Accelerating Science
with connective services for cyber infrastructure

Rob Gardner
Computation Institute
University of Chicago
Science is Collaborative

- On campus
- Inter-campus
- National
- International
National Cyber Infrastructures

• XSEDE
• Open Science Grid
• Worldwide LHC grid
• Services ... e.g. Globus
• How do campuses connect to these, and to each other?
Challenges: limited budgets, operational costs, scale of science
Commodification, Centralization
Even so, distributed resources remain
Off-campus too
(the national ecosystem)
Need: transparency for users, cost effective for providers
Campus Data Centers
and the national cyber infrastructure
Globus data transfer, sharing
(and now publication)
Cycle sharing ecosystem?
The OSG fabric of services

- The leading **distributed** high throughput computing service in the US
  - 104k cores, 75.6 PB, 123 compute endpoints
- Well connected
  - Most sites on OSG have 10 Gbps or greater to I2 or ESnet, many upgrading, plus SciDMZs
  - (at least) 2M transfers/day, 1 PB/day
- Friction free
  - The OSG VO with GlideinWMS offers transparent access to these resources for small groups

A great way to share empty job slots
Growing number of campus users

70,000,000 cpu-hours/month

- Opportunistic use at ~>25%
- Mostly “Campus users”
- Diverse group of sciences

Lothar Bauerdick
OSG ED

2007 - 2014
Login to the OSG
(direct access to those cycles)

Globus endpoint
user#stash

osgconnect.net

stash
portal
login

UChicago UC3
Open Science Grid
Amazon EC2
Bringing resources together: Campus, OSG, Cloud

Scheduling to OSG via flocking to the GlideinWMS service
OSG Connect Suggested Campus Grids as Service

ci-connect.net built on Globus Platform and HTCondor
Services for Connecting

campus, HPC...
grid, cloud...
..bridging
Duke Grid → the OSG
(No significant effort needed by Duke SCSC staff)
Connecting communities to the national ecosystem

- **FAXbox**: rccf.usatlas.org (glidein factories)
- **portal**: login.usatlas.org
- **Campus Grids Off-grid Tier3**
- **ATLAS T1 (dev) Tier2**
- **TACC Stampede (dev)**
- **Cloud (AWS)**

Various campus resources

- XSEDE cloud

Users from 44 institutions
Strategy for XSEDE

TACC Stampede
http://www.tacc.utexas.edu/stampede/

P. Onysis, UTexas

Science Gateway for ATLAS
gather lessons for campus bridging

- 6,400 nodes
- 102,400 Sandy Bridge cores
- 390,400 Xeon Phi cores
- 32 GB RAM/node
- Special highmem, NVIDIA nodes
- 100 Gbps WAN
- 48 hr max job runtime
Minimize footprint @ TACC

- Key is minimizing Stampede admin involvement while hiding complexity for users
- Simple SSH to Stampede SLURM submit node
- ATLAS software mounted using CVMFS and Parrot
- ATLAS squid cache configured nearby
- Wide area federated storage access
- Leverage Globus, HTCondor, Glidein Factory, CCTools, OSG accounting
- (CI Connect Services)
Campus Connect Services
In Preparation
Summary

• Discovered we could leverage the Globus Platform to connect users to the Open Science Grid

• Suggested a model to couple users, data and distributed compute cycles as a service

• Easy path for campuses to “connect” or “bridge” to the national ecosystem
Acknowledgements

Steve Teucke, Rachana Ananthakrishnan + Globus Team!
Dave Lesny – UIUC (Midwest Tier2)
Lincoln Bryant, David Champion – UChicago (MWT2)
Suchandra Thapa (UChicago OSG)
Peter Onysis – UTexas
Jim Basney (CI-Logon) & InCommon Federation
XSEDE Science Gateway Team (Raminder Jeet, Suresh Maru, Marlon Pierce, Nancy Wilkins-Diehr)