Campus Grids

Rob Gardner

Mission

- Deliver distributed high throughput computing capabilities to campuses while enabling campus researchers with DHTC ready computing tasks

- Create a campus infrastructures community with workshops, online webinars and website to share best practices

Effort deployed

- I was appointed area coordinator in October 2013
- Effort
  - 0.5 FTE (Gardner)
  - 0.5 FTE (Thapa, effective 2/2014, shared w/ Software)
  - 1.0 FTE (Scientific Computing Consultant, search in progress)

- I have leveraged effort from overlapping OSG areas and aligned external projects
  - OSG User Support, Accounting, Security
  - glideinWMS core team + OSG Operations
  - Globus, US ATLAS Computing Facility, HTCondor
  - CI-Logon and InCommon team

Key Accomplishments

- Proposed “Platform of Services” model to facilitate delivery of needed DHTC services for campus researchers

- Delivered an OSG Connect platform, an interactive job and data service for the OSG

Key Accomplishments, cont.

- CI Connect proposed as a template for creating campus grids and campus bridges

- Two new OSG campus grids built with CI Connect
  - Duke CI Connect (Duke grid, bridged to UC3, OSG)
  - ATLAS Connect (LHC “Tier3” batch analysis pilot)

- Planning for Michigan next

The OSG Connect Platform

- "Login to the OSG using your campus identity"
- Launched at OSG Campus Infrastructures Community sponsored Workshop at Duke
  - August 26-27, 2013
  - About 25 users signed up using Duke credentials
  - Thousands of tutorial jobs executed on OSG
  - Several Duke OSG Projects created

Components

- Leverages Globus, HTCondor, CI-Logon, U-Bolt, Bosco technologies
  - Bundled as instance of a CI Connect service portfolio
  - *Provided as a **Service** to reduce Campus IT load*
- Submit host
  - Flocks to OSG VO front-end, UC3 grid, & Amazon if needed
- Object storage service (90 TB usable)
  - POSIX, Globus Online, http, chirp access protocols
- Accounting (Gratia) and monitoring (Cycle Server) services

Components, cont.

- Inside Science DMZ with 80 Gbps uplink
  - Direct peering I2 & ESnet at CIC OmniPoP (100 Gbps); PerfSonar
  - Co-located with major ATLAS Tier2 center and OSG opportunistic cycle provider
- Full advantage of Globus for reliable file transfer, data sharing
- Integrated web portal & group organization
- User-focused knowledge base (ConnectBook)

CPU Hours Delivered

Project Name	Field of Science	Usage Class	Contact	Institution	CPU hours
osg.KnowledgeSys	Psychology	Significant	Michael Culbertson	University of Illinois	383,182
osg.NESCent	Cross disciplinary Evolution Studies	Significant	Fabricia Nascimento	National Evolutionary Synthesis Center	220,547
osg.Staff	N/A	Significant	Rob Gardner	University of Chicago	143,685
osg.PathSpaceHMC	Computational Condensed Matter Physics	Significant	Frank Pinski	University of Cincinnati	60,288
osg.ConnectTrain	N/A	Significant	Various	Various	23,793
osg.Swift	Computer Science	Significant	Mike Wilde	University of Chicago	11,837
osg.SouthPoleTelescope	Astrophysics	Significant	John Carlstrom	University of Chicago	7,940
osg.EvoTheory	Evolutionary Biology	Some	Christina Burch	University of North Carolina Chapel Hill	264
osg.CompChem	Chemistry	Some	Chaoren Liu	Duke University	229
osg.DBConcepts	English Literature	Some	Richard Jean So	University of Chicago	4
osg.RADICAL	Computer Science	Experimented	Shantenu Jha	University of Chicago	1
osg.GlassySystems	Chemistry	Experimented	David Reichman	Columbia University	1
osg.CompNeuro	Neuroscience	Experimented	Po-He Tseng	Duke University	0
osg.AMFORA	Computer Science	Experimented	Zhao Zhang (Ian Foster)	University of Chicago	0
osg.UChicago-RCC	Research Computing Center	Experimented	Birali Runesha	University of Chicago	0
osg.NRELMatDB	Material Science	Experimented	Steve Sullivan	National Renewable Energy Laboratory	0
osg.BioStat	Bioninformatics	No accounting	Janice McCarthy	Duke University	0
osg.PlantBio	Plant Biology	No accounting	Joy Bergleson	University of Chicago	0
osg.RDCEP	Economics	No accounting	Ian Foster	University of Chicago	0
OSG Connect	All	Significant	Rob Gardner	All	851,772
Duke CI Connect	Campus grid	Significant	Tom Millege	Duke University	1,176,498
ATLAS Connect	Community grid	Significant	Rob Gardner	US ATLAS	255,077
				Total usage	**2,283,347**

CI Connect-based Campus Grids

- ATLAS Connect

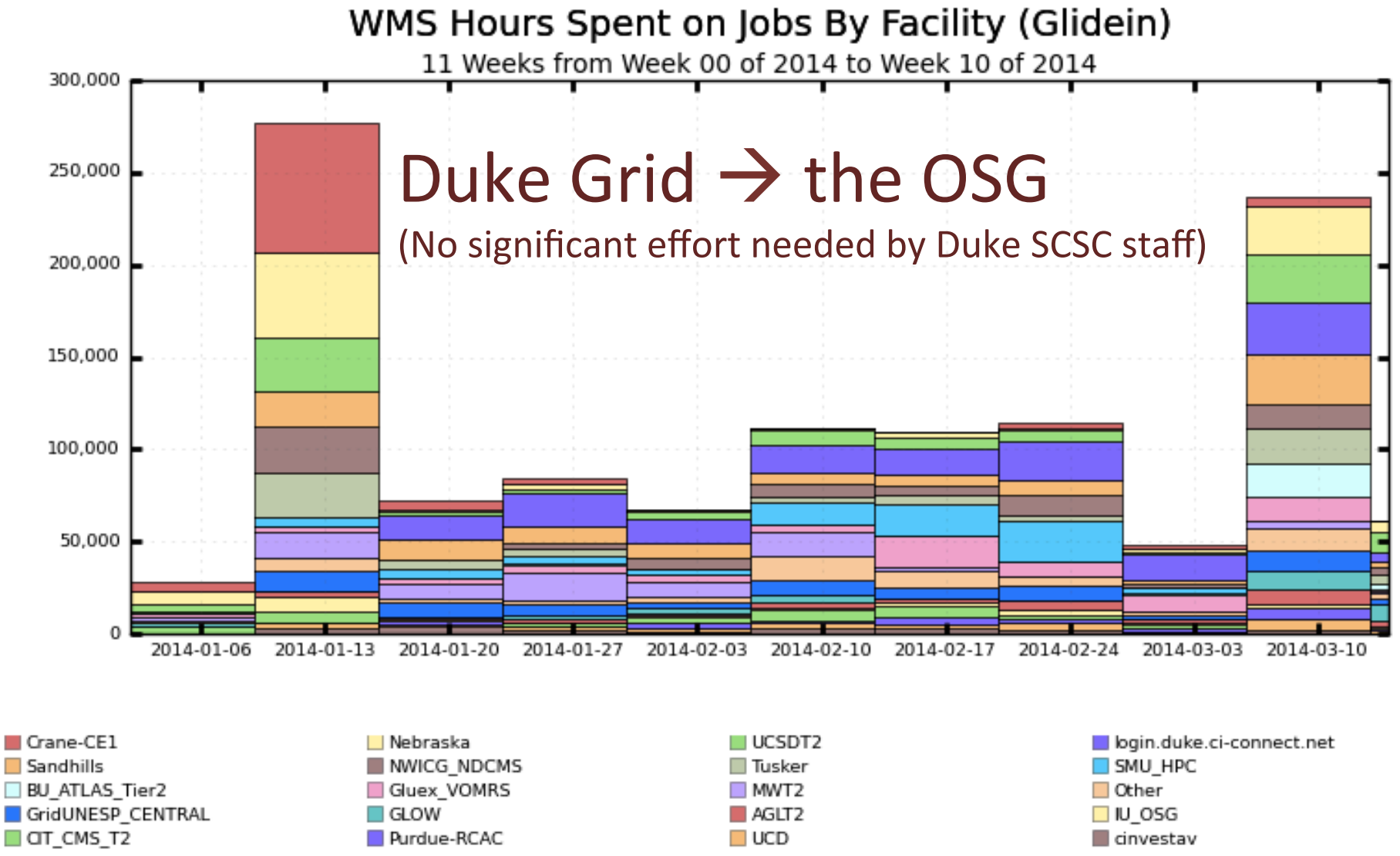
Potential to connect 44 US institutions: users & Tier 3

- **infrastructure**

Project Name	Usage Class	CPU hours
ATLAS-ORG-UCHICAGO	Significant usage	157,840
ATLAS-WG-HIGGS	Significant usage	30,059
ATLASCONNECT	Significant usage	24,037
ATLAS	Significant usage	20,042
ATLAS-ORG-FRESNO-STATE	Significant usage	7,246
ATLAS-ORG-UTEXAS	Significant usage	6,478
ATLAS-WG-SUSY	Some usage	1,857
CONNECT	Training	1,277
STAFF	Test and benchmark	6,242
	Total	**255,077**

- Duke CI Connect

Project Name	Usage Class	CPU hours
DUKE-4FERMION	Significant usage	747,931
DUKE	Significant usage	422,230
ATLASCONNECT	Significant usage	5,289
DUKE-QGP	Experimented	0
CONNECTTRAIN	Training	932
OSG-STAFF	Test and benchmark	116
	Total	**1,176,498**


WMS Hours Spent on Jobs By Facility (Glidein)
11 Weeks from Week 00 of 2014 to Week 10 of 2014
300,000
250,000
200,000
150,000
100,000
50,000
0
2014-01-06
2014-01-13
2014-01-20
2014-01-27
2014-02-03
2014-02-10
2014-02-17
2014-02-24
2014-03-03
2014-03-10
Duke Grid → the OSG
(No significant effort needed by Duke SCSC staff)
Crane-CE1
Sandhills
BU_ATLAS_Tier2
GridUNESP_CENTRAL
CIT_CMS_T2
Nebraska
NWICG_NDCMS
Gluex_VOMRS
GLOW
Purdue-RCAC
UCSDT2
Tusker
MWT2
AGLT2
UCD
login.duke.ci-connect.net
SMU_HPC
Other
IU_OSG
cinvestav
Maximum: 276,604 , Minimum: 28,237 , Average: 110,106 , Current: 61,593

Storage Accounting

Stash Server Project	Space Used (MB)	Files
osg.SouthPoleTelescope	16,080,000	22,451
osg.DBConcepts	2,300,000	6,400,000
osg.Staff	375,243	116,614
osg.AtlasConnect	217,575	809
osg.NESCent	179,000	1,237,816
osg.BioStat	125,000	1,296
osg.KnowledgeSys	7,000	129,647
osg.Duke-QGP	152	1,268
osg.ConnectTrain	126	1,425
osg.CompNeuro	120	1,290
osg.Swift	249	12
Total	19,284,465	7,912,628

	# users	# files	# TB
OSG Connect (Stash)	23	7,912,628	19
ATLAS Connect (Faxbox)	14	80,571	15
Total	**37**	**7,993,199**	**34**

Cyber Ecosystem Metrics

Campus grids created	3
(Total >=Level 3 OSG campus grids previously	6)
Campuses bridged	7
Campus grids connected to OSG	2
Off-grid campus research clusters connected	2
XSEDE sites connected	1
OSG Connect users	53
OSG Connect user institutions	12
New OSG Projects	7
Publications forthcoming	5

Challenges

- Encouraging campus cluster users to adopt DHTC
- Emulating home cluster look-n-feel on remote resources, e.g. software utilities
  - /home-cluster, /xsede, /osg environments
- Several workflow options: guiding users to the best fit
- Several data management and access options
- Wrapping off-grid campus clusters with Parrot tools (for software access)
- Compute site data storage or caching

Opportunities, Ideas, Qs

- Develop “standard recipes” for popular workflows, toolkits, science domains
- An OSG http federation and site-level http data caches?
- Giving campus researchers a single environment
  - Multi-user Bosco factory service to reach campus accounts (condo allocations); same for XSEDE
  - Distribute XSEDE campus bridging roll via OASIS
  - OSG user software toolkit distributed via OASIS