Sharing Data at NERSC
Sharing Data Internally
The /project directory

- Your project will have a shared project directory at /global/project/projectdirs/<dirname>
- Shared group permissions for your project repo
- https://www.nersc.gov/users/storage-and-file-systems/sharing-data/
Give/take

• Give a file to a user:
  – give -u <recipient-username> <file-or-directory>
  – This sends is into a holding area for pickup and notifies the recipient over email

• Accept a file from a user
  – take -u <sender-username> (-d <destination folder>) <filename>
  – Picks up the file (use –a for all files)
Sharing Data Externally
NERSC Science Gateways

- Web portals that allow you to interface with your data and computation at NERSC
- Interfaces built around your science
- Science-As-A-Service
Services

• Simple data publishing capabilities
• Rich web interfaces and complex portals
• Backend databases and message queues
• NEWT API to access NERSC resources
• Virtual machines and “designer” URLs
Publish Data On the Web

• Every repo now has a project directory
• Trivial to wire up your project directory to make it accessible over the web
• Create a file in your www directory
  – mkdir /global/project/projectdirs/<yourproj>/www
  – cd /global/project/projectdirs/<yourproj>/www
  – vi index.html
    <html>Hello World</html>
• Make sure all the above files and directories are world readable
  – chmod 775 /global/project/projectdir/<yourproj>/ etc.
• Voila:
  – http://portal.nersc.gov/project/<yourproj>/
Build Full Stack Web Applications

• Build full stack web applications for your science at NERSC
  – Python/Django, PHP, Ruby on Rails, Java Backends
  – JavaScript + AJAX Frontends

• Databases
  – MongoDB, MySQL, PostGreSQL, SciDB
  – http://tinyurl.com/nerscdfs

• Public or Authenticated Gateways

• OpenDAP and MQ services
Some Examples

- http://materialsproject.org
- https://spot.nersc.gov
- https://openmsi.nersc.gov
- https://portal-auth.nersc.gov/atc
Use NEWT API to access NERSC

• NEWT – the NERSC REST API
• Use the NEWT HTTP API to access NERSC HPC resources directly from your web apps.
Basic NEWT Example

- HTTP verb + URL returns structured JSON data

eg.

GET https://newt.nersc.gov/newt/status/

[{
  "status": "up",
  "system": "hopper"
},
{
  "status": "up",
  "system": "carver"
},
{
  "status": "up",
  "system": "edison"
},
{
  "status": "up",
  "system": "pdsf"
},
{
  "status": "up",
  "system": "genepool"
},
{
  "status": "up",
  "system": "archive"
}]
NEWT Features

- Run Commands on any system
- Transfer files
- Authentication
- Submit/Query Jobs directly through NEWT
- Persistent Store
  - Store JSON objects in the NEWT storage
- Access to NIM info
  - Information about Users, Repos etc.
- System Status
NEWT demos

• See https://newt.nersc.gov/ for documentation and live demos
Virtual Machines and Custom URLs

- We typically build portals on shared hosts:
  - portal.nersc.gov and portal-auth.nersc.gov

- We can add designer URLs that point specifically gateway eg.
  - http://cxidb.org/
  - http://deepskyproject.org/
  - https://openmsi.nersc.gov/

- If you need dedicated resources for your project we can also allocate Virtual Machines to host your gateways
Engagement

• There are various levels of engagement and collaboration
  – NERSC provides building blocks and backend infrastructure, science groups build their own gateways.
  – Immersive development - science groups work in collaboration with NERSC engineers to build gateways. (But requires more resources from both sides).

• If you are interested in building a portal please come and talk to us. We can help customize our offerings to meet user needs.
Other tools for Sharing
Globus

- Useful for both sending and receiving data
- Future: Globus Sharing to enable access for non-NERSC users
Anonymous FTP Upload Service

- For external collaborators to be able to send data to NERSC users
- Create a temporary FTP account to upload files that will be delivered to a NERSC user
- NERSC user can retrieve the file with "take -u ftpup <file_name>".