



# Transferring Your Data



Research Computing  
UNIVERSITY OF COLORADO **BOULDER**

**Be Boulder.**

---

Brandon Reyes

- *email: rc-help@colorado.edu*
- *RC Homepage: <https://www.colorado.edu/rc>*

Slides available on GitHub:

[Summer\\_Camp\\_2023/Day\\_Two/Transferring\\_your\\_data/slides](#)

# Outline

---

- Ways to access your data
- Data transfer using the command line
- Data transfer using Open OnDemand
- Data transfer using Globus
- Sharing Data
- Getting A Petalibrary Allocation

# Accessing Data on RC Resources

---

- When you use RC resources the data is not on your local machine
- Ways to access the data from your local machine
  - Command line (a variety of tools)
  - Open OnDemand (straightforward GUI interface)
  - Globus (GUI interface with some set up required)

# Access through the Command Line

---

- If you don't need a *fancy* GUI
- Provides a larger variety of tools
  - SCP
  - SFTP
  - RSYNC
  - RCLONE
  - SSHFS
  - SMB
- The tools provided can improve your data workflow (more on this later)



# General Filesystem Structure

## /home (2GB)

- Small important data
- Backed up frequently
- Not for sharing files or job output

## /projects (250GB)

- Medium sized important data
- Software
- Can be shared with others
- Backed up, but less frequently
- Not for job output

## /scratch/alpine (10TB)

- Large data
- Can be shared with others
- Fast Data transfer to compute nodes
- Not backed up!
- Purged after 90 days!

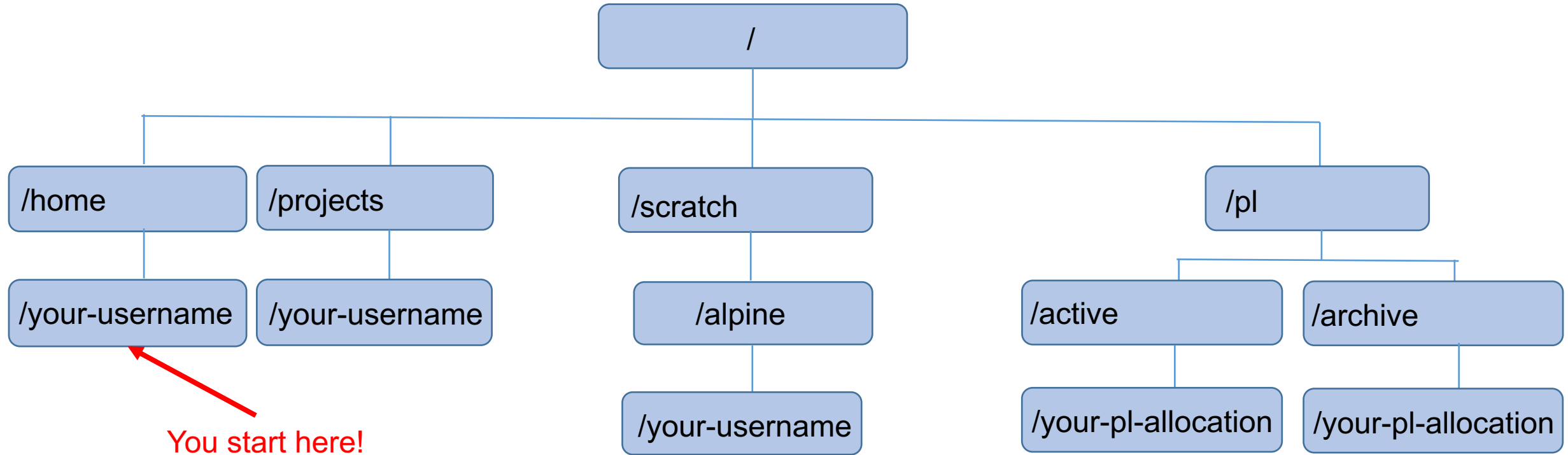
Filesystem documentation: <https://curc.readthedocs.io/en/latest/compute/filesystems.html>

# Let's get on a login node!

---

```
ssh <your-username>@login.rc.colorado.edu
```

# RC Filesystem Map





# Basic Navigation Commands

---

- Change directories

```
cd <relative-or-full-path>
```

- List contents of a directory

```
ls <optional-path>
```

- Print current working directory

```
pwd
```

# RC endpoints

---

Endpoint – one of the two file transfer locations i.e., it is either the source or the destination we want to copy data from or to.

- For data on RC resources, we have two endpoints

- The **login\*** nodes

- Only use for small transfers!!

`<your-username>@login.rc.colorado.edu`

- Data transfer nodes (DTNs)

`<your-username>@dtn.rc.int.colorado.edu`

- CSU

`<your-username>@dtn.rc.colorado.edu`

# RC Data transfer nodes (DTNs)

---

- Command line use of DTNs only available if you are on CU Boulder or CSU's network or VPN
- Dedicated nodes for transferring data
  - Faster transfers
  - More stable transfers
- Suitable for
  - Large and frequent transfers
  - Automated (passwordless) transfers
    - Only for CU Boulder folks
- Cannot ssh into the DTNs!

# Command line option - SCP

---

SCP (Secure Copy Protocol) is a command line tool to transfer files/directories to, from, or between remote locations.

- Simple, but useful!
- Copying a local file to RC resources using a login node:

```
scp file1 <username>@login.rc.colorado.edu:<remote-path>
```

- Copying a directory from RC resources to local path via a DTN:

```
scp -r <username>@dtn.rc.int.colorado.edu:<path-to-directory> <local-path>
```

# Command line option - SFTP

---

SFTP (Secure File Transfer Protocol) a command line tool that is similar to SCP, but provides an sftp session where both the local and remote filesystems are available

- Slightly more advanced than SCP
- Useful for multiple file/directory transfers
- Starting a SFTP session on a local machine

```
sftp <username>@login.rc.colorado.edu
```

- Demo time!

# Command line option - Rsync

---

Rsync (remote sync) a command line tool that offers remote and local file synchronization.

- Only copies the portion of the files that have changed!
- Already installed on most Linux distributions and macOS
  - Needs to be installed on Windows
- Sync RC resources to local computer

```
rsync -av <username>@login.rc.colorado.edu:<remote-path> <local-path>
```

- Flags:

- v    # verbose mode
  - a    # archive mode

# Command line option - Rclone

---

Rclone is a command line tool used to manage files on cloud storage.

- It is compatible with all major cloud storage solutions
  - Supported by over 40 cloud storage products!
- Created as a cloud equivalent to the UNIX commands:
  - rsync, cp, mv, mount, ls, ncdu, tree, rm, and cat
- Needs to be downloaded on your local machine
- Requires a more involved setup process but works great!
  - <https://curc.readthedocs.io/en/latest/compute/data-transfer.html#rclone>

```
rclone copy rclonetest.csv aws_s3:testbucket/
```



# Command line option - mounting

---

Mounting is the process of attaching a file system to a directory on another system.

- SSHFS (secure shell filesystem)
  - Needs to be installed on Mac and Windows (available on most Linux distributions)
  - You need to be on the campus network or VPN!

```
sshfs <username>@login.rc.colorado.edu:<path> <local-mountpoint>
```

- SMB (server message block)
  - Built into all major operating systems
  - You need to be on the campus network or VPN!
  - Contact us if you want to use this

---

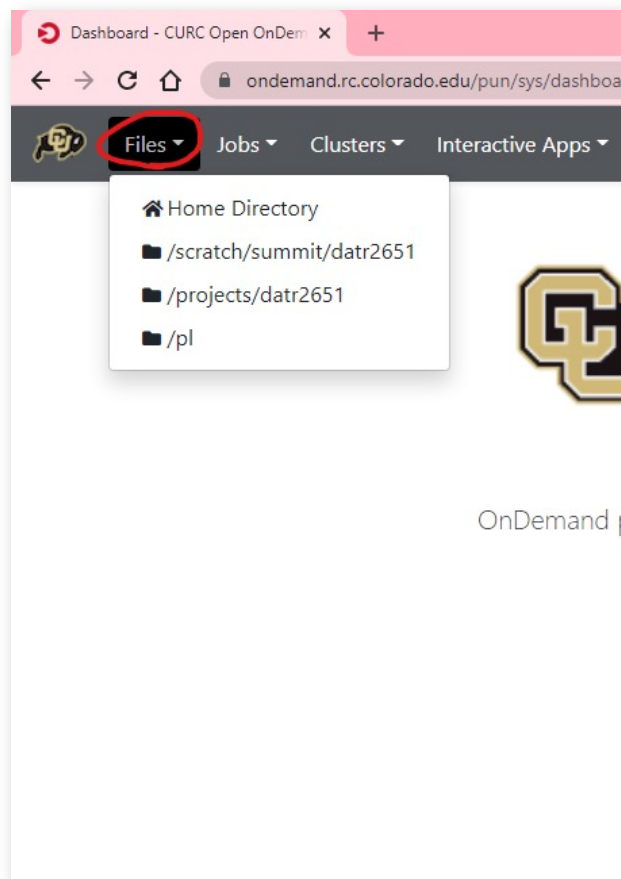
# GUI based options

# GUI option - Open OnDemand

---

- No command line required!
  - <http://ondemand.rc.colorado.edu/>
  - <http://ondemand-rmacc.rc.colorado.edu/>
- File management
  - Create, Delete, Move, and Rename
- File transfers
  - Upload and Download





Open in Terminal New File New Directory Upload Download Copy/Move Delete

/ projects / datr2651 / Change directory Copy path

☐ Show Owner/Mode ☐ Show Dotfiles Filter:

Showing 7 of 11 rows - 0 rows selected

	Type	↑ ↓ Name	↑ ↓ Size	↑ ↓ Modified at
<input type="checkbox"/>	Folder	bench	-	12/16/2021 4:46:14 PM
<input type="checkbox"/>	Folder	mana	-	12/15/2020 10:13:20 AM
<input type="checkbox"/>	Folder	private	-	8/30/2020 2:51:51 PM
<input type="checkbox"/>	Folder	public	-	12/10/2021 3:19:48 PM
<input type="checkbox"/>	Folder	scripts	-	2/9/2022 2:38:58 PM
<input type="checkbox"/>	Folder	software	-	8/10/2021 1:21:33 PM
<input type="checkbox"/>	File	NAMD_2.14_Source.tar.gz	55.1 MB	1/12/2022 3:41:54 PM

# Let's take a look!

# GUI option - Globus

---

Globus is a service that allows for users to reliably move, share, and discover data

- Command line version is also available
- Our recommended way to transfer data
  - Stable and fast data transfers
  - Transfers continue if a user disconnects
  - Web GUI or Globus Connect Personal GUI
- Supported on all major operating systems
  - Works well with cloud storage providers

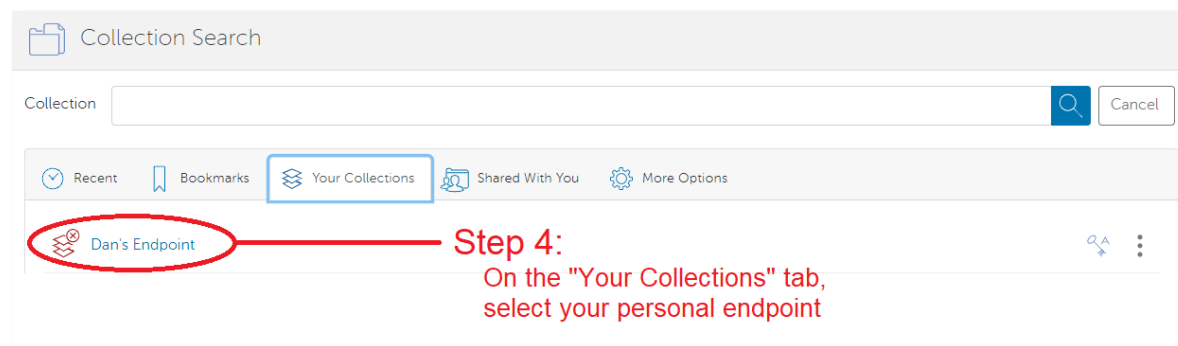
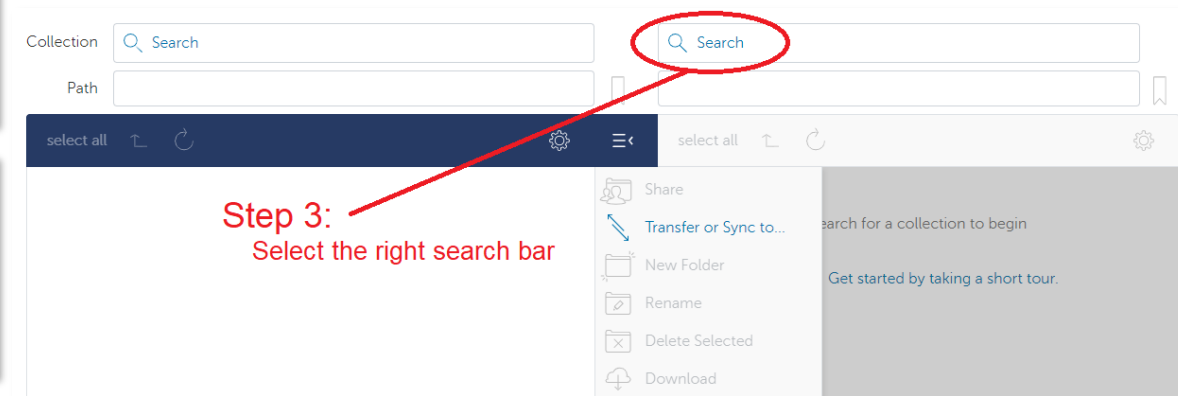
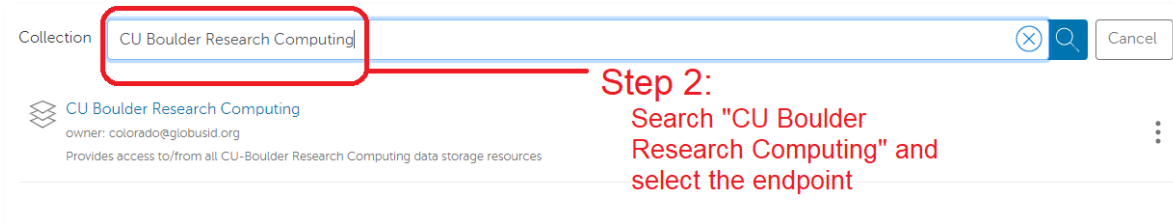
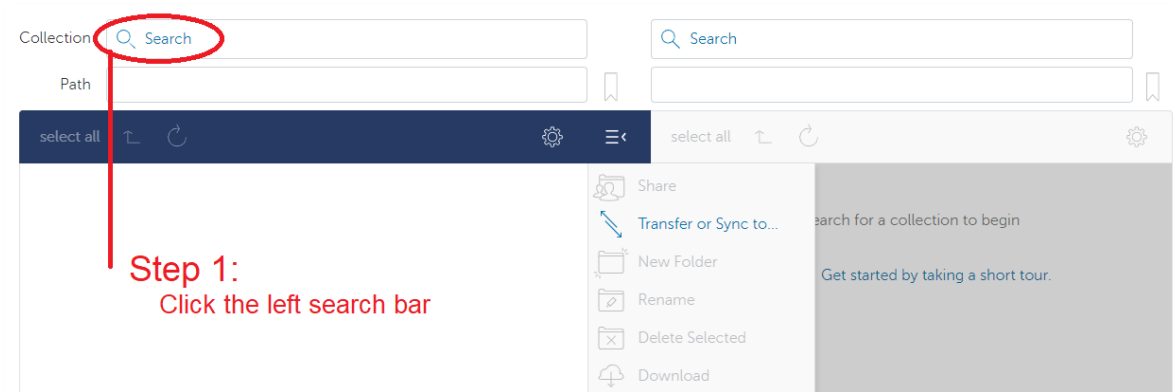


# Globus Demo

---

- Globus login is simple and quick: <https://app.globus.org>
  1. CU Boulder users - select “University of Colorado at Boulder” in the dropdown menu
    - Other institutions should select “ACCESS”
  2. Login with your credentials
  3. Continue with onscreen prompts until you are brought to the Globus WebGUI
- Installing a Globus Endpoint on your local machine
  - Required if you want to transfer data to your machine
  - Navigate to <https://www.globus.org/globus-connect-personal>
    - Click on operating system specific version and follow install instructions





# Let's check it out!

# The PetaLibrary

---

The PetaLibrary is a CU Boulder Research Computing service

- Expands the amount of storage space available to you
  - Confidential data should not be stored on PetaLibrary!!
- Aims to work seamlessly with all RC resources
- Supports the storage, archival, and sharing of data
- Available at a subsidized cost for researchers affiliated with University of Colorado
- New customer's initial upper limit:
  - 200 TB for Active storage (available to compute resources)
  - 100 TB for Archive storage (**not** available to compute resources)

# Sharing Data

---

- RC Users on RC resources
  - Send a request and a list of the users to [rc-help@colorado.edu](mailto:rc-help@colorado.edu)
    - RC will place the chosen users in your Linux group
      - Allows them to see your scratch and project directories
      - You can set permissions in the space, so items are hidden
  - On-premise collaborators can also access Petalibrary files with Globus Shared Endpoints
- Off-premise collaborators
  - Data sharing is only available if you have a PetaLibrary allocation
    - Data transfer is done through Globus Shared Endpoints

# Globus Shared Endpoints

---

- Globus offers ‘shared endpoints’ which don’t require a user to have an account with RC.
- RC provides this capability for easy access of Data.
- PetaLibrary exclusive!
- Generates a shared collection that can be accessed with a link.
  - See <https://scholar.colorado.edu/concern/datasets/9593tw13k>
  - Can assign various permissions to specific users or all users withing Globus
  - More information on here: <https://docs.globus.org/how-to/share-files/>

# Thank you!!

For more help contact [rc-help@colorado.edu](mailto:rc-help@colorado.edu)

Additional documentation: <https://curc.readthedocs.io/en/latest/compute/data-transfer.html>

# Unix Groups

---

- Unix Groups
  - 3 Levels of permissions:
    - User
    - Group
    - Other
  - All users have a group associated with their username
  - Permissions can be set for an individual file with the `chmod` command

```
chmod g+rx file.exe
```