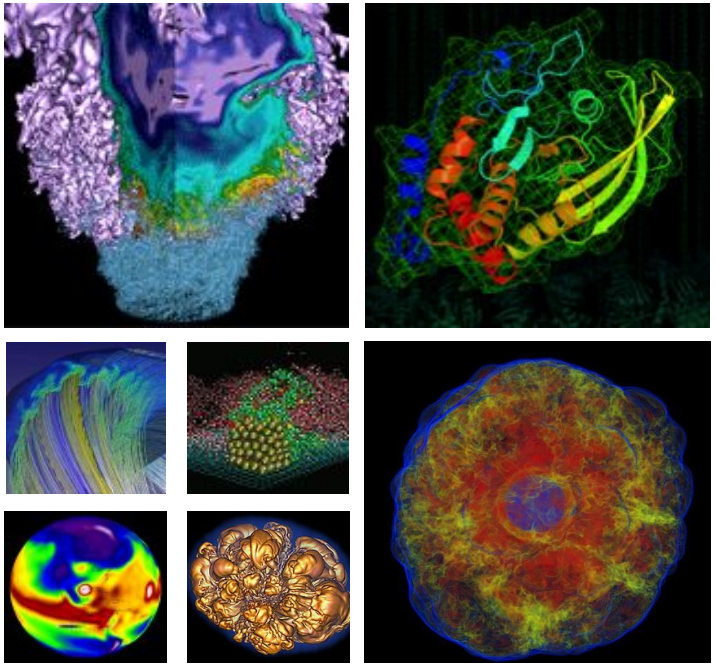
The Community File System (CFS)



NERSC

CFS Team

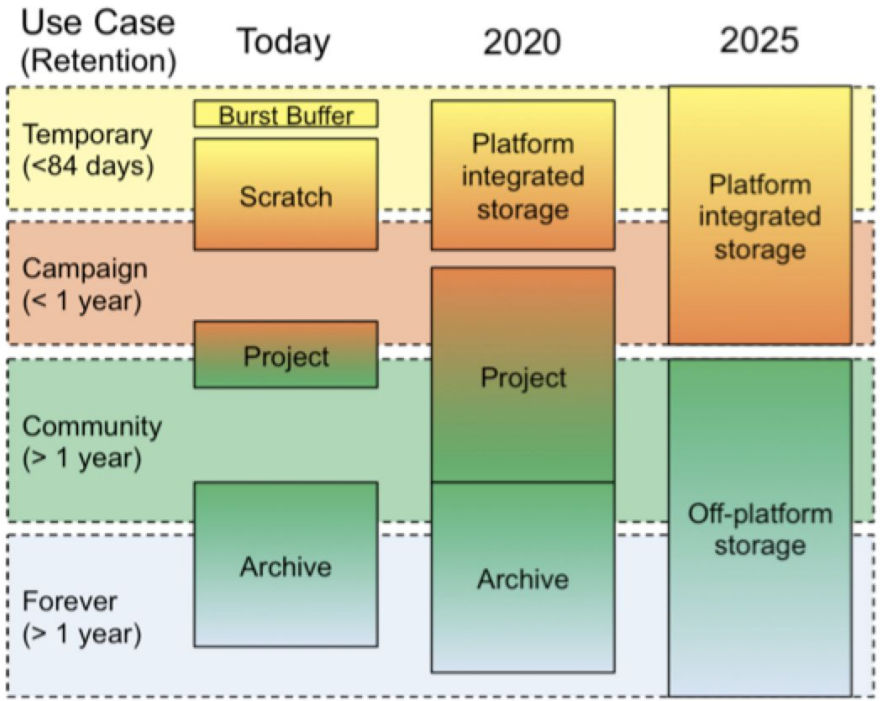
December 12, 2019

U.S. DEPARTMENT OF ENERGY | Office of Science

BERKELEY LAB

Community File System Purpose

- Essential part of NERSC's storage plan
- Layer between scratch and HPSS for storage and sharing of long-lived scientific data
  - Inside (group permissions) and outside NERSC (portals and parallel data transfers)
- This will replace /project (and /projecta eventually)



What's new with CFS?

- Space!
  - An order of magnitude increase in space to ~60PB, will increase to 200PB over the next five years
  - Default quota for directories will go from 1TB to 20TB (and 20M inodes)
- Better quota management for subprojects in a repo
  - PIs can request separate directories and have *individual* quotas for them
- New allocation model
  - Quotas for each repo are granted by DOE allocations managers as part of the ERCAP process
- File system features
  - Faster rebuilds from distributed raid
  - End to end checksums (client to servers to ensure data integrity)
  - Subblocks allow more efficient use of capacity especially for small files

What's the same with CFS?

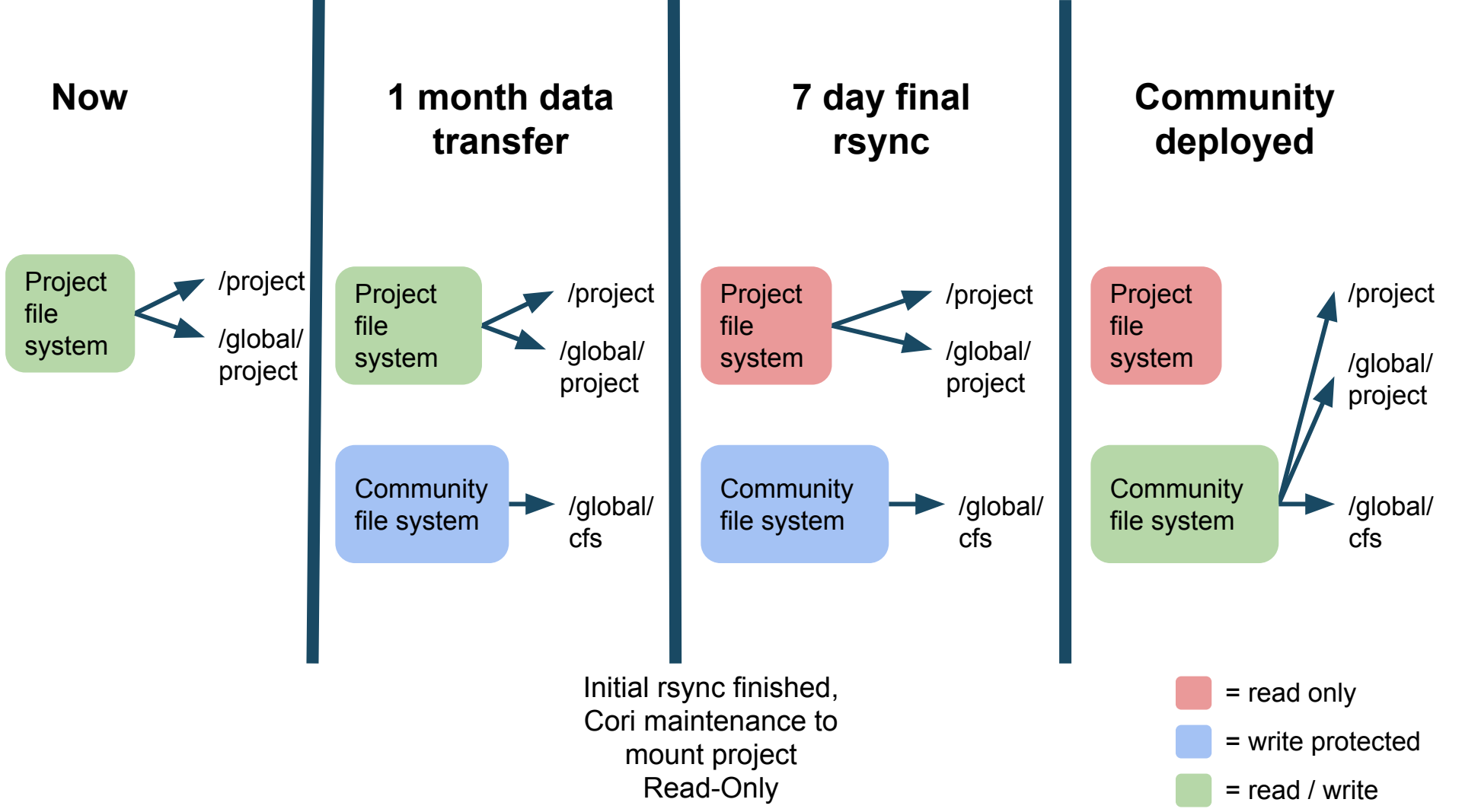
- Every repo will have a directory of the same name by default
  - Created automatically for new repos
  - Repos can request multiple directories
- Group read and writable permissions
  - Owned by PI and RW by linux group of the same name as the repo

    dtn10.nersc.gov> ls -ld /global/cfs/cdirs/nstaff

    drwxrws--x 30 sudip nstaff 131072 Aug 8 10:54 /global/cfs/cdirs/nstaff
  - Group permissions are the same as on the Project File System
- Mounted on every system
- Retirement policy: Directories from inactive repos are migrated to HPSS after 1 year
- CFS will not be purged and has 7 days backup capability from snapshots

Deployment Details

- NERSC staff will migrate your data from the Project File System
- Data migration is going on now
- The Project File System will be set to read-only from January 14, 2020 to January 21, 2020 for the final sync of data
  - Please reach out to NERSC consulting if this will cause major hardship for your group
- /global/cfs/cdirs/<repo_name> will continue to be accessible as /global/project/projectdirs/<repo_name> until mid-2020
  - Services in Spin, science gateways, and scripts will continue to work
  - NERSC dot files will have a new shell variable “CFS”, which you can use with ‘cd $CFS/repo1’
  - We recommend you migrate your scripts to using the new path


Now
1 month data transfer
7 day final rsync
Community deployed
Project file system
/project
/global/ project
Community file system
/global/ cfs
Initial rsync finished, Cori maintenance to mount project Read-Only
= read only
= write protected
= read / write

Sponsored Storage

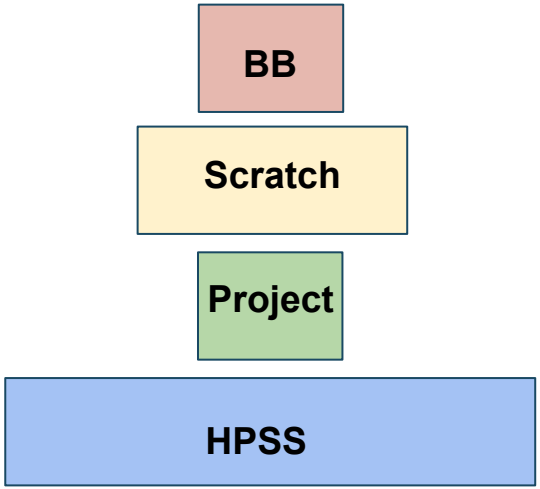
- Existing sponsored storage purchases will be honored
  - At a later date, they may be migrated onto the community file system
- New threshold for purchases will be 1 PB
  - New purchases will be on CFS
- Only buy twice a year in January and June to help with consolidating purchases

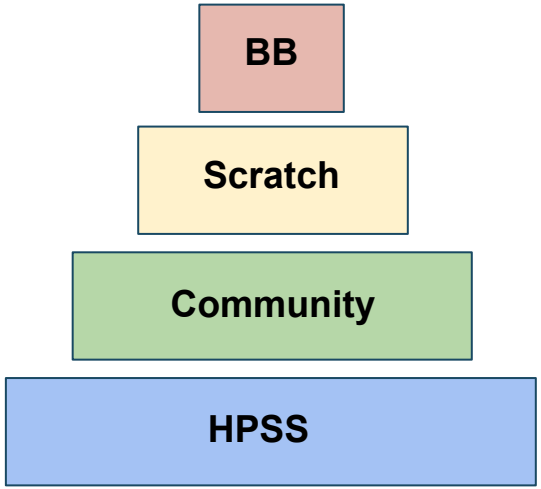
CFS Usage

- CFS is designed for capacity; for optimal I/O performance continue to use either Cori scratch or the Burst Buffer.
- NERSC has command line data transfer tools to facilitate this:

https://docs.nersc.gov/services/globus/#command-line-globus-transfers-at-nersc

A New Order





Thank You

Tunable Quotas

A long requested desire to have individual quotas for separate working groups or purposes

Example: ACME wants three different directories. Total Quota (from ERCAP) is 100TB

1. **/global/cfs/cdirs/e3sm-inputdata: quota 50TB**
2. **/global/cfs/cdirs/e3sm-performance-data: quota 30TB**
3. **/global/cfs/cdirs/e3sm: quota 20TB**

Split will be controllable by PIs and PI Proxies via Iris

Limit of 5 per repo

The default 20TB will apply to all directories a repo has. No printing money!