DATA TRANSFER WORKING TEAM

Lukasz Lacinski, DOE/ANL
lukasz@uchicago.edu
Rachana Ananthakrishnan, DOE/ANL
ranantha@uchicago.edu

ESGF installer pre-2.1

- IdP node:
  - MyProxy server with PAM module to authenticate users against ESGF database
- Data node:
  - End user and Bulk Data Movement GridFTP servers
  - Custom authorization callouts and modifications to use ESGF authorization infrastructure
- Everything was installed from source code – C, Java (about 1 hour)

Globus Connect Server

- Globus Connect Server is a set of scripts and configuration files that make installing and configuring GridFTP servers, MyProxy, MyProxy OAuth server easy, and set it up to work with Globus (globus.org)
  - Choose components that need to be installed
- RPMs:
  - globus-connect-server
  - globus-connect-server-id, globus-connect-server-io, etc.
  - myproxy-server, globus-gridftp-server, etc.

Work done for 2.1

- IdP node:
  - Globus Connect Server MyProxy component only with PAM module to authenticate users against ESGF database
- Data node:
  - Globus Connect Server IO (GridFTP server)
  - Updated custom authorization callouts
- Everything is installed as RPMs and Java (less than 2 minutes)
- Requires registration with Globus to create a regular endpoint corresponding with the data and idp nodes
- Better certificate management

Globus integration with CoG

- Helped NASA JPL integrate Globus download option with CoG
- Use case targets data accessible to public
  - Web based download using redirection to globus.org
  - Command line client for users to download (Python client that uses Globus APIs)
- Globus Transfer and Sharing was used
  - NASA JPL Globus endpoint: `jplnasagov#esgf-node`
  - Shared dataset with Globus: `jplnasagov#public`

    path: /

    ACL: all Globus users – read only

Work planned

- ESGF Transfer node:
  - Globus Connect Server IO (GridFTP server) with custom authorization callouts only
  - Supports Science DMZ architecture for replication and downloads
- Integration of Globus with replication tools
- Globus download in CoG for restricted data
- Better delegation model to get ESGF certificates for transfer
- Updates needed to support data format transformation (e.g. data in netCDF4 but downloaded as netCDF3)