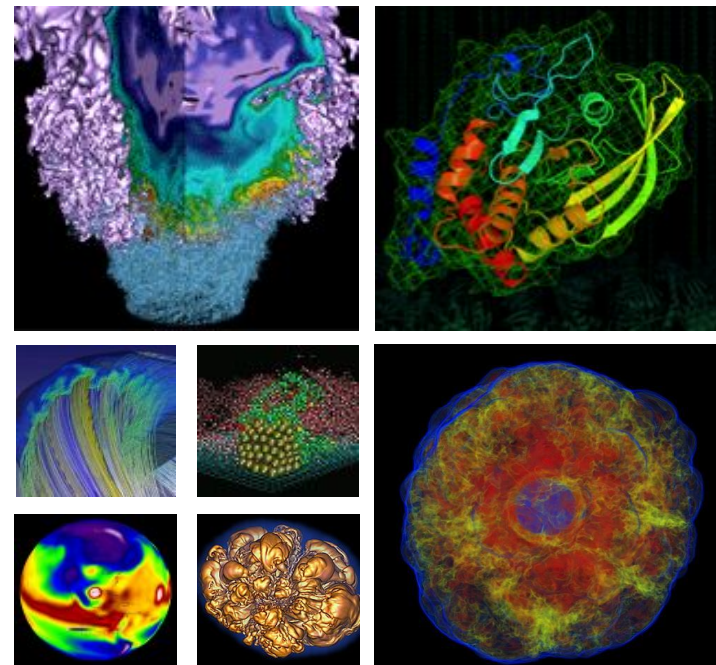


Data Sharing at NERSC



Annette Greiner

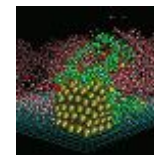
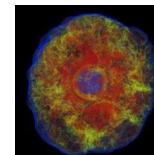
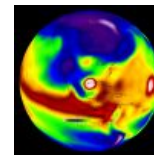
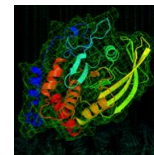
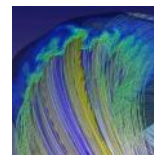
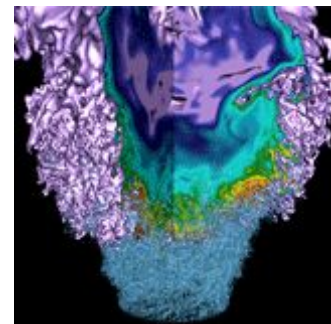
**NERSC Data and Analytics Services
NUG June 2022 Monthly Meeting**

- **Notes on permissions**

- **Internal Data Sharing**
 - **Within a project**
 - **Outside a project**

- **External Data Sharing**
 - **Science gateways**
 - **Data transfers**

Notes on Permissions



User (u)

- Only the owning user can change the perms.

Group (g)

- Default group on NERSC systems named as username.
- A user can be added to many groups.

Other (o)

- Others on the system may or may not be collaborators.

Mode

- `rx` and friends
- `+x` on a dir enables entering, `+r` on a dir enables `ls`

Least Privilege



Private files should be set aside in a dir with

g-rwx

o-rwx

Remember Paths



Don't open the door but close the road.

`o+x` for directories in the path, allows traversal

`o+rx` allows navigation

Setgid on a directory



Preserves group ownership

Helps keep shared dirs accessible

g+s

We also provide a script to fix permissions on a dir

```
make-dir-group-writable -g groupname dir ...
```

Your umask (along with system default permissions) controls default permissions for files you create.

Default umask at NERSC is 007.

```
-rw-rw----
```

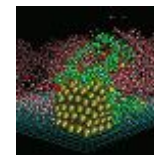
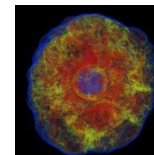
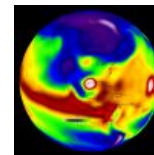
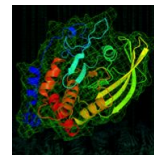
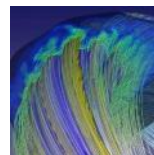
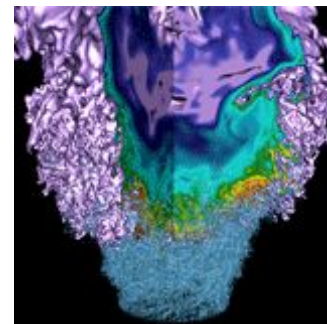
```
drwxrwx---
```


Permissions in docs.nersc.gov



<https://docs.nersc.gov/filesystems/unix-file-permissions/>

Sharing Within NERSC



Sharing Within a Project



- Permissions
- Project unix group
- Shared project dir on CFS
- PIs can use
 - [PI Toolbox](#) for changing perms in CFS
 - [Iris](#) for creating/managing unix groups

Sharing Outside a Project



Internal to NERSC, but across projects

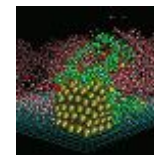
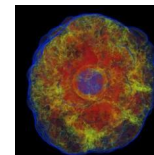
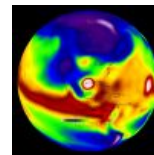
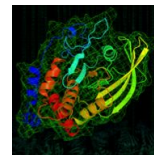
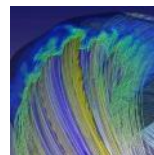
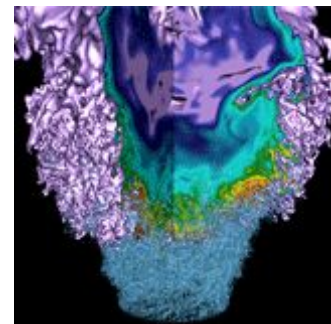
- **Permissions**
- **give/take**

```
give -u recipient filename
```

```
take -u giver -a
```

```
take -u giver -d newlocation filename
```

Sharing Outside NERSC



File Transfers



- **Globus (recommended)**
- **DTNs**
- **scp**
- **bbcp**
- **FTP (not recommended)**



<https://docs.nersc.gov/services/globus/>

Globus Guest Collections



- Formerly “Globus Sharing”
- Enables read-only sharing with unauthenticated users
- Create a dir, file a ticket, add endpoint at globus.org

/global/cfs/cdirs/<myproject>/gsharing

```
nerscuser@cori03>ls  
about-logs.051715.pdf  
about-mutrino1yr-v122016.pdf  
logs.051715.cr.tgz  
mutrino1yr-v122016.tgz
```



The screenshot shows the Globus File Manager interface. At the top, there is a notification: "Globus services will be unavailable for up to 1 hour beginning at 9:00 AM US Central on Saturday, planned maintenance." Below this, the "File Manager" section is visible. The "Collection" is set to "HMDR Datasets" and the "Path" is "/". A "Start" button is present. A list of files is displayed:

File Name	Created	Size
about-logs.051715.pdf	4/10/2019, 04:13 PM	371.79 KB
about-mutrino1yr-v122016.pdf	4/10/2019, 04:13 PM	79.93 KB
logs.051715.cr.tgz	4/10/2019, 04:14 PM	679.17 MB
mutrino1yr-v122016.tgz		

On the right side, a context menu is open with options: Permissions, Transfer or Sync to..., New Folder, Rename, Delete Selected, Download, Open, and Upload.

- **Web-based scientific resources for collaboration**
- **A form of computer-supported cooperative work**
- **Can be shared data, shared tools, shared services**
- **“value-added interfaces to access . . . shared resources”
(sciencegateways.org)**

Two major approaches:

- **Share a /www directory on CFS or HPSS**
 - HTML and js, file downloads, client-side programming
 - Can use js APIs like Superfacility API
 - No server-side framework
 - 8 x 5 support
- **Build a service in Spin**
 - Any server-side framework, custom domain name
 - Connect directly to DBs, CFS, other NERSC resources
 - Can use js APIs like Superfacility API
 - 99%-uptime infrastructure, self-administered services

Sharing a /www Directory

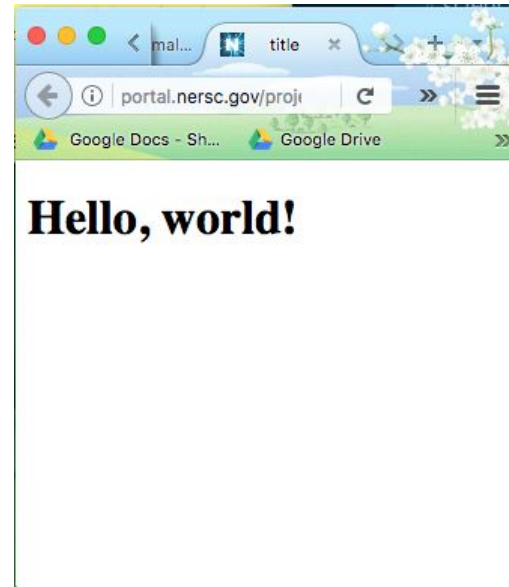


World-readable files in /global/cfs/cdirs/<myproject>/www

/global/cfs/cdirs/myproject/www/index.html

http://portal.nersc.gov/project/myproject/index.html

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <title>title</title>
  </head>
  <body>
    <h1>Hello, world!</h1>
  </body>
</html>
```



Spin



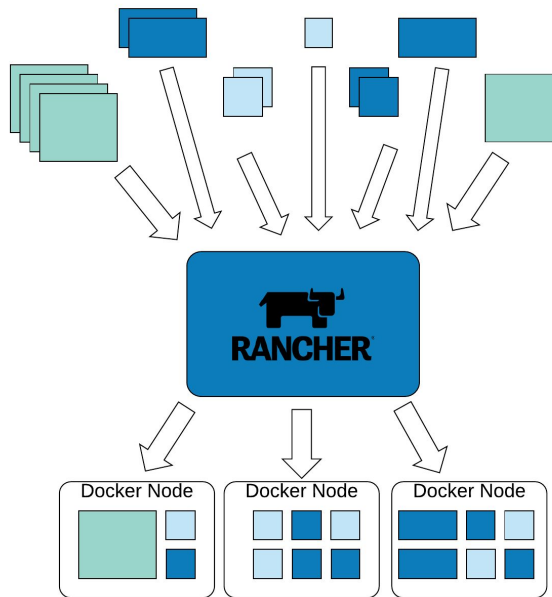
Spin is NERSC's cloud platform for container-based services

Based on Rancher 2

Kubernetes under the hood

- **Build** x86 image on a laptop with Docker
- **Ship** with DockerHub or NERSC registry
- **Run** with Rancher

Without orchestration, a pool of servers and no coordination for users



Managed and assigned to Docker nodes, enabling holistic management, failover, service ownership.

Spin is for apps, not compute

- **database-backed web apps that access project data**
- **workflow orchestration tools running outside of HPC**
- **API servers for real-time or distributed projects**
- **or something else!**

In-depth documentation at <https://docs.nersc.gov/services/spin/>



Spin Training Workshops



All workshops delivered over Zoom. Registration required! <https://www.nersc.gov/users/training/spin/>

June

- Seminar: Wed Jun 22, 2022, 9:00a - 12:00p
- Hack-a-thon; choose either
 - Option A: Wed Jun 29, 9:00a - 12:00p
 - Option B: Thu Jun 30, 1:00p - 4:00p

August

- Seminar: Wed Aug 10, 2022, 9:00a - 12:00p
- Hack-a-thon; choose either
 - Option A: Wed Aug 17, 9:00a - 12:00p
 - Option B: Thu Aug 18, 1:00p - 4:00p

October

- Seminar: Wed Oct 5, 2022, 9:00a - 12:00p
- Hack-a-thon; choose either
 - Option A: Wed Oct 12, 9:00a - 12:00p
 - Option B: Thu Oct 13, 1:00p - 4:00p

November

- Seminar: Wed Nov 30, 2022, 9:00a - 12:00p
- Hack-a-thon; choose either
 - Option A: Wed Dec 7, 9:00a - 12:00p
 - Option B: Thu Dec 8, 1:00p - 4:00p



Thank You



U.S. DEPARTMENT OF
ENERGY

Office of
Science

