Module 1: Getting Familiar with RC

Meet the User Support Team

Layla
Freeborn

Brandon
Reyes

Andy
Monaghan

Michael
Schneider

John
Reiland

Dylan
Gottlieb

Mohal
Khandelwal

Ragan
Lee

Learning Goals

1. Understand CURC Resources & the Alpine cluster
2. Getting an account & logging in
3. Navigate the RC system

Ask Questions

Discuss Ideas

RC Resources

Tech Support

- High Performance Computing
- Data Management
- Cloud Computing
- Secure Research

Human Support

- Training Materials & Workshops
- Consultations & Office Hours
- Help Tickets

High Performance Computing (HPC)

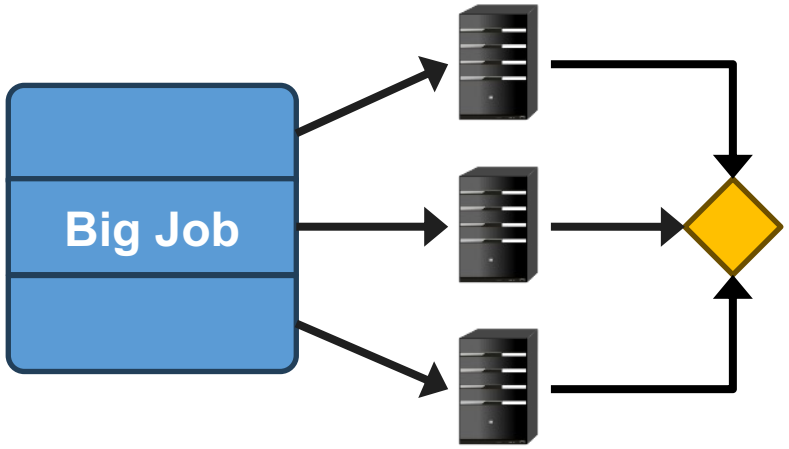
Traditional vs HPC

Fewer Computing Resources, Fast Sprinter*

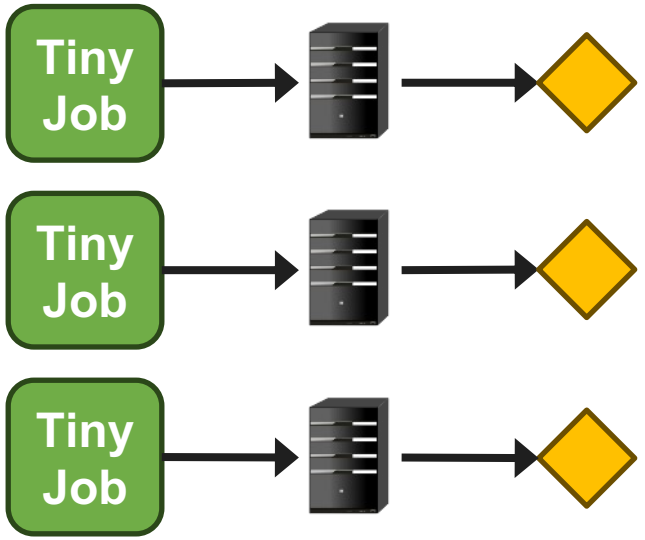
More Computing Resources, Steady Marathon Runner*

***Potentially**

What can / use HPC for?

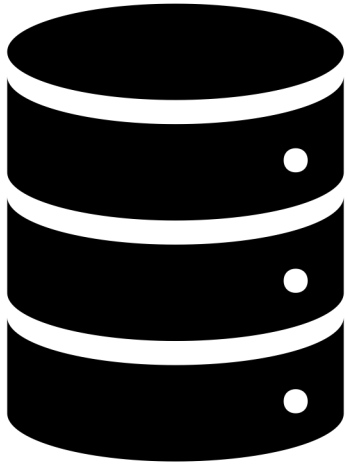


Parallel Jobs

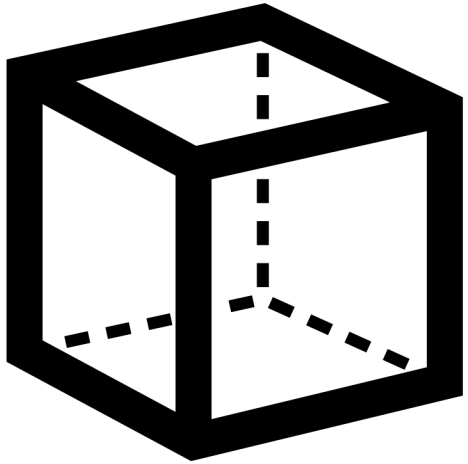


Serial Jobs

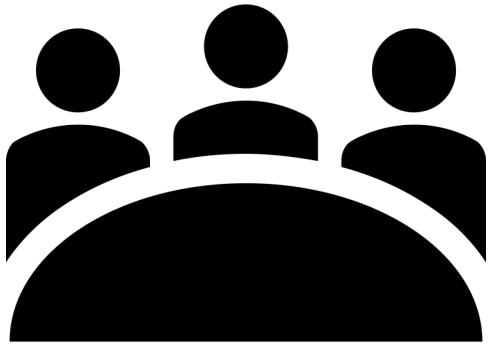
What can / use HPC for?



Big Data



Viz/Rendering



Shared Work Environment

HPC Cluster: Alpine

- 3rd-generation HPC cluster
  - Janus (2012)
  - RMACC Summit (2017)
- Heterogeneous cluster with hardware currently provided by CU Boulder, CSU, and Anschutz
- Access available to CU Boulder, CSU, AMC and RMACC users

Alpine Partitions

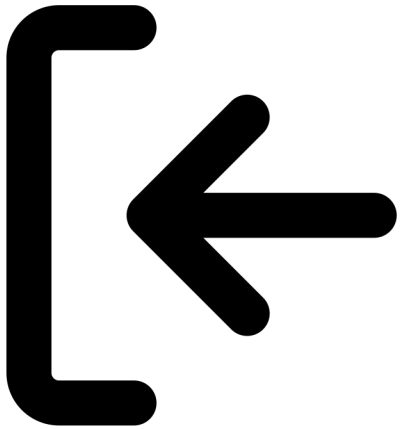
Partition	Description	# of nodes	RAM/core (GB) (qty.)	cores/node (qty.)	GPUs/node
amilan/amilan128c	General Compute Node: AMD Milan	390 (amilan), 16 (128c)	3.74 (390), 2.01 (16)	128 (16), 64 (313), 48 (28), 32 (49)	0
ami100	GPU Node: 3x AMD MI100	8	3.74	64	3
aa100	GPU Node: 3x Nvidia A100	12	3.74	64	3
amem	High-memory node	24	21.5 (12), 16 (12)	128 (2), 64 (10), 48 (12)	0
al40	GPU Node: 3x Nvidia L40	3	3.74	64	3

HPC Cluster: Alpine

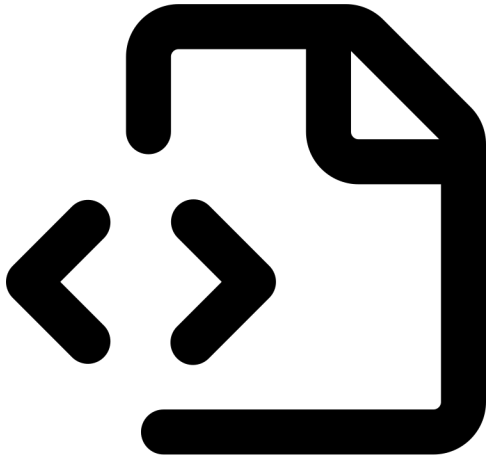
- Interconnect
  - **CPU nodes**: HDR-100 InfiniBand* (200Gb inter-node fabric)
  - **GPU nodes**: 2x25 Gb Ethernet +RoCE
  - **Scratch Storage**: 25Gb Ethernet +RoCE
- Operating System
  - RHEL 8 - **R**ed**H**at **E**nterprise **L**inux Version 8

* Not all CPU Nodes have Infiniband

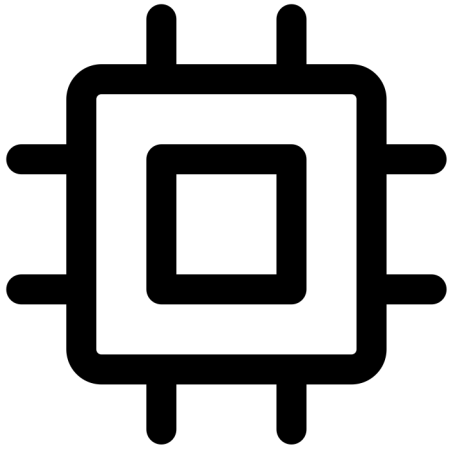
Node Types



LOGIN

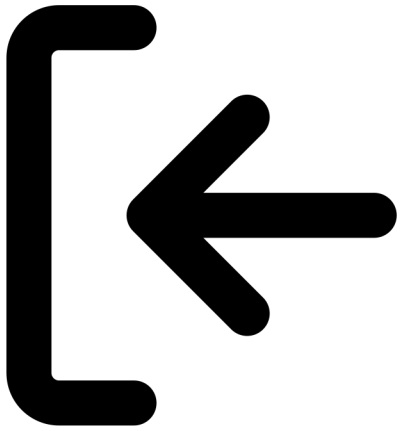


COMPILE

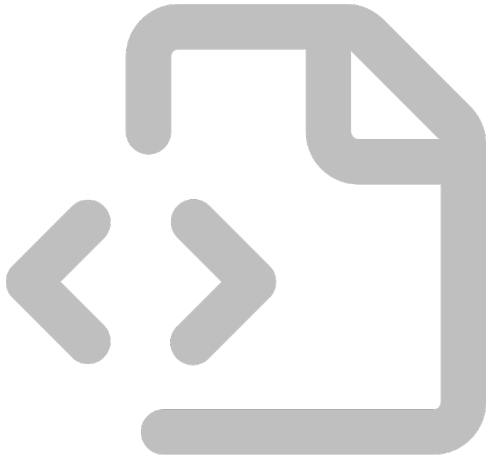


COMPUTE

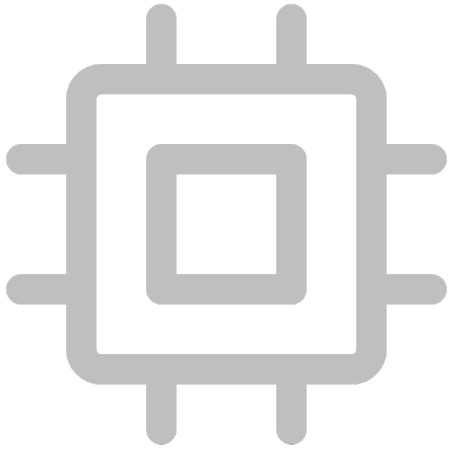
Node Types



LOGIN



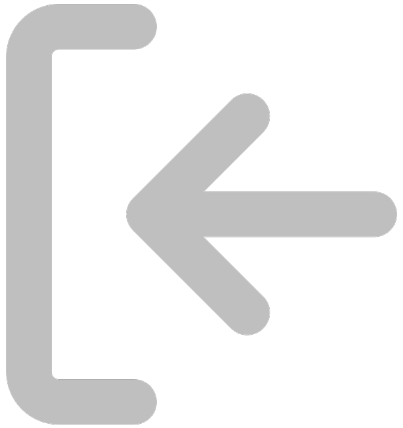
COMPILE



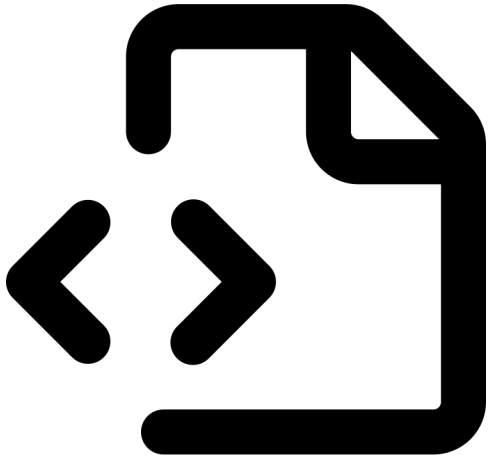
COMPUTE

- Entry to system
- View or edit files
- Submit jobs

Node Types

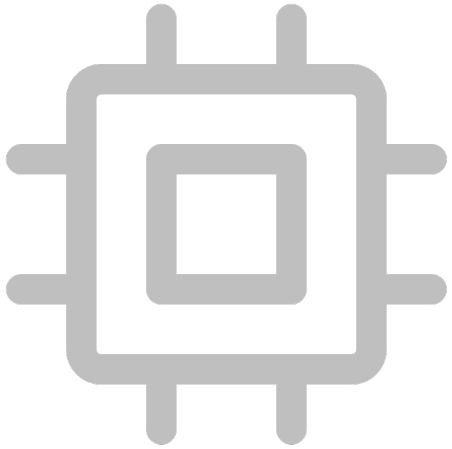


LOGIN



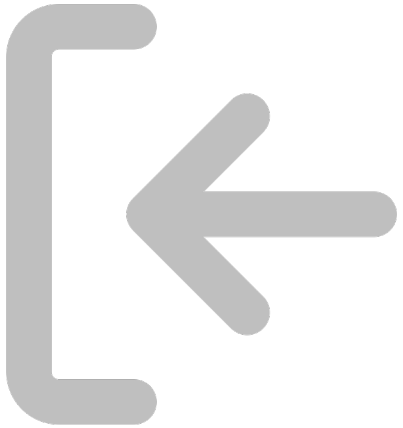
COMPILE

- View or edit files
- Submit jobs
- **Compile code**
- **Install software**

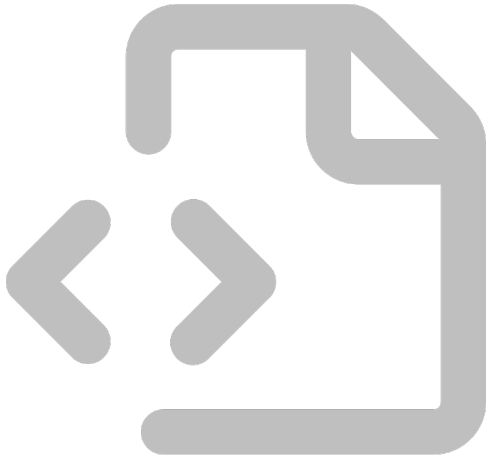


COMPUTE

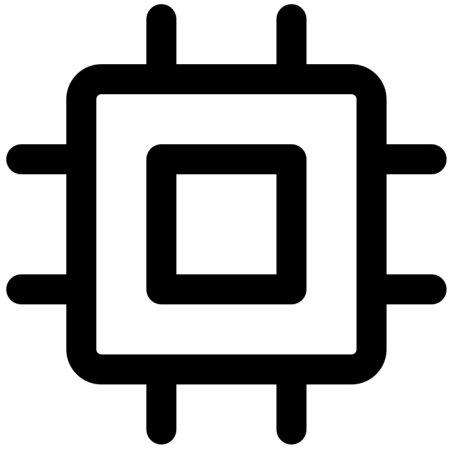
Node Types



LOGIN



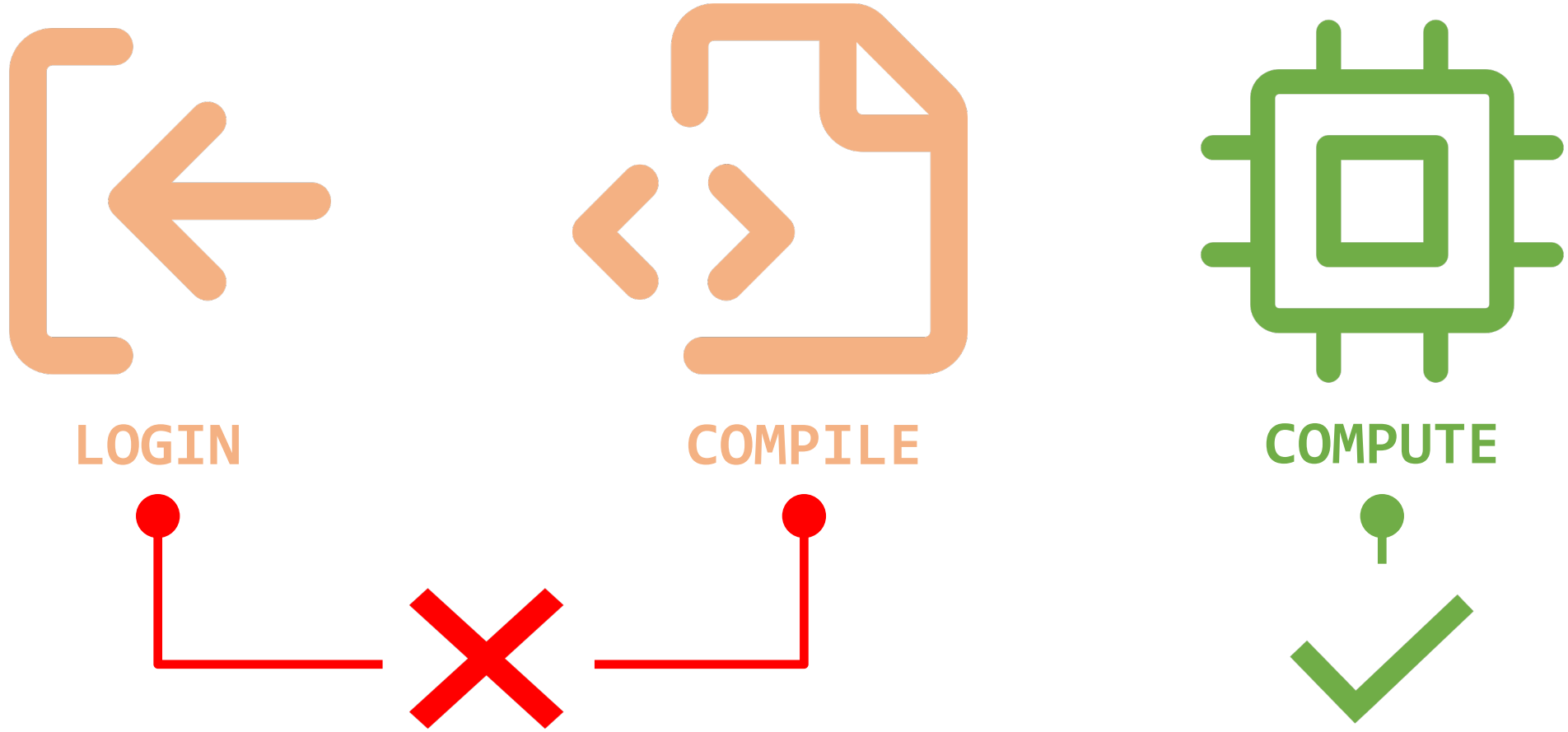
COMPILE



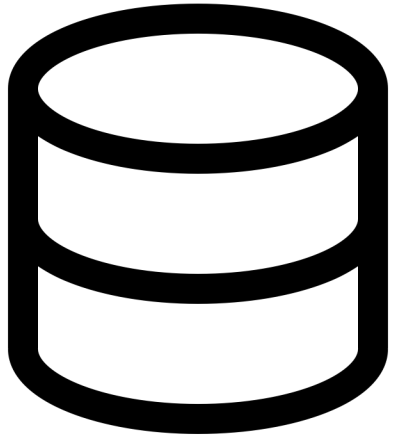
COMPUTE

- Run scheduled jobs
- Handle calculations

Where do I run jobs?



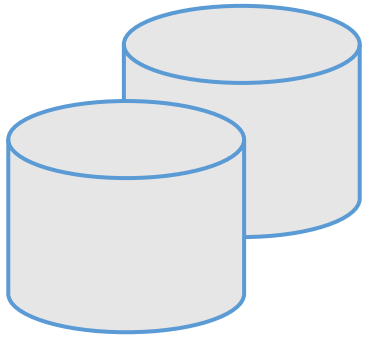
Node Types



DATA (DTN)

- Support data transfers
- Can be selected when using scp, sftp, or ssh transfers

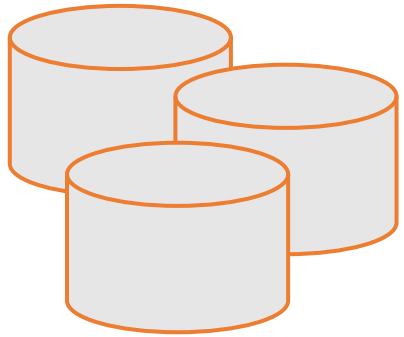
Storage at CURC



Core

Included with RC account

- /home
- /projects
- /scratch



PetaLibrary

Paid Service for large data

- Storage
- Archival Backup
- Sharing

Core Filesystem Structure

/home (2GB)	/projects (250GB)	/scratch/alpine (10TB)
• Scripts and Code • Important files (small)		
Not for: • File Sharing • Large Files		
Configuration Files, Notes		

Core Filesystem Structure

/home (2GB)	/projects (250GB)	/scratch/alpine (10TB)
• Scripts and Code • Important files (small)	• Code/files/libraries • Software you are installing • Sharing files	
Not for: • File Sharing • Large Files	**Not for** • Job Output • Temporary Files	
Configuration Files, Notes	Job Scripts, Shareable Files	

Core Filesystem Structure

/home (2GB)	/projects (250GB)	/scratch/alpine (10TB)
• Scripts and Code • Important files (small)	• Code/files/libraries • Software you are installing • Sharing files	• Output from running jobs • Large files/datasets • Sharing files
Not for: • File Sharing • Large Files	**Not for** • Job Output • Temporary Files	**Not for** • Long Term Storage
Configuration Files, Notes	Job Scripts, Shareable Files	Job Output Files, Data Files

PetaLibrary Tiers

Active	Archive	Active + Archive	Archive + DR
• Performance Tier[1] • Accessible by All Nodes • No File Limit • Double Parity • lz4 compression			

[1] Slower than /scratch, [2] Per Terabyte

PetaLibrary Tiers

Active	Archive	Active + Archive	Archive + DR
• Performance Tier[1] • Accessible by All Nodes • No File Limit • Double Parity • lz4 compression	• Integrity Tier • Accessible by Login Nodes Only • 10,000 File Limit[2] • Triple Parity • zstd compression		

[1] Slower than /scratch, [2] Per Terabyte

PetaLibrary Tiers

Active	Archive	Active + Archive	Archive + DR
• Performance Tier[1] • Accessible by All Nodes • No File Limit • Double Parity • lz4 compression	• Integrity Tier • Accessible by Login Nodes Only • 10,000 File Limit[2] • Triple Parity • zstd compression	• Replicate Active into Archive • Updated every 15 minutes • No File Limit	

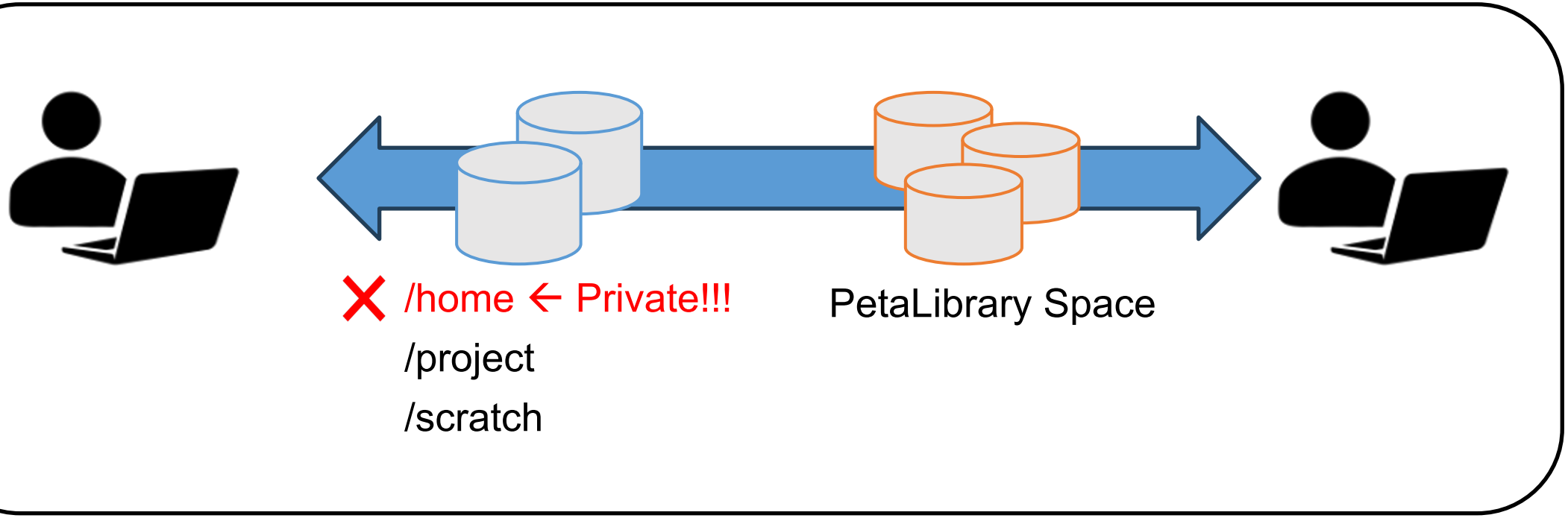
[1] Slower than /scratch, [2] Per Terabyte

PetaLibrary Tiers

Active	Archive	Active + Archive	Archive + DR
• Performance Tier[1] • Accessible by All Nodes • No File Limit • Double Parity • lz4 compression	• Integrity Tier • Accessible by Login Nodes Only • 10,000 File Limit[2] • Triple Parity • zstd compression	• Replicate Active into Archive • Updated every 15 minutes • No File Limit	• Disaster Recovery • Monthly backup to offsite location

[1] Slower than /scratch, [2] Per Terabyte

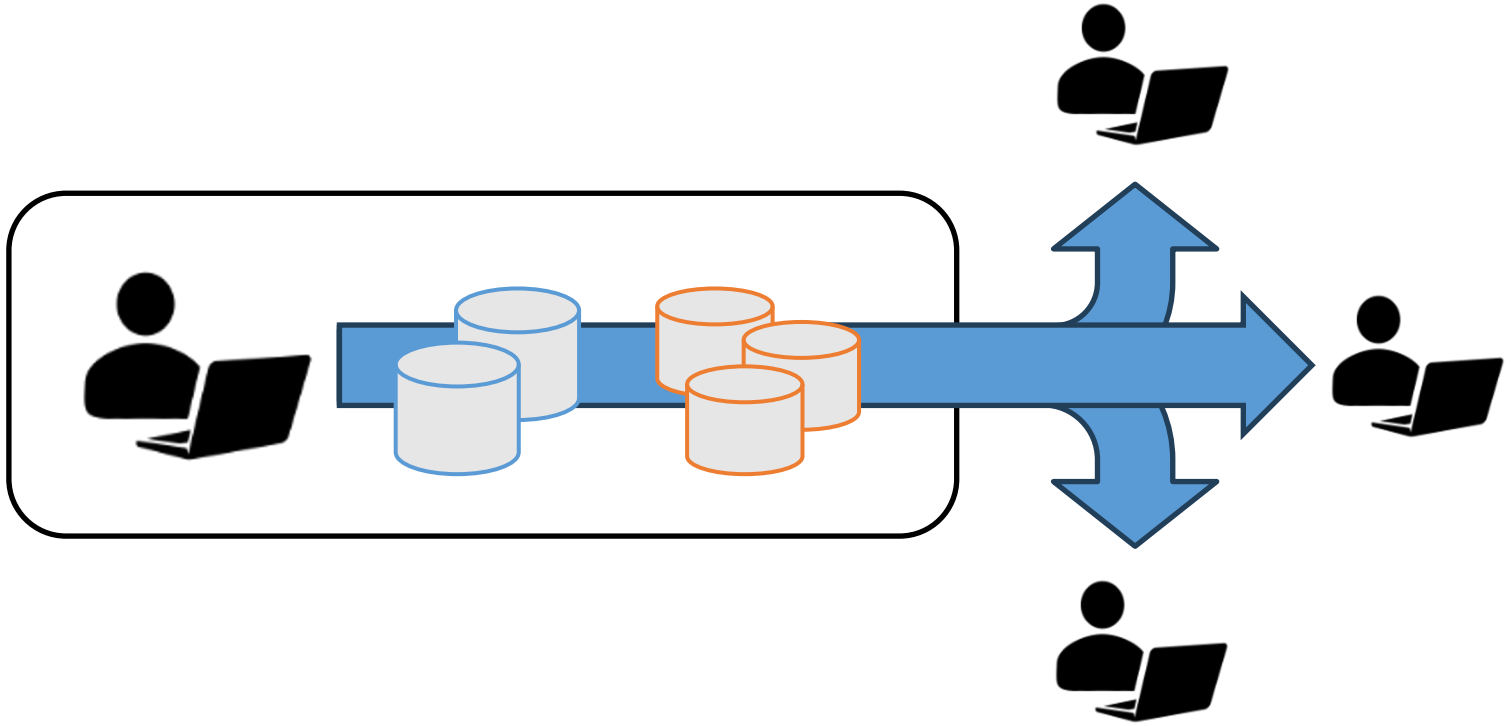
Data Sharing: Within RC



Data Sharing: Outside RC

Large Data Transfers:

- Globus (Recommended)
- Data Transfer Nodes (DTN)
- Terminal/Command Line:
  - rsync
  - rclone
  - sftp
  - scp



Acceptable data storage and use

CURC systems and services **should not be used to store** any data that is US government Classified, nor any Controlled Unclassified Information.

For users requiring storage for sensitive data types, please see the secure research computing resources:

https://www.colorado.edu/rc/secure-research-computing-resources

Cloud Computing

- CURC supports AWS, Azure, and GCP
- Alternative to HPC
- Enhance HPC

How to Access RC Resources?

1. Get an RC account
2. Set up two-factor authentication with Duo
3. Log in
4. Create greatness! (responsibly)

Getting an Account

- CU Boulder, CSU users and affiliates:
  - Request an account through the RC Account request portal: https://rcamp.rc.colorado.edu/accounts/account-request/create/organization
- AMC, RMACC users and affiliates:
  - Request an account through the ACCESS-CI User Registration Portal: https://identity.access-ci.org/new-user.html

Your RC Account

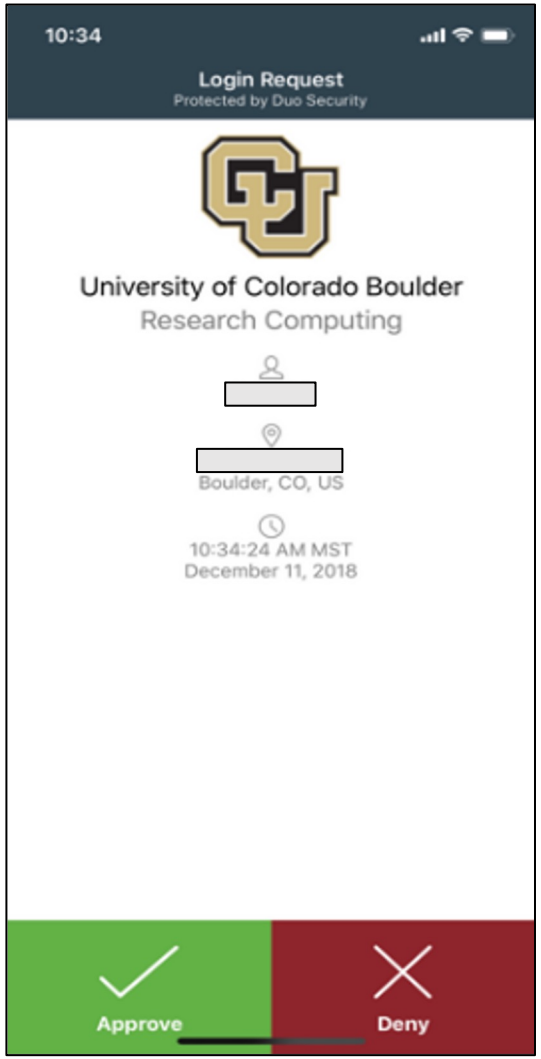
Access to:

1. Alpine Cluster
2. Core Storage
3. PetaLibrary Storage*
4. Open OnDemand
5. Approximately 2,000 Service Units (SUs) per month

*If purchased

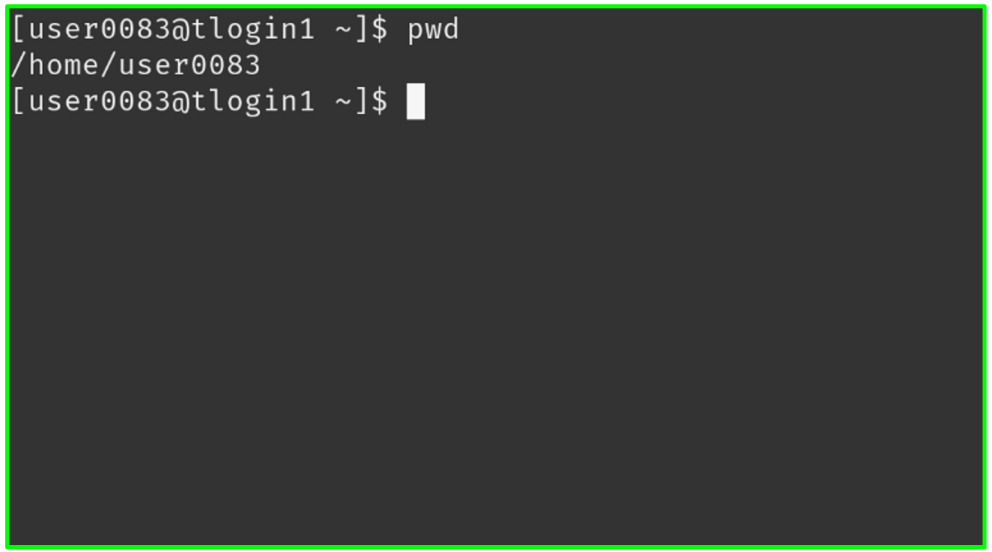
Duo Authentication

1. **Duo smartphone app (recommended)**
2. Phone Call/Text is an alternatives



Terminal Access

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



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

Demo: Logging in via Terminal

To login to an RC login node:

```
$ ssh <username>@login.rc.colorado.edu
```

Supply your IdentiKey* password and your Duo app will alert you to confirm the login

* Exclusive to CU and CSU accounts

Demo: logging in with OnDemand

CURC Open OnDemand is a browser based, integrated, single access point for all of your HPC resources at CU Research Computing.

- CU Boulder: Visit https://ondemand.rc.colorado.edu.
- Other RMACC Institutions: Visit https://ondemand-rmacc.rc.colorado.edu/

Asking for Help

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

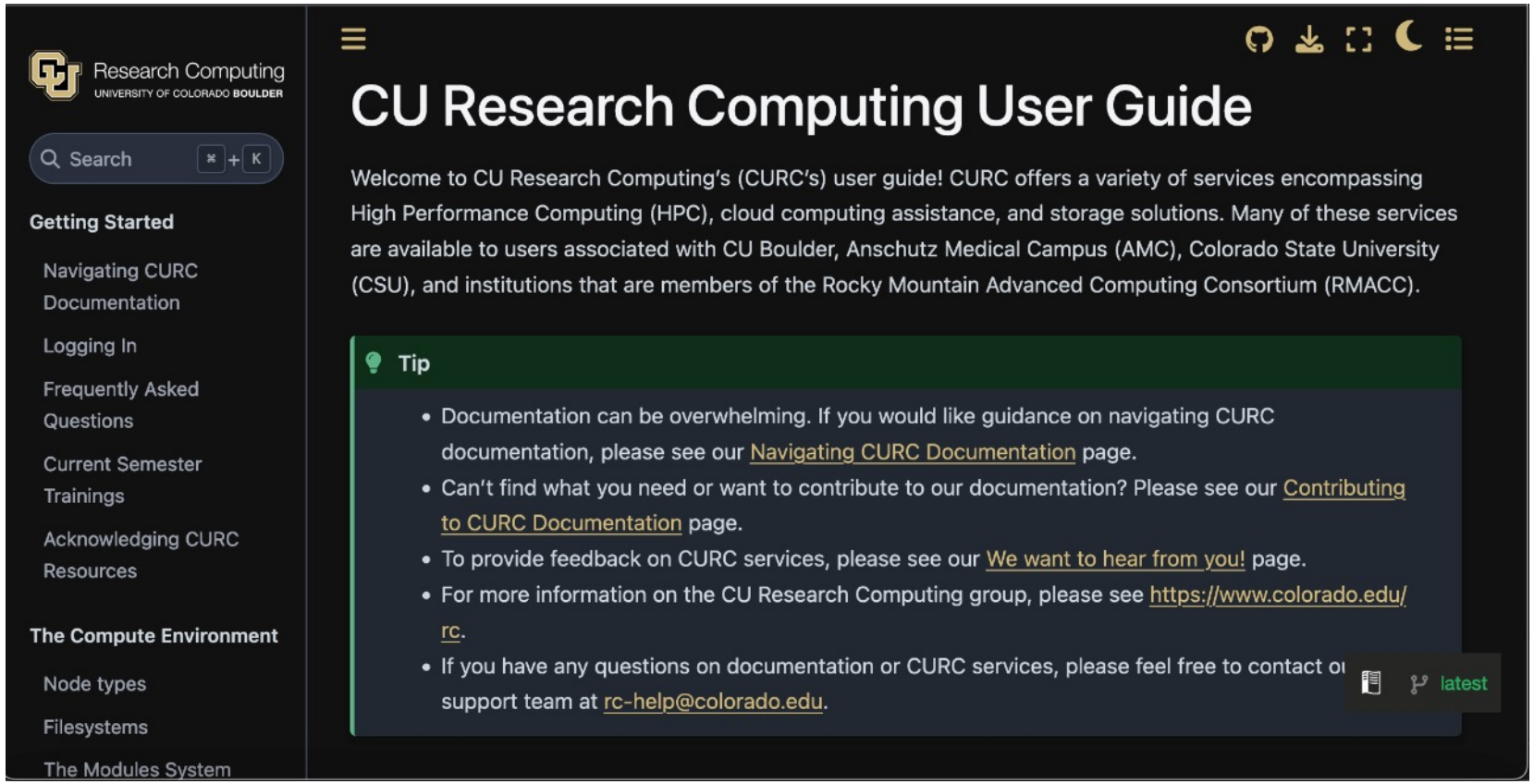
- **CURC Documentation**: curc.readthedocs.io
- **External Resources**
  - The Internet! (Stack Overflow, YouTube, etc.)
- **Trainings & Consults with Center for Research Data and Digital Scholarship (CRDDS)**
- **CURC Helpdesk**: rc-help@colorado.edu

When should I use these?

- **Documentation**: curc.readthedocs.io
  - Useful at any time! Check the documentation first when you run into issues.
- **External Resources**
  - Useful for learning a new skill or initial troubleshooting.
- **Trainings with Center for Research Data and Digital Scholarship**
  - Useful for broad, long-term learning
  - Drop-in consult hours are held Tue (12-1p) and Thu (1-2p) during the Fall and Spring semesters
- **CURC Helpdesk**: rc-help@colorado.edu
  - Useful for quick, personalized assistance. We can schedule Zoom consults if needed.

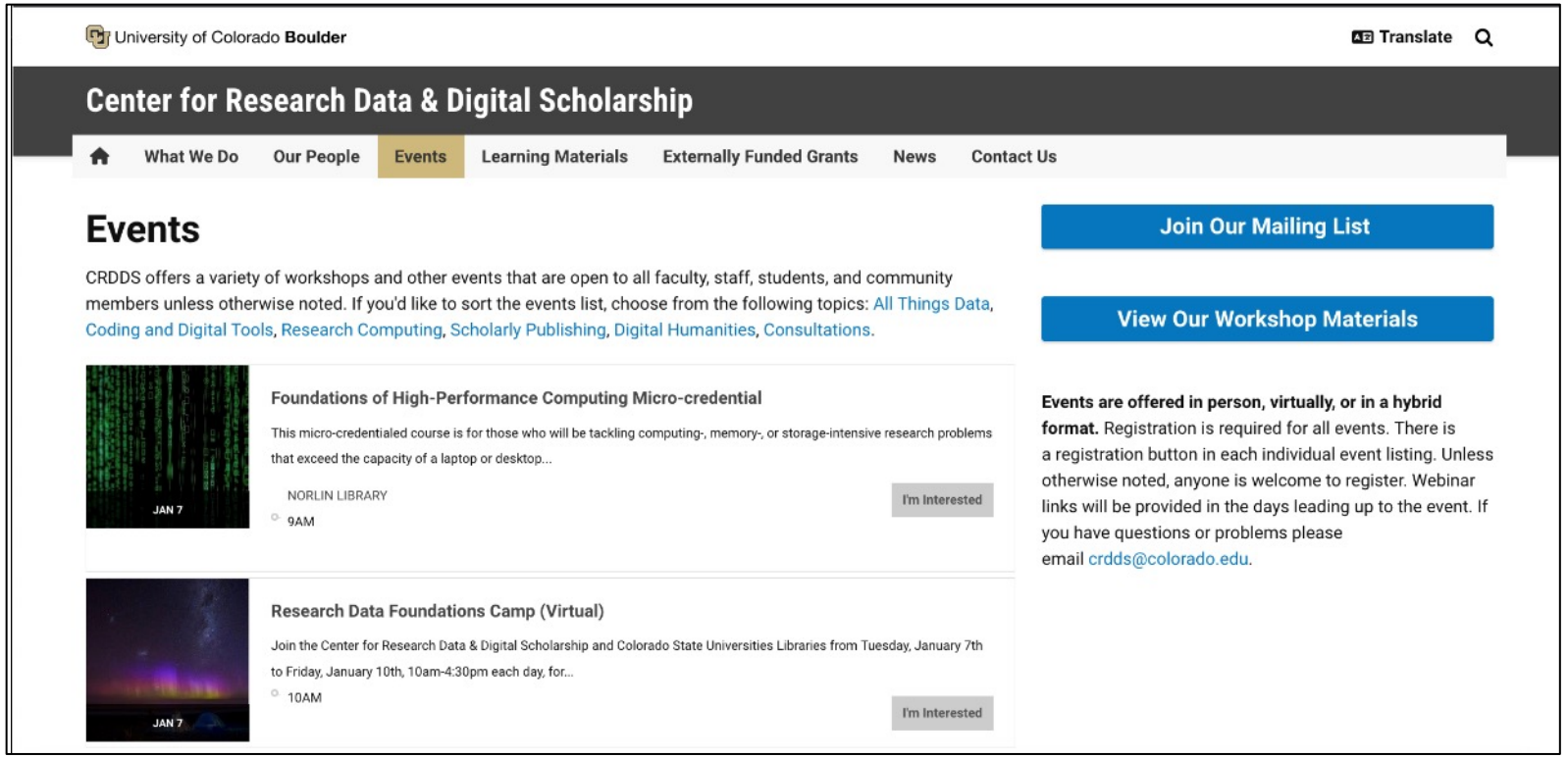
Our Documentation

Located at: https://curc.readthedocs.io



CRDDS trainings and consult hours

View upcoming events at: https://buff.link/crddsevents



Helpdesk Tickets: sub-optimal vs optimal

To: rc-help@colorado.edu

Dear Research Computing,

Help! My code won't run! Help!

Help please,
Andy

To: rc-help@colorado.edu

Dear Research Computing,

I am running into issues running my Python script. I am using a conda environment called my_python_env with the pytorch software, and I am receiving the following error. I am not sure how to troubleshoot. My job ID is 620350. Let me know what I can try!

srun: fatal: SLURM_MEM_PER_CPU, SLURM_MEM_PER_GPU, and SLURM_MEM_PER_NODE are mutually exclusive.

Thanks,
Andy

Helpdesk Tickets: sub-optimal vs optimal

To: rc-help@colorado.edu

Dear Research Computing,

Can you install pytorch for me?

Thanks,
Andy

To: rc-help@colorado.edu

Dear Research Computing,

I am looking to utilize PyTorch to use in conjunction with AMD GPUs. I have tried an anaconda installation and have so far been unsuccessful. Could you please help me complete this install?

Thanks,
Andy

How can I compose an effective ticket?

- Provide detail!
- Scale down your workflows for testing!
- Email our helpdesk!
- Try a few things and let us know what you've tried!
  - We are not just being lazy – it helps us contextualize the issue.
  - We would likely try the same things as you – if you can eliminate potential solutions, it will help us get to a solution more quickly.

Documentation

https://curc.readthedocs.io/en/latest/