Building Open Compute Systems using Globus Identity

David Champion
GlobusWorld, April 14-15, 2015
Argonne National Laboratory
Open Science Grid

- A national cyberinfrastructure for distributed high throughput computing
- 140+ resource providers in the Americas
- Properties:
  - loosely coupled, serial jobs
  - 800M CPU hours delivered in 2014
  - opportunistic resources accessed by HTCondor “glidein” pilots
OSG: 140 resource endpoints

... with campus users far and wide

... with x.509 AuthZ for **virtual organizations**

- Augment with simple sign-up and access for individual researchers
- Use campus identity services
- Reduce time between sign-up and job submission

→ OSG Connect
x.509 virtual organization validation occurs during wide area job distribution.

**trust relationship** between the resource provider and the OSG VO.

**users not required to use x.509 certificates directly.**
OSG Connect

- A central entry point for campus-based users and individual PI’s
- Access to resources using the OSG VO and glidein service
- An *identity bridge*: campus identity (CILogon) → OSG Connect identity (Globus) → virtual organization roles (OSG)
Establishing identity

- How do we get from <campus researcher> to user=angus on OSG Connect?
  - Globus Nexus provides an answer.
IdM with Nexus

Customer-directed team management

Self-service profile and credential control

SSH key upload (for login shell)
Python Nexus Client

- Nexus exposes a REST Web Service (API)
- Globus provides Nexus API client for Python
  - [https://github.com/globusonline/python-nexus-client](https://github.com/globusonline/python-nexus-client)
- Using the API, we:
  - search our group hierarchy for relevant changes
  - store these changes locally to track current state
  - provision user accounts into directory (nss_nis)
  - provision user filestores into storage systems
  - populate local group space as a pure reflection of Nexus groups
  - define access rights to GridFTP and job submission
Python Nexus Client (example)

```python
from nexus import GlobusOnlineRestClient

config = {
    'server': 'nexus.api.globusonline.org',
    'client': 'osgconnect',     # service account
    'client_secret': 'password', # secret!
}

gc = GlobusOnlineRestClient(config=config)
headers, response = gc.get_group_members(groupuuid)
members = response['members']
members = [member for member in members if member and member['username']]
members.sort(lambda a, b: cmp(a['status'], b['status']) or cmp(a['username'], b['username']))
for member in members:
    print '%s (%s) %s' % (group, member['status'], member['username'])
headers, profile = gc.get_user_profile(member['username'])
if profile.has_key('credentials'):
    keys = sorted([cred['ssh_key'] for cred in prof['credentials'] if cred['credential_type'] == 'ssh2'])
    # store ssh keys into ~/.authorized_keys
```
Summary: Data flow architecture

OSG Connect
- ~340 users
- ~60 projects
- ~40 campuses

Architecture extended to other campus integrations via CI Connect:
- ATLAS
- CMS
- Duke University
- UChicago
Thank you!

And our thanks to the Globus, CILogon and OSG teams. In particular:

- Rachana Ananthakrishnan (Globus)
- Mattias Lidman (Globus)
- Stephen Rosen (Globus)
- Mats Rynge (OSG)
- Jim Basney (CILogon) and the InCommon federation
Further information

Open Science Grid

http://opensciencegrid.org/

OSG Connect

http://osgconnect.net/

Python Nexus Client

https://github.com/globusonline/python-nexus-client