

Globus @ NERSC

Shreyas Cholia
scholia@lbl.gov
NERSC



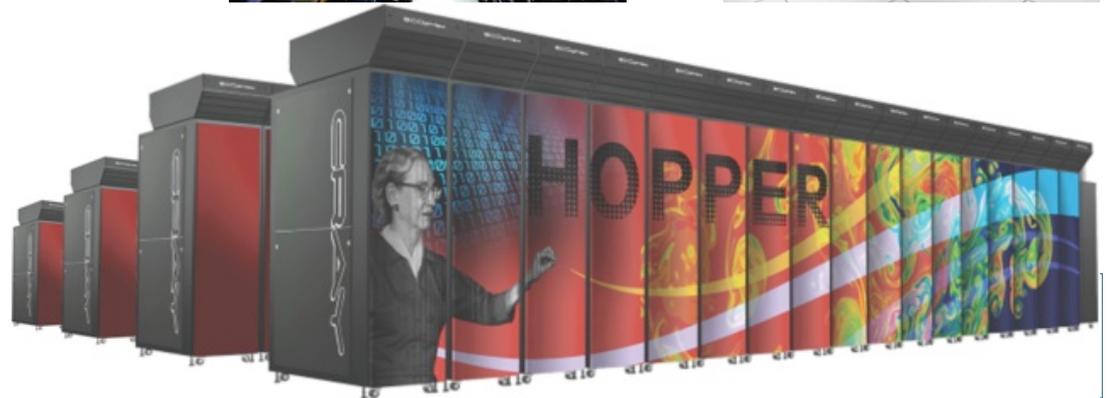
Powering **scientific discovery** since 1974.

- **National Energy Research Scientific Computing Center is a DOE Office of Science User Facility at Lawrence Berkeley Lab**
- **Accelerate the pace of scientific discovery by providing high-performance computing, information, data, and communications services.**



NERSC Resources

- **Multiple compute clusters**
 - Hopper, Franklin, Carver, Euclid, PDSF
- **HPSS archival storage system**
- **Global File System**



U.S. DEPARTMENT OF
ENERGY



Broad Range of Computing Needs

- **More than 3000 users, 400 projects, 500 codes**
- **Focus on unique resources**
 - **Different systems to solve different classes of problems**
 - **Highly Parallel HPC, Mid Range, Data Parallel, Serial**
- **Science driven**
 - **Real science problems used in machine procurements and metrics**
 - **Services For Science**



Grid Infrastructure

- **NERSC is part of OSG**
 - **OSG/VDT software stack**
- **Globus interfaces on all major systems**
 - **GridFTP**
 - **Globus Gatekeeper**
- **Dedicated Data Transfer Nodes for High Performance WAN Transfers**
- **Mass Storage (HPSS) with GridFTP interface**
- **NERSC Online CA to simplify certificate management**
- **NERSC Globus Online Endpoints**
- **NEWT – REST API to Access NERSC resources**



Challenges

- **Ease of use**
- **Performance Tuning**
- **Certificate Management**
- **Reliable Data Movement**



Then ... and Now

- **Users were using a combination of scp, bbcp, globus-url-copy, SRM**
- **Heavy admin assistance and intervention for large transfers**

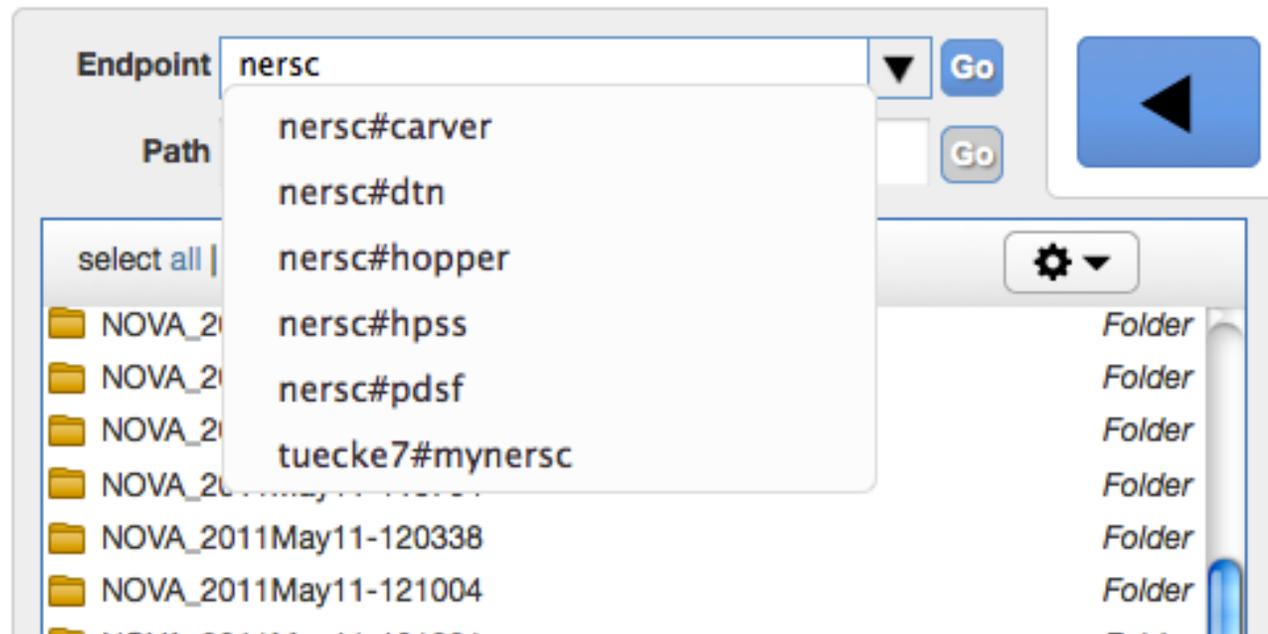
- **Enter Globus Online**



Globus Online Endpoints

- All NERSC systems supported:

Transfer Files





NERSC CA

- **Users login with LDAP password to get a cert**
- **We manage our own CA, and all users are automatically added to grid-mapfiles**
- **NERSC CA is IGTF accredited and part of the distributed CA bundle**



Globus Online Demo

- **Drag and Drop Archiving to/from HPSS is finally ready!**
- **Globus Online makes it very easy to move data across sites**
- **Globus Connect has been an enormous help since it doesn't require infrastructure on the other side**



It Works

- **Globus Online is now the officially recommended way to get data in and out of NERSC**
- **Number of tickets has gone down significantly!**
 - **People are just using this transparently without intervention.**
- **Multitple Communities**
 - **Climate, Astrophysics, Materials Science etc.**



Features We Like

- **Different Certs For Endpoints**
 - Don't have to worry about cross site trust
- **Globus Connect**
 - No need for users to manage GridFTP servers
- **Tight Coupling with NERSC CA Identities**
 - Allows our users to login with NERSC username/
password. No messing with certs
- **Detailed Logging and Error Reporting**
- **Clean UI**
- **HPSS Support**



NEWT

- **Nice and Easy Web Toolkit**
- **REST JSON API to access NERSC resources**

GET <https://portal-auth.nersc.gov/newt/queue/hopper/>

POST <https://portal-auth.nersc.gov/newt/queue/hopper/>

GET <https://portal-auth.nersc.gov/newt/file/datatran/>

- **NEWT uses globus under the covers**
- **Offload data movement to GO?**
- **eg. NERSC Mobile: <http://m.nersc.gov>**



Wishlist

- **Administrator tools for GO**
 - Ability to throttle or kill rogue jobs
 - Per User access controls for endpoint
 - Configurable # of streams, buffers
- **Non GSI based GO/GridFTP transfers**
 - OAuth 2.0 delegation
- **Integrate GO with collaboration web portals**
- **Firewall traversal**