# Usable Distributed Cyberinfrastructure: Campus Grids, OSG Connect & User Support

slides online: http://bit.ly/1uM3qEz

Rob Gardner • Campus Grids Coordinator





### **Campus grids mission**

- Create an open forum for broad discussions of DHTC practices (users & campus research computing center professionals)
  - $\circ \rightarrow$  Campus Infrastructures Community
- Provide an easy on-ramp to OSG for campus researchers  $\rightarrow$  OSG Connect
- Deliver campus grids as a service

### **Campus Infrastructures Community**

#### CIC: Workshops, webinars, archive of demos



#### campusgrids.org

#### **OSG Connect for users**



Make OSG seem like a virtual campus cluster:

- login
- software
- job scheduling
- job data

### **OSG Connect approach**

- Leverage existing work, don't develop
  - OSG VO & GlideinWMS overlay services for resource access; OASIS, Modules for software access; Gratia accounting & OIM information service
  - Globus (ID & transfer), CI-Logon
  - HTCondor (& Bosco-as-service) for other clusters
- Deliver as service
  - Leverage automated build & configuration practices from LHC Tier 2 Centers





### **OSG Connect service elements**

- Identity & group management service
- Auto account provisioning to login host
- Flocking services
- Data service (for quasi-transient job data)
- Job monitoring, accounting service
- User facing monitoring & documentation
- Continuous validation services

## **Identity and group management**

- Leverage the Nexus service from Globus
- Use home campus identity with CI-Logon & InCommon federation services
- Users authorized in standard way but without grid certificate and VO service overhead
- Associate users with Groups (OSG Projects) for accounting and to control access

# Auto provisioning of user accounts

- The service periodically checks for newly authorized members
- Creates Unix account, and users can quickly login and start work (no admin needed)
- Goal is to have users running jobs on OSG within the hour

#### Acceptable Use Policy

By registering with OSG Connect as a GRID user you shall be deemed to accept these conditions of use:

- You shall only use the GRID to perform work, or transmit or store data consistent with the stated goals and policies of the VO of which you are a member and in compliance with these conditions of use.
- 2. You agree that any publications resulting from the computation performed on OSG will include the following citation:

This research was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science."

Additional information is available at <a href="http://osgconnect.net/citations-see">http://osgconnect.net/citations-see</a> also

https://twiki.opensciencegrid.org/bin/view/Management/ScientificPublications.

- 3. You shall not use the GRID for any unlawful purpose and not (attempt to) breach or circumvent any GRID administrative or security controls. You shall respect copyright and confidentiality agreements and protect your GRID credentials (e.g. private keys, passwords), sensitive data and files.
- You shall immediately report any known or suspected security breach or misuse of the GRID or GRID credentials to the incident reporting locations

#### I Accept

#### estimate of the second seco

Support - Resources - Connect - Sign In/Sign Up -

#### Sign up for OSG Connect

You can sign up for OSG Connect in a few basic steps:

- Visit the OSG Connect signin portal (this will open in a new tab or window).
- 2 Click Proceed to authenticate with your campus NetID. Your browser will redirect you to a cilogon.org site.
  - . In the Select an Identity Provider area, find OSG Connect and select it.
  - Check the remember this selection box, then click Log On. Your browser will redirect you to the OSG Connect authentication page. It should look familiar to you. If you have recently signed in, you may not need to reauthenticate.
  - Sign in as you normally would, using your campus NetID and password. Your browser will take you briefly past cilogon.org again, before returning you to the signin portal. These steps allow you to sign in to the web portal any time using your OSG Connect credentials.
- 3 Now you will see a page entitled Need to Make a Connection. This links your campus NetID to a OSG Connect account. If you aready own a Globus account; that will also be your OSG Connect account: sign in with your Globus credentials here, then move ahead to step 5.
- Otherwise, click Create a new Globus account.
  - Enter your full name and your OSG Connect email address. It's important to use your OSG Connect address so that OSG Connect administrators can approve your access.
  - Enter a username and password to use for your OSG Connect account. The site will interactively let you know if your chosen name is unavailable.
  - You must indicate your acceptance of the Globus terms of service by checking the appropriate box. (Other boxes are optional.)
  - Click Register to create your account. You will need to validate your email address before proceeding:
    - Wait for an email from support@globus.org containing your validation URL.
  - · Click the link, or paste it into your browser where you are.
  - · Your account will be confirmed.
- 5 To complete signup, you will join the osg group. This is automated, but requires some additional information:
  - your first name
  - your last name

Signing up

- · your field of science
- · your organization (university, institution, agency, etc)
- your department within that organization

This information is used for resource utilization reports, to build longitudinal analyses of how grid resources are used.

6. After a moment, you will have joined the osg group, and your browser will bounce over to the Manage Identities screen. It is advisable that you add an SSH public key now, but it is not required — you may do this at any time. The SSH key will grant you passwordess access to the OSG Connect login server, <u>login .osgconnect.net</u>. If you do not know how to generate an SSH key, you may ind our SSH Key Generator helpful.

<b></b>	osg connect	Support	- Resources - Cor	nect - Transfer - rwg -	, OSG Home Members Subgroups
My Groups	Search	osg			
INVITATIONS (MEW & RES	SPOND)	Hama	Subgroups		connect
MEMBERSHIP APPROVAL	L PENDING	Home Members	Subgroups	≡	osg
University of Chicago Rese	earch Computing Center	Managers are listed in bold.			osg.AMFORA
MANAGER OF		CI Connect 🥸	Garhan Attebury	OSG Training 1	osg.AlGDock
		OSG Connect Operations	Gerard Bernabeu Altavo	OSG Training 2	osg.BioMolMach
* MEMBER OF		Aaron Garow	Giovanni G. Baez Flores	OSG Training 3	osg.BioStat
atlas.org.uchicago		Alain Devino	Glen Hocky	OSG Training 4	osg.CompChem
atlas.wg.Exotics		Alan Sill	Gonzalo Merino	OSG Training 5	osg.CompNeuro
atlas.wg.Higgs		Alden Stradling	Harinder Singh Bawa	OSG test	osg.ConnectTrain
atias.wg.SUSY		Amanda Lund	lan Foster	Ole Weidner	osg.ConnectTrain
connect		Amanda Whitlock	Igor Sfiligoi	Patrick A Reeves	osg. Continuousintegration
duke		Amin Nikakhtar	lgor Sfiligoi	Patrick Malsom	osg.Doconcept8
duke.4fermion		Amit Kumar	Ilija Vukotic	Paul Alexander	osg. Duke-order
duke.campus		Andra Luckow	JUAN DIEGO HERNANDEZ J	Peter F Couvares	osg. EvoTheory
osg		Andre Luckow	Jacob Roberts	Peter Hadlaw	osg.ExhaustiveSearch
osg.ConnectTrain		Andrea Manzi	James Christopher Westland	Philip W. Smith	osg.FFValidate
osg.OSG-Staff		Angrea Manzi	James Zhang	Philippe Grassia	osg.GRASP
umich		Anna Dizabeth Woodard	Jan Balewski	Po-He Tseng	osg.GlassySystems
		Anna Oison	Janina Krumbeck	Bachana Ananthakrishnan	osg.HTCC
		Annika Eriksson	Jay Fowler	Bachel Killackey	osg.HealthInformatics
		Antoop Mayampuratin	Jeff Dandov	Rafael Francisco Pagan Vargas	osg.lceCube
Create New Group »	admin action required	Anton Betten	Jerry Perez	Raffaele Montella	osg.KnowledgeLab
		Anton Yu Delementer Desirate:	Jimmy Dorff	Bick Irving	osg.KnowledgeSys
		Balamurugan Desingnu	Joakim Olsson	Bob Gardner	osg.NESCent
		Ben Esenbraun	John Blischak	Bob Quick	osg.NHELMatDB
		Benjamin Brachi	John wang	Robert Cowles	osg.OSG-Stall
		Bing Xe	Jonah Bernhard	Bobert Hopie	osg.Octops frain
		Bo Hodda	Jonathan Schroeder	Bohan Jaitpal	osg.PathSpaceHMC
		Bockjoo Kim	Jordan Govreau	Buna Sravani Kommineni	osg.PhysStat
		Brendan Albano	Jose Ramón Alvarez de los	Buth Marinshaw	osg.PlantBio
		Brennan Govreau	Joseph Haley	Saad Nasser	osg.ProtFolding
		Brian Mitchell	Joseph Kleinhenz	Samuel Meehan	osg.Proteomics
		brij Kishor Jashai	Julian Valay	Samuel van Loon	osg.RADICAL
		Brooklin Gore	Karan Vahi	Scott Shi	osg.RDCEP
		Bryan Howell	Karl Jablonowski	Scott Teine	osg.SBGrid
		Cesar Pollack	Karthikevan Saravanan	Shantenu Jha	osg.SNOplus
		Chaoren Liu	Kanichi Hatakayama	Shia lian Ding	osg.SouthPoleTelescope
		Chaowen GUO	Konnoth Aird	Shila LI	osg.StanfordRCC
		Chen Li	Kenneth Airu	Singe Laseba	osg.Switt
		Chetan Raj Rupakheti	Ketan Mehanhwari	Sinon Jacobs	osg. Ucnicago-HUU
		Choa-lin Chou	Keuan Maneshwan	Soren Hayashi	osg.userScribbi2014
		Christian Ross	Nexus Fill	ouun Nat Lau	Uou.duiiiid

#### Groups and users



### Job service: keeping it simple

- Not a gateway or web portal
  Users must be shell-literate
- Resource targets advertise attributes back to the Condor schedd as usual
- Usual ClassAds and Condor submit scripts
  - Users are responsible for anything more in terms of workflow tools
- Recipes, templates for the "DHTC way"

#### **Resource targets**

- OSG GlideinWMS is the workhorse
- UChicago campus grid (UC3)
- In progress:
  - Syracuse University HTCondor pool (OrangeGrid)
  - St. Louis University SGE compute pool
- Any cluster running Condor, PBS, SLURM, or SGE can be directly added to OSG Connect (using Bosco) requiring only a user account on the resource
- Optionally can add Amazon EC2 instances (e.g. workshops, bootcamps, etc, for guaranteed execution)

## **Storage service: Stash**

- Provide a quasi-transient storage service for job input/output data
- Implemented with Ceph object store
- POSIX access provided to the login host
- Globus Online Server for managed transfers from campus data services
- Personalized http service endpoint



Easy, reliable upload of data to Stash for processing on OSG, and downloading output data back to campus.

#### Here we leverage Globus transfer



C 🕯 🗎 https://po	A 🔒 https://portal.osgconnect.net/xfer/ManageEndpoints					
∍⊜⊜ osg co	nnect					
lanage Endpoints						
			add Globus Connect Personal	add an endpoint		
recently used in use in a shared v	/ith me 🎢 shared by me	Administered by me	all miter list by endpoi	nt name		
endpoint		status	credential			
atlasconnect#faxbox		in use	never expires	-		
connect#faxbox		ready	never expires	•		
connect#stash		in use	never expires	-		
osgconnect#stash		ready	never expires	-		
Globus Connect Personal		ready	never expires	-		

### **Application software**

#### • Legacy DHTC methods for software access: HTCondor file transfer + http access Ś module avail OSG OASIS+Modules \$ module load R

$\odot \odot \odot$	D ☆ rwg — rwg@login01:~ — ssh						
[rwg@login01 ~]\$ module avail							
/cvmfs/oasis.opensciencegrid.org/osg/modules/modulefiles/Core							
R/3.1.1 SitePackage SparseSuite/4.2.1 atlas blast blender curl/7.37.1	fftw/3.3.4-gromacs fftw/3.3.4 gcc/4.6.2 glpk/4.54 gromacs/4.6.5 hdf5/1.8.13 jpeg	(D)	lapack 1mod/5.6.2 matlab namd/2.9 octave/3.8.1 pcre/8.35 python/2.7 (D)	python/3.4 qhull/2012.1 settarg/5.6.2 wget/1.15			
Where: (D): Default Module							
Use "module spider" to find all possible modules. Use "module keyword key1 key2" to search for all possible modules matching any of the "keys".							
[rwg@login01 ~]\$							

- \$ module load namd
- \$ module purge

#### Common OSG user collections

- Curating based on user demand
- Will include XSEDE rpms
- Tracking by user, project & site for analytics

#### **User-facing OSG usage & data**



#### Storage used by me, by my projects

۰۰	serve osg connect					
			Search:			
Type 🔺	Name	÷ Ki	nd	÷ c	hanged	+ Size
>	temp		directory	y <sup>Th</sup>	u, 17 Jul 20 12:	14 0 item 51
	RegistrantsList (1).xls	application	on/vnd.ms exce	- Mon	, 26 Aug 20 21:	13 31.00 01
	chart_pie_30805.html		text/htm	nl Tue	, 24 Sep 20 11:	13 522 40
	chart_pie_5000.html		text/htm	nl Tue	, 24 Sep 20 11:	13 129 41 129
	user.kkrizka.01242101284.output.root	applica	ation/octet strean	- Tue	, 24 Sep 20 14:	13 20 970.00 i

← Cycles used by me, my projects, and by others

#### Job Table

User	Project	Site	Number of Jobs ▲	Wall Time (hours)	CPU Time (hours)	CPU Efficiency (%)
David Schultz	IceOube	Project Total	247956	50458.79	27605.83	37.78
		CIT_CMS_T2	37699	7323.69	4245.65	57.97
		uc3-mgt.mwt2.org	35850	10080.07	6802.68	67.49
		MWT2	25871	8413.21	5568.44	66.19
		Tusker	18977	2679.87	1381.09	51.54
		UCSDT2	17899	2675.77	1295.81	48.43
		UConn-OSG	15346	1869.64	1013.89	54.23
http://bit.lv/	http://bit.ly/1sPlxrX		11892	3629.85	1908.63	52.58
<u>mup.//bit.ty/</u>			9316	1282.32	59.50	4.64
		AGLT2	8743	2520.62	1798.46	71.35
		Nebraska	8575	1800.42	1161.09	64.49
By user	By user, project, site, #jobs, wall time,		8515	1170.20	262.08	22.40
Dy user, j			7585	1265.95	0.00	0.00
site, #job			7160	781.22	11.34	1.45
wall time			6832	418.38	87.66	20.95
wan une,			5686	755.31	477.31	63.19
cpu time, cpu efficiency		BU_ATLAS_Tier2	3735	1126.05	654.30	58.11
		FLTECH	3106	359.10	31.09	8.66
		NUMEP-OSG	2853	147.27	56.85	38.60
			2149	100.28	15.50	15.46
Tailored to	Tailored to user: can optionally see where		2139	290.97	0.00	0.00
optionally s			1934	198.29	0.00	0.00
jobs run ar	nd how	UCD	1076	305.09	129.79	42.54
efficiently	More job	SPRACE	966	350.93	209.78	59.78
info ploppo	info planned		750	215.72	125.86	58.35
nilo planne			726	109.36	68.00	62.18
(success/fa	allures, by	IU_OSG	628	193.03	117.13	60.68
task, etc.)		Sandhills	485	59.10	35.86	60.67
		GRASE-CSE-MAGIC	354	108.11	8.60	7.96
		VT_OSG	348	87.40	6.84	7.83
		LITA SWT2	294	57.42	30.90	53.82

### **Job Monitoring**



16

#### **OSG Connect usage in 2014**

#### OSG-Connect Projects 2014/01/01 - 2014/08/12

Project Name	PI	Institution	Field of Science	Wall Hours
AlGDock	David Minh	Illinois Institute of Technology	Chemistry	154,236
aprime	Jan Balewski	MIT	Nuclear Physics	348
atlas-org-uchicago	Robert Gardner	University of Chicago	High Energy Physics	1,165
AtlasConnect	Robert Gardner	University of Chicago	High Energy Physics	768
CompNeuro	Po-He Tseng	Duke University	Neuroscience	67,048
ConnectTrain	Robert Gardner	University of Chicago	HTC Training	441,422
ContinuousIntegration	Robert Gardner	University of Chicago	Technology	1,503
DBConcepts	Richard Jean So	University of Chicago	Computer Science	26
FFValidate	Vijay Pande	Stanford University	Chemistry	34,035
HTCC	Robert Quick	Indiana University	Community Grid	142
IceCube	Francis Halzen	University of Wisconsin	Astro-particle Physics	534,653
KnowledgeLab	James Evans	University of Chicago	Other	100,537
KnowledgeSys	Michael J. Culbertson	University of Illinois, Urbana-Champaign	Educational Psychology	218,451
Orbiter	Anton Betten	Colorado State University	Mathematics	50
OSG-Staff	Chander Sehgal	Fermilab	Computer Science	117,543
OSGOpsTrain	Robert Quick	Open Science Grid	Community Grid	43
Proteomics	Sam Volchenboum	University of Chicago	Bioinformatics	320
ProtFolding	Jinbo Xu	Toyota Technological Institute at Chicago	Bioinformatics	531,865
RADICAL	Shantenu Jha	Rutgers University	Computer Science	11
scicomp-analytics	Robert Gardner	University of Chicago	Mutli-Science Community	11,574
SouthPoleTelescope	John Carlstrom	University of Chicago	Astrophysics	7,507
Swift	Michael Wilde	University of Chicago	Computer Science	280
UserSchool2014	Tim Cartwright	OSG	Mutli-Science Community	739
Total				2,224,264



#### **Campus grids as service**



A generic pattern for connecting campuses to national cyber infrastructure

#### Duke CI Connect: 1<sup>st</sup> campus grid as service



Minimize entry barrier for campus grids & bridging

Users submit locally as before, but can overflow to OSG or other targets

No local IT building: just a standard login account to get started

## **Duke Campus** → **OSG Open Facility**



Maximum: 276,604 , Minimum: 28,237 , Average: 110,106 , Current: 61,593

### **CI Connect for communities**



Virtual LHC Tier-3 service for analysis batch processing

Flexible mix of resource targets, including "off-grid" Tier-3s and campus clusters

Lessons for software access on non-OSG, non-LHC resources (e.g. XSEDE)

## **Campus grids in progress**

• Expect these to be in production very soon

Each presents specific challenges that provide lessons for new campuses.

Service upgrades provided centrally, no local updates, re-installs, etc.

THE UNIVERSITY OF Research CHICAGO Computing Center





Revisiting submit client as module (rpm installed by local admin) for home cluster  $\rightarrow$  connect service  $\rightarrow$  OSG + CI submit mode { smodule load connect }

# **Back to Syracuse OrangeGrid**



- HTCondor campus grid:
  - ~3000 desktops with virtual machines running in the background supporting SL6 and Ubuntu
  - Condor Virtual Machine Coordinator interfaces with HTCondor to launch and manage virtual machines. Grown from 6000 to 10000+ cores.
- First Phase:
  - OrangeGrid as an OSG Connect resource target
- Second phase:
  - CI Connect instance, virtually extending OrangeGrid to the OSG for Syracuse researchers

## **User support & engagement**

- CIC workshops
- OSG Connect →
  OSG User School
- With User Support, developing joint boot camp with Software Carpentry for the fall



Strategy will be to partner with campus HPC / computing center consultants



## **Summary and Conclusions**

 Campus Grids and OSG Connect are effectively incorporating core OSG principles o shared access, overlays, trust model, openness,... Will evolve methods to integrate campus infrastructures for the benefit of small research groups and campus IT providers Providing them as value-added services for campus IT means multiple instances benefit from innovation