RMACC New User Seminar

1

Be Boulder.

Website: www.rc.colorado.edu

Documentation: https://curc.readthedocs.io

Helpdesk: rc-help@colorado.edu

RC Resources for RMACC

Tech Support

- High Performance Computing

- Data Management

Human Support

- Training Materials & Workshops
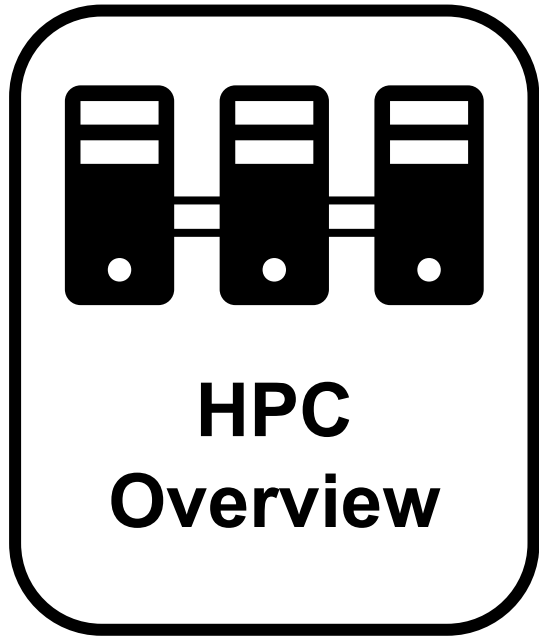
- Consultations & Office Hours

- Help Tickets

Research Computing
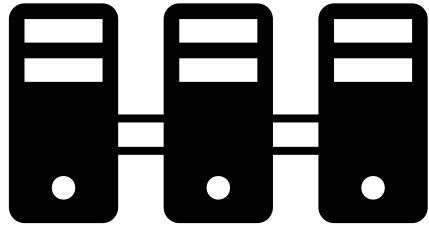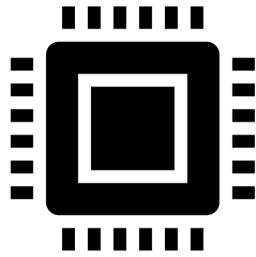UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Ask Questions

Discuss Ideas

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Learning Goals

HPC Overview

Research Computing
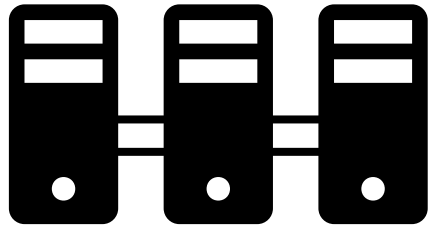UNIVERSITY OF COLORADO **BOULDER**
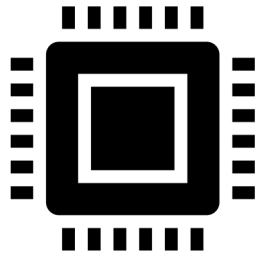
Be Boulder.

Learning Goals

HPC Overview

Nodes & Partitions

Learning Goals

HPC Overview	Nodes & Partitions	Data Storage

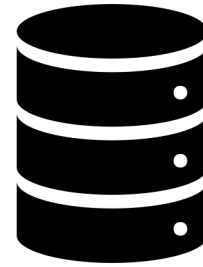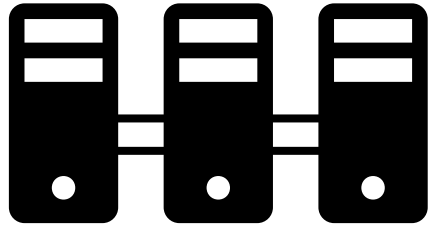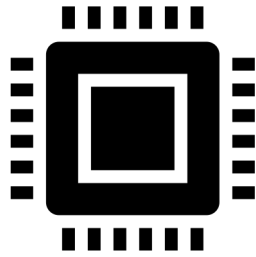Research Computing
UNIVERSITY OF COLORADO **BOULDER**
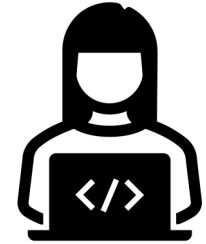
Be Boulder.

Learning Goals

HPC Overview

Nodes & Partitions

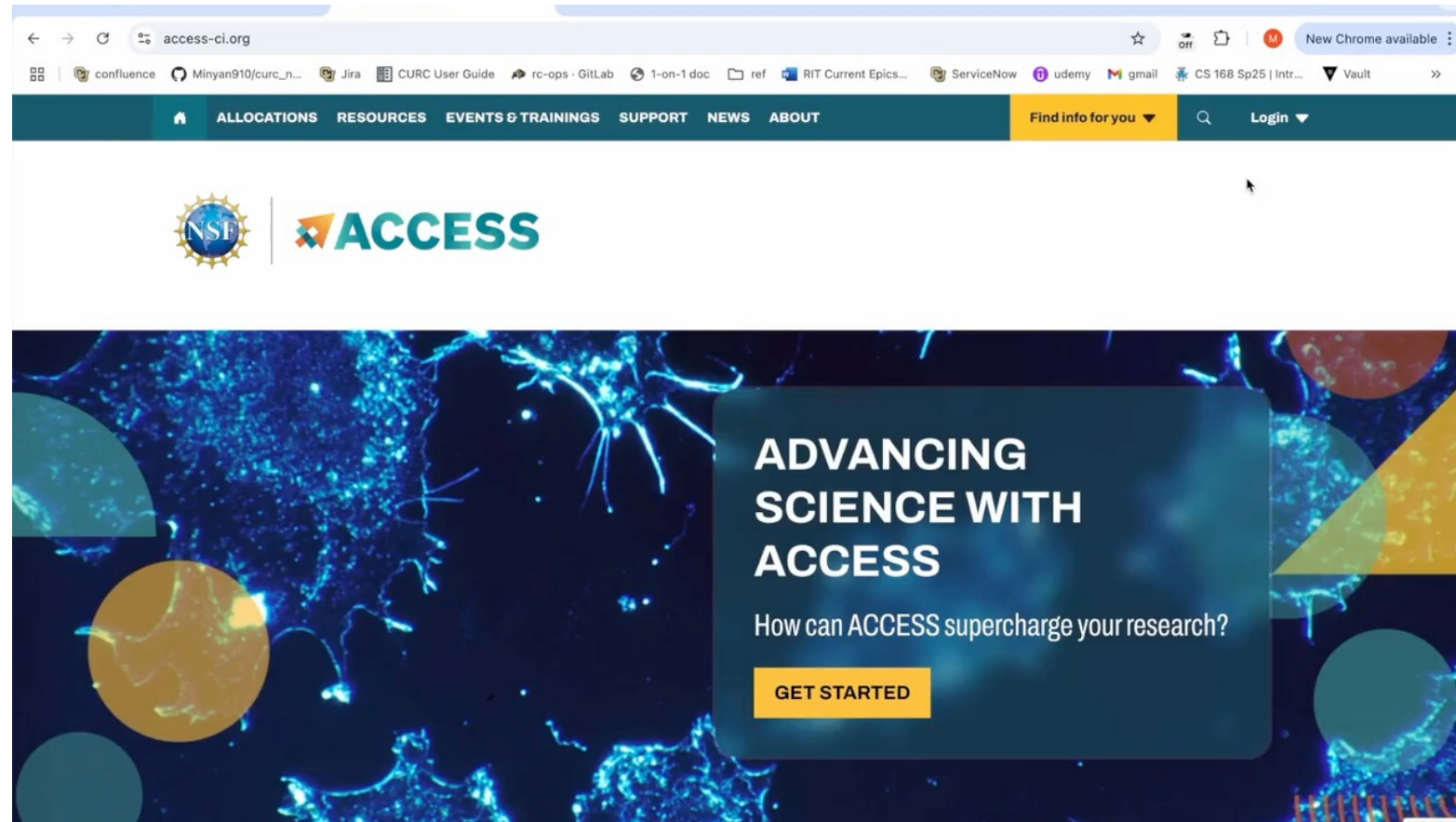**Data Storage**

System Access

How to Access RC Resources?

1. Get an RC account

2. Log in

3. 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

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Overview   Nodes   Data   **Login**

Be Boulder.

ACCESS Account Creation (Part 1)
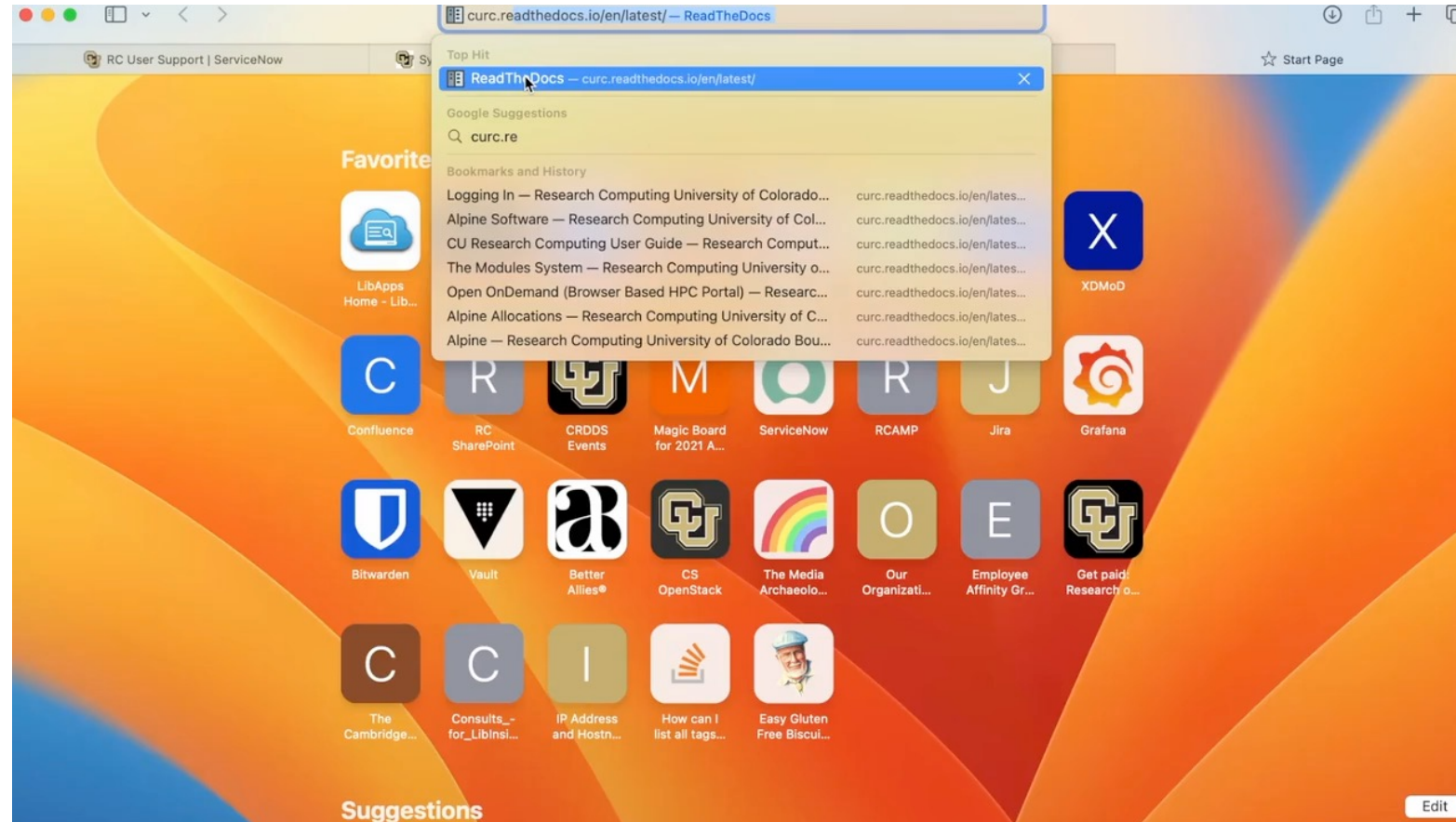
ACCESS Account Creation (Part 2)
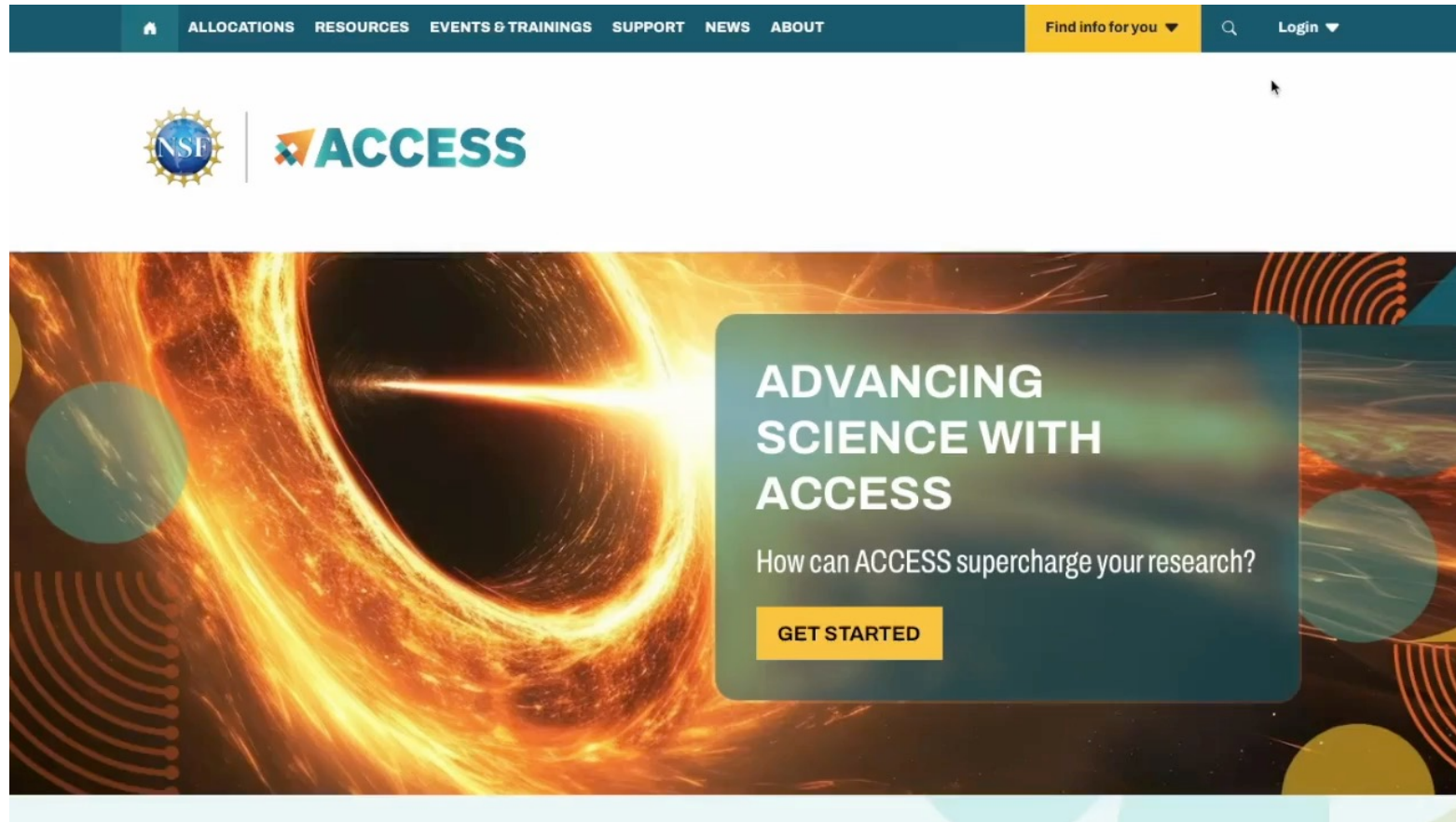
RMACC Account Form

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

*Only Active tier available to non-CU users, if purchased

High Performance Computing

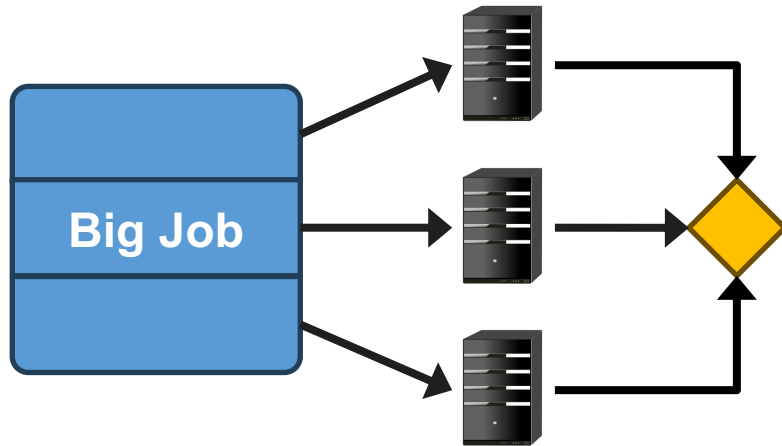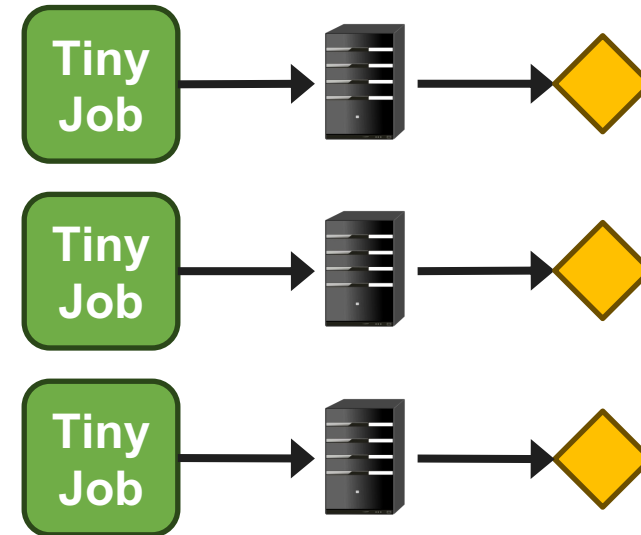**Scale**   **vs**   **Speed**

What can *I* use HPC for?

Parallel Jobs

Serial Jobs

What can *I* use HPC for?
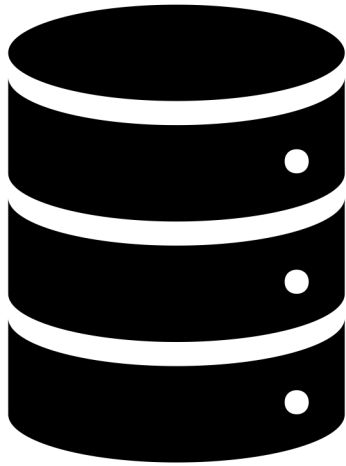
Big Data
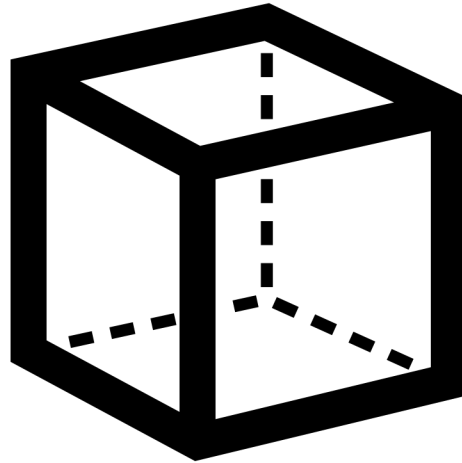
Viz/Rendering
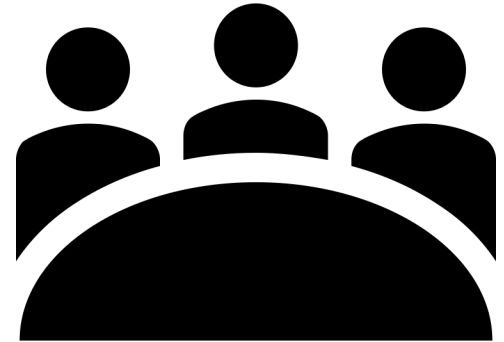
Shared Work Environment

HPC Cluster: Alpine

- Heterogeneous cluster

- Hardware provided by CU Boulder, CSU, and AMC

- Access available to CU Boulder, CSU, AMC and RMACC users

Node Types
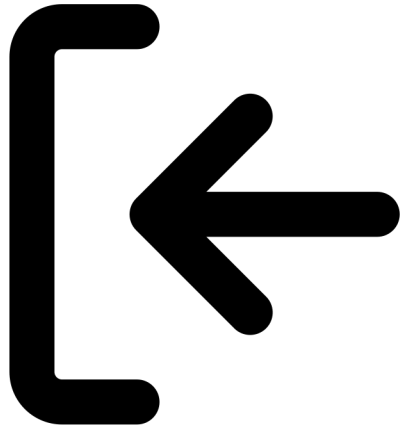
LOGIN
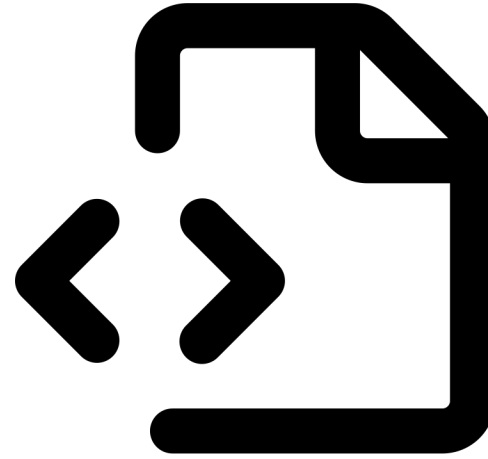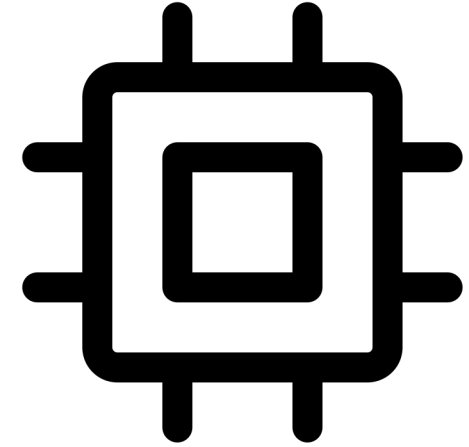
COMPILE

COMPUTE

Node Types

LOGIN

COMPILE

COMPUTE

- Entry to system
- View or edit files
- Submit jobs
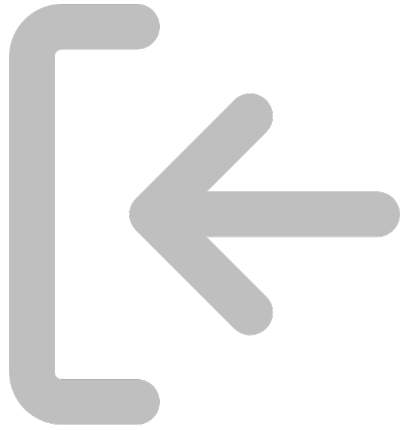
Node Types
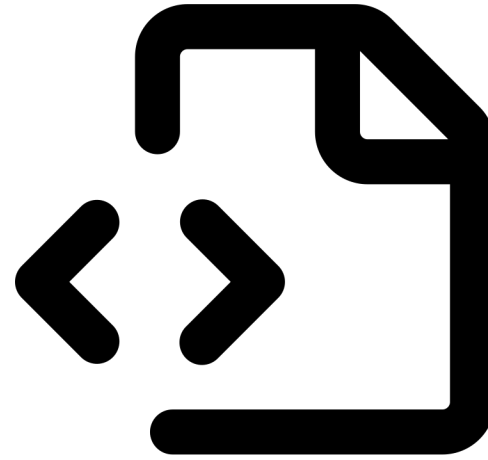
LOGIN

COMPILE

COMPUTE

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

Node Types

LOGIN
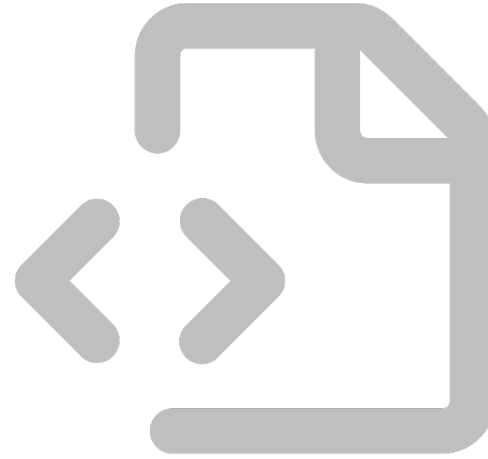
COMPILE
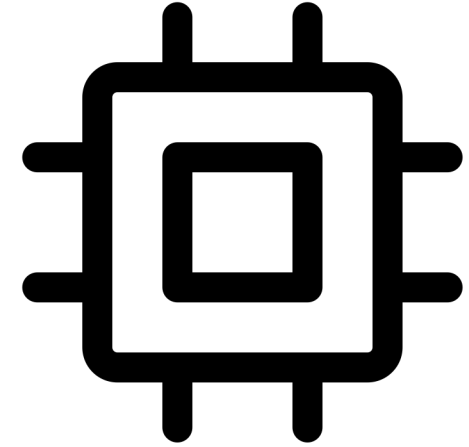
COMPUTE

- Run scheduled jobs
- Handle calculations

Alpine Partitions

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.
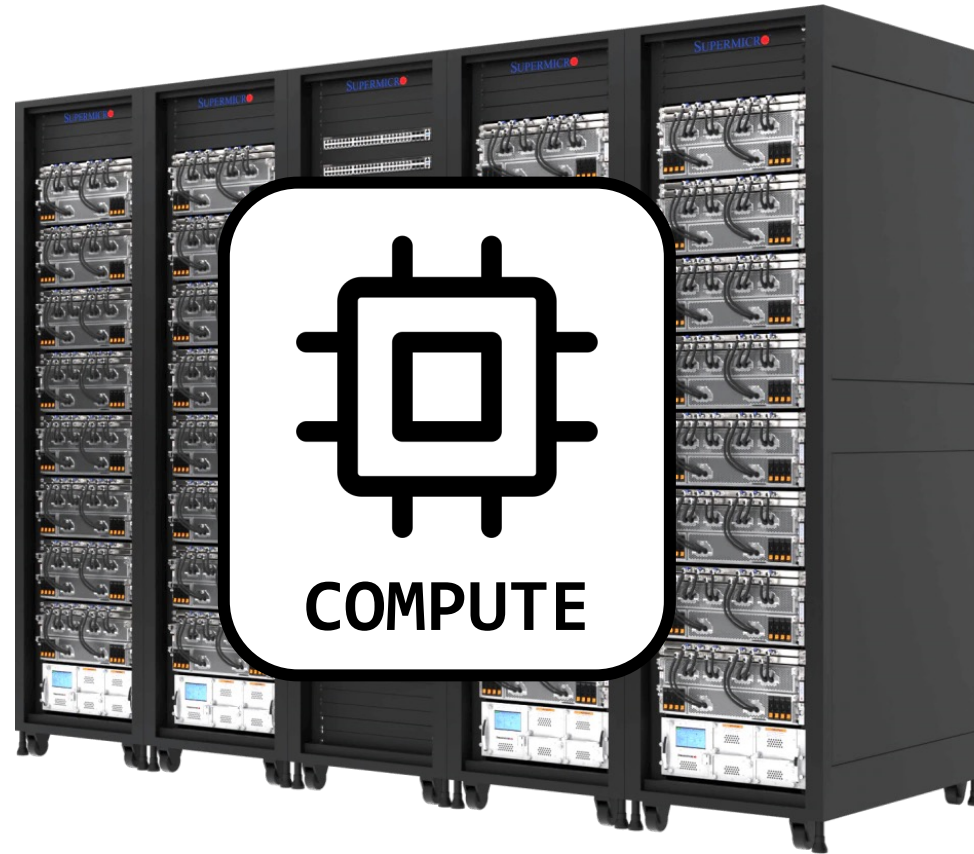
Alpine Partitions

amilan
General Usage

Alpine Partitions

amilan
General Usage

amem
High Memory

Alpine Partitions

amilan
General Usage

amem
High Memory

aa100
Nvidia GPU's

Alpine Partitions

amilan
General Usage

amem
High Memory

aa100
Nvidia GPU's

ami100
AMD GPU's

Data Storage

Core

- Personal Storage
- Includes 3 Directories
  - /home (2 GB)
  - /projects (250 GB)
  - /scratch (10 TB)

Data Storage

Core

- Personal Storage
- Includes 3 Directories
  - /home (2 GB)
  - /projects (250 GB)
  - /scratch (10 TB)

PL

- PetaLibrary
- Tiered Storage
  - Active, Archive
- Requires Funding
- Starts at 1 TB

Core

Home

- Personal data
- Config files
- OnDemand data
- Do not share!

2 GB

Core

Home

- Personal data
- Config files
- OnDemand data
- Do not share!

2 GB

Projects

- Code, job scripts
- Installed software
- Shared data

250 GB

Core

Home

- Personal data
- Config files
- OnDemand data
- Do not share!

2 GB

Projects

- Code, job scripts
- Installed software
- Shared data

250 GB

Scratch

- Job data (in/out)
- Shared data
- Deleted after 90 days

10 TB

PetaLibrary

Active

- Performance tier

- Accessible by all nodes

- No file limit

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

PetaLibrary

Active

- Performance tier
- Accessible by all nodes
- No file limit

Archive*

- Integrity tier
- Accessible by data transfer (DTN) and login nodes only
- 10,000 File Limit

*only available to CU System users

PetaLibrary

Active
- Performance tier
- Accessible by all nodes
- No file limit

Archive*
- Integrity tier
- Accessible by data transfer (DTN) and login nodes only
- 10,000 File Limit

Active + Archive*
- Same features as Active
- Copy is synced to separate on-campus datacenter every 15 mins.

*only available to CU System users

CU Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

PetaLibrary

Active

- Performance tier
- Accessible by all nodes
- No file limit

Archive*

- Integrity tier
- Accessible by data transfer (DTN) and login nodes only
- 10,000 File Limit

Active + Archive*

- Same features as Active
- Copy is synced to separate on-campus datacenter every 15 mins.

Archive + DR*

- Same features as Archive
- Copy is synced to offsite datacenter monthly

*only available to CU System users

Data Sharing: Within RC

❌ /home ← Private!!!

/projects

/scratch

PetaLibrary Space
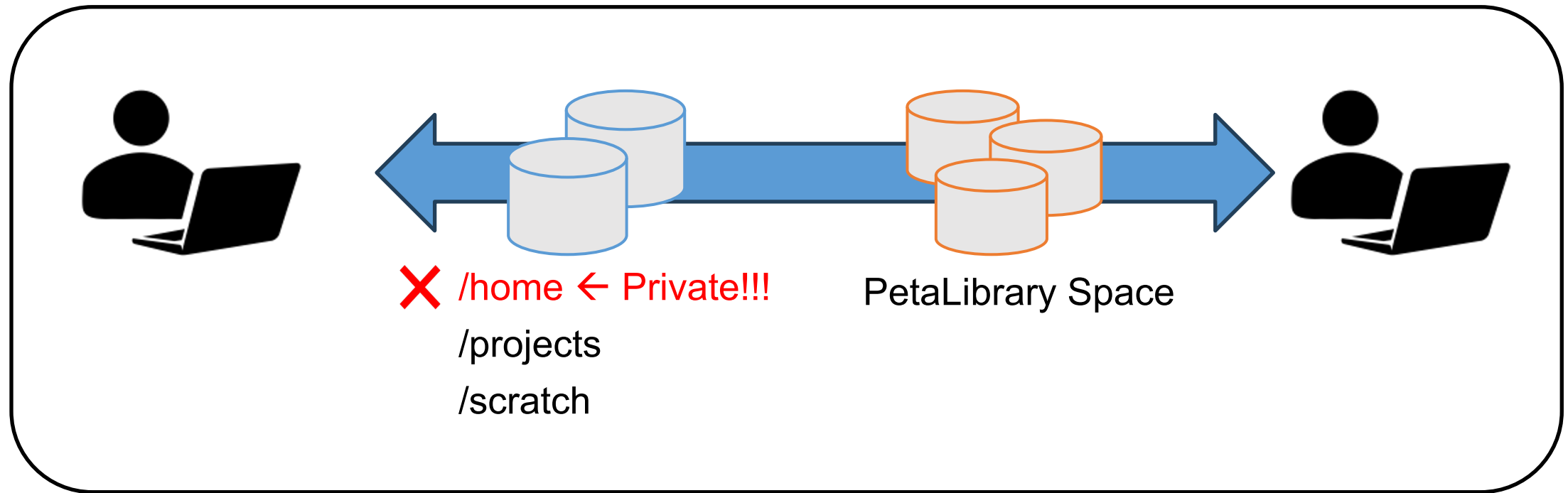
Data Sharing: Outside RC

Large Data Transfers:

- Globus (Recommended)

- Data Transfer Nodes (DTN)

- Terminal/Command Line:
  - rsync
  - rclone
  - sftp
  - scp

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Overview Nodes **Data** Login

Be Boulder.

Data Transfer Nodes
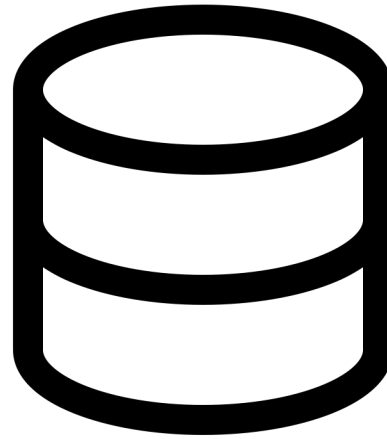
DATA (DTN)

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

Acceptable data storage and use

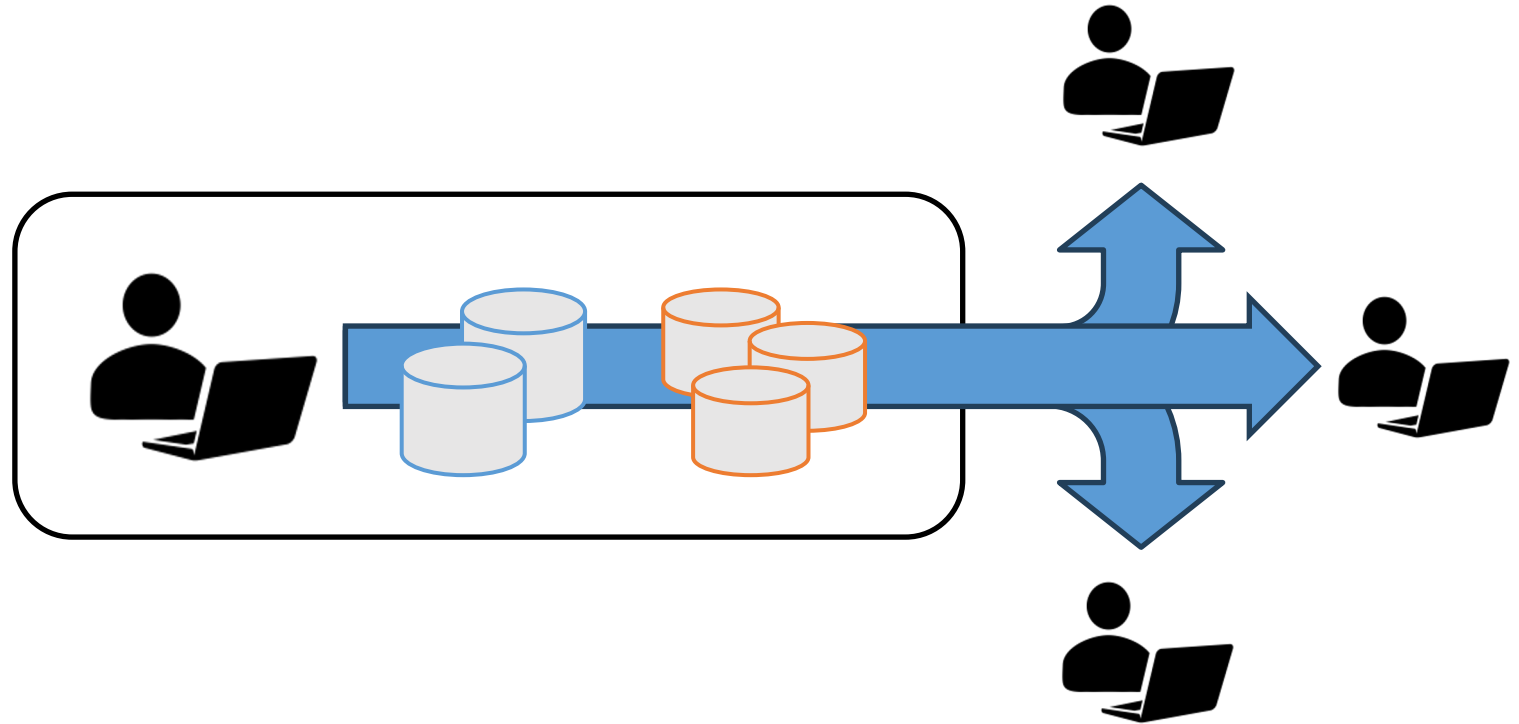CURC systems and services **should not be used to store** any data that is US government Classified, nor any Controlled Unclassified Information.

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

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

Demo: logging in with OnDemand

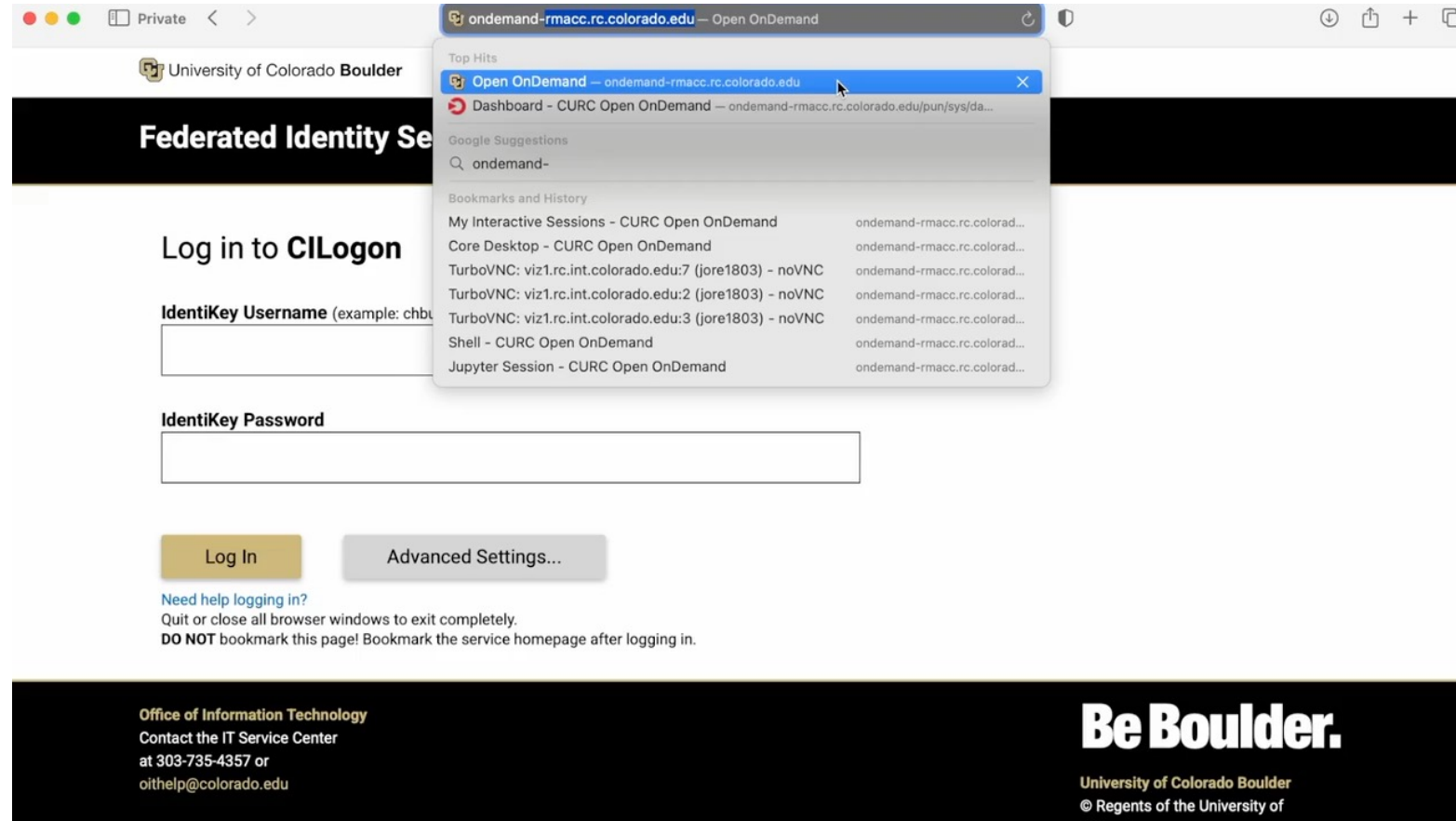CURC Open OnDemand is a browser based, integrated, single access point for all of your HPC resources at CU Research Computing.

- CU Boulder: Visit https://ondemand.rc.colorado.edu.

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

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Overview   Nodes   Data   **Login**

Be Boulder.

Demo: logging in with OnDemand

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/

- **Helpdesk**: rc-help@colorado.edu

- **Consult Hours** (Tuesday 12:00-1:00 in-person or virtually, Thursday 1:00-2:00 virtually)

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Help Desk Tickets

To: rc-help@colorado.edu

Dear Research Computing,

Help! My code won't run!
Help!

Help please,
John

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 1620350. Let me know what I can try!

sbatch: error: Batch job submission failed: Invalid partition name specified.

Thanks,
John

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.