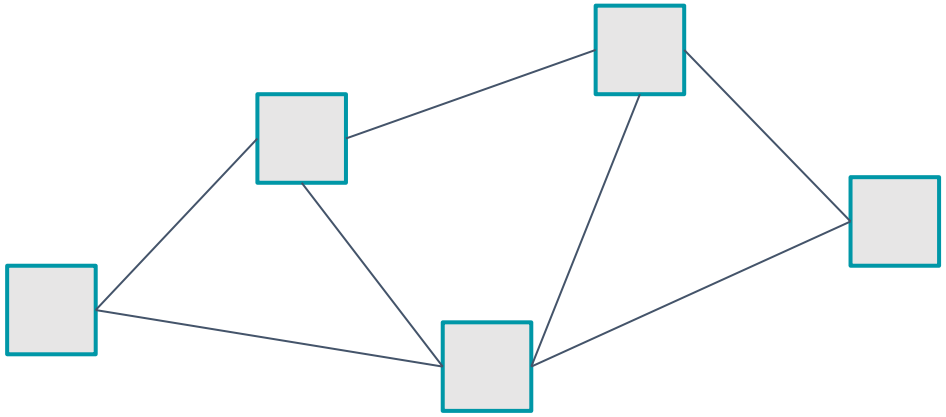
- Open OnDemand (*alternative for CU affiliates*)
  - For those less familiar with Linux (ondemand.rc.colorado.edu/)

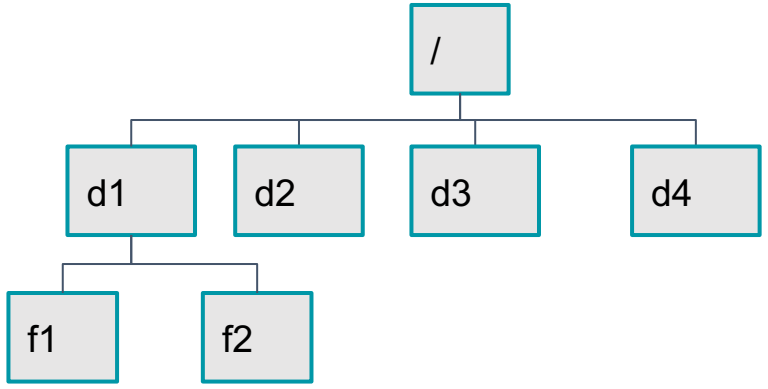
Node Types

Login	Compile	Compute
• Where you log in to • For editing code, job submission • No heavy computation	• Where you compile code, install packages • Explore the Alpine software environment • Edit code, submit jobs • No heavy computation	• Where scheduled jobs run • Intended for heavy computation
Ex. edit job script	Ex. Install python libs	Ex. Running Matlab

Navigating CURC

Node	File System
• One computing server • Physical hardware • Work together in parallel	• The basic tree-like layout • From most nodes* you have access to most file systems





Alpine Compile Nodes

- If you have used Summit in the past, compile nodes work *slightly* differently:
  - Instead of having dedicated hardware (2 nodes) which are oversubscribed for users to `ssh` into
  - Alpine's **`acompile`** command starts an interactive job which users can compile in which provides the following benefits:
    - Users can request specific resources (i.e. more cores to compile with)
    - Limits dedicated hardware set aside

Alpine Software

- Once logged in, type:

```
$ acompile
```

  To start an Alpine compile job.

- Once on a compile node, type:

```
$ module avail
```

  To list currently available software

- Alternatively, type:

```
$ module spider <software>
```

  To search for a specific software

Filesystem Structure

/home (2GB)	/projects (250GB)	/scratch/alpine (10TB)
• Scripts, Code, Small, important files/directories • Not for sharing files or job output	• Code/files/libraries • Software you are installing • Sharing files • Not for job output	• Output from running jobs • Large files/datasets • Sharing files • Cluster specific • Not for long term storage
Ex .bashrc	Ex. Shared job scripts	Ex. Data

Exploring the Filesystem

- Once logged in use the following commands to navigate to your different workspaces

```
$ cd  /home/<user>
$ cd  /projects/<user>
$ cd  /scratch/alpine/<user>
```

Moving Data to/from CURC

- Globus
  - Web-based file transfer
- FileZilla
  - File transfer GUI application that can be used on Windows, Mac, and Linux.
- Secure Copy (scp)
  - Can send data to and fetch data from a remote server.
- Rsync
  - Can be used to synchronize files and directories across two locations, which can often lead to efficiencies in repeat-transfer scenarios.
- SFTP
  - An interactive alternative to scp that allows multiple, bi-directional transfer operations in a single session.
- Rclone
  - Command line program to manage files on cloud storage.

Check out our documentation on File Transfer for guides to using each of these tools:
https://curc.readthedocs.io/en/latest/compute/data-transfer.html

Using Alpine compute: Jobs

What is a "***job***"?

- Work for the cluster to perform on
- Has a unique ID

1. **Batch jobs**
   - Submit job script which will be executed when resources are available
     - Create script containing information about the job
     - Submit the job file to a queue
2. **Interactive jobs**
   - Work interactively at the command line of a compute node

SLURM

- **S**imple **L**inux **U**tility for **R**esource **M**anagement

- Through SLURM users can:
  - Schedule jobs on specific compute resources
  - Run jobs interactively or hands off
  - Query job statistics

Alpine Partitions

Partition	Description	# of nodes	RAM/core (GB)	cores/node	GPUs/node
amilan	General Compute Node: AMD Milan	64	3.74	64	0
ami100	GPU Node: 3x AMD MI100	8	3.74	64	3
aa100	GPU Node: 3x Nvidia A100	8	3.74	64	3
amem	High-memory node	4	21.5	48	0

QOS: normal (24-h), long-7-d

Alpine Allocations

How can I use more computational time?:

- Trailhead Allocation (Default)
  - ~2,000 SUs / Month
- Ascent Allocation
  - 250,000 SUs
- Peak Allocation
  - >250,000 SUs

Request an allocation at https://curc.readthedocs.io/en/latest/clusters/alpine/allocations.html

Help! I'm stuck, where do I go?

- **Documentation**: curc.readthedocs.io/
- **Trainings with Center for Research Data and Digital Scholarship (CRDDS)**: https://www.colorado.edu/crdds/
- **RC Summer Camp Slides**: https://github.com/ResearchComputing/Summer_Camp_2023
- **Helpdesk**: rc-help@colorado.edu
- **Consult Hours (now!)**

Questions

CURC User Policies: https://curc.readthedocs.io/en/latest/additional-resources/policies.html?highlight=policies#curc-user-policies

Survey and feedback

http://tinyurl.com/curc-survey18