



Research Storage Simplified

John Reiland

View the Slides



https://github.com/ResearchComputing/research_storage_simplified

Meet the User Support Team



Layla
Freeborn



Brandon
Reyes



Andy
Monaghan



Michael
Schneider



John
Reiland



Dylan
Gottlieb



Mohal
Khandelwal



Ragan
Lee

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, Active+Archive, Archive+DR
- 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

More Details on Scratch

/alpine

- Available from all nodes
- Accessible even when no jobs are running

/local

- Specific to each compute node
 - Accessible only on that node, while a job is running on it

PetaLibrary

Active

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



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

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.

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.

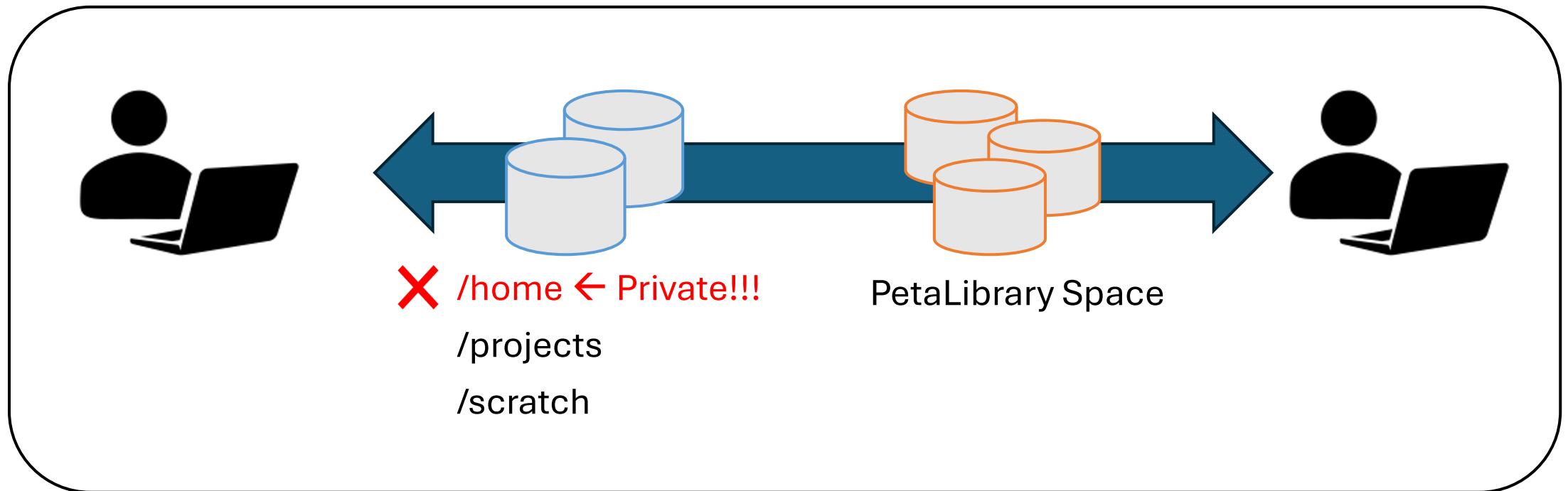
Archive + DR

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

Speed Comparison

- **/scratch/local: Fastest**
- **/scratch/alpine: 2nd Fastest**
- **PetaLibrary (all tiers): 3rd Fastest**
- **/projects: 4th Fastest**
- **/home: Slowest**

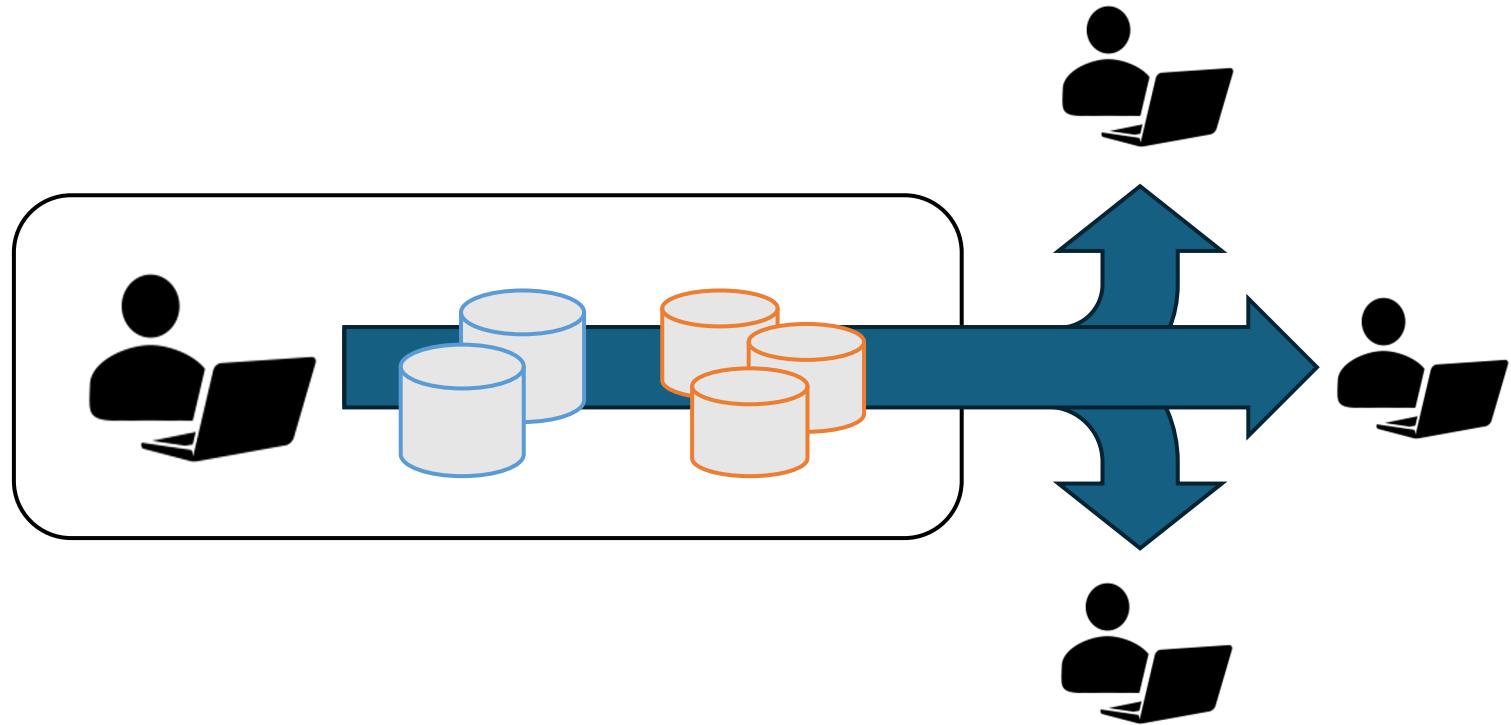
Data Sharing: Within RC



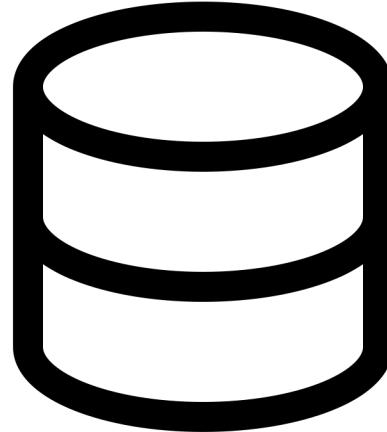
Data Sharing: Outside RC

Large Data Transfers:

- Globus (Recommended)
- Data Transfer Nodes (DTN)
- Terminal/Command Line:
 - rsync
 - rclone (connect to 3rd party cloud services like OneDrive)
 - sftp
 - scp



Data Transfer Nodes



DATA (DTN)

- Support data transfers
- Most robust data transfer method
- 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>

Documentation



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

Survey and feedback



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