



Open Science Grid ED Report to Council Meeting

October 2, 2013

Lothar Bauerdick



Main Areas OSG Delivers On

Open Science Grid

◆ Production/Operations

- ★ including an evolving bouquet of services at the GOC, UCSD, elsewhere
 - ◆ DigiCert, SHA-2, OSG-XD host, OASIS, OSG Connect, network monitoring, ...

◆ User Support

- ★ consulting on technologies, architectures and user support
- ★ spreading knowledge on HTC as a science problem solver

◆ Technologies

- ★ including software packaging, system testing, patching

◆ How about Campus Grids? Clouds?

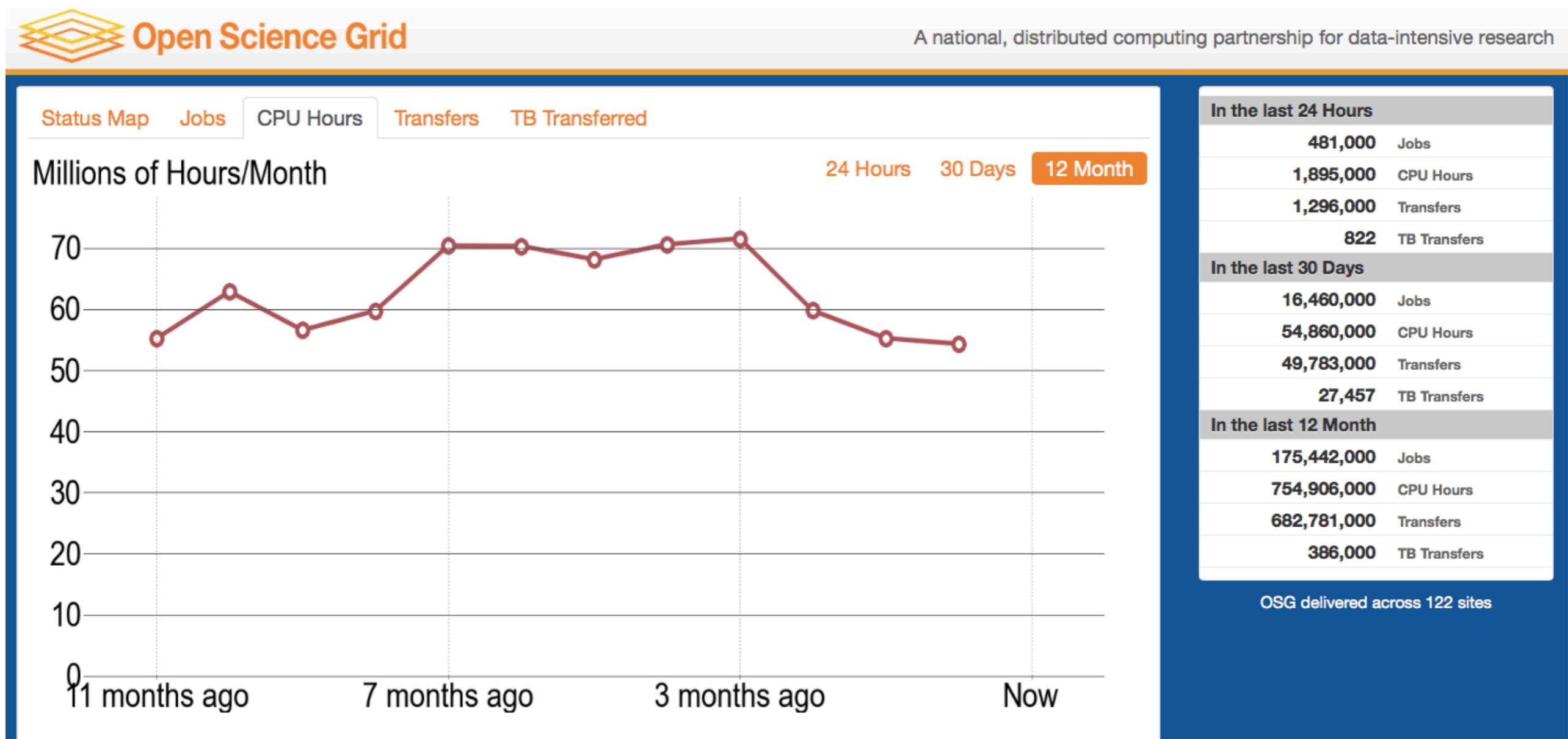
- ★ This talk will concentrate on reporting status, and outlining strategies and plans as discussed at the recent OSG ET retreat



Operations/Production

Open Science Grid

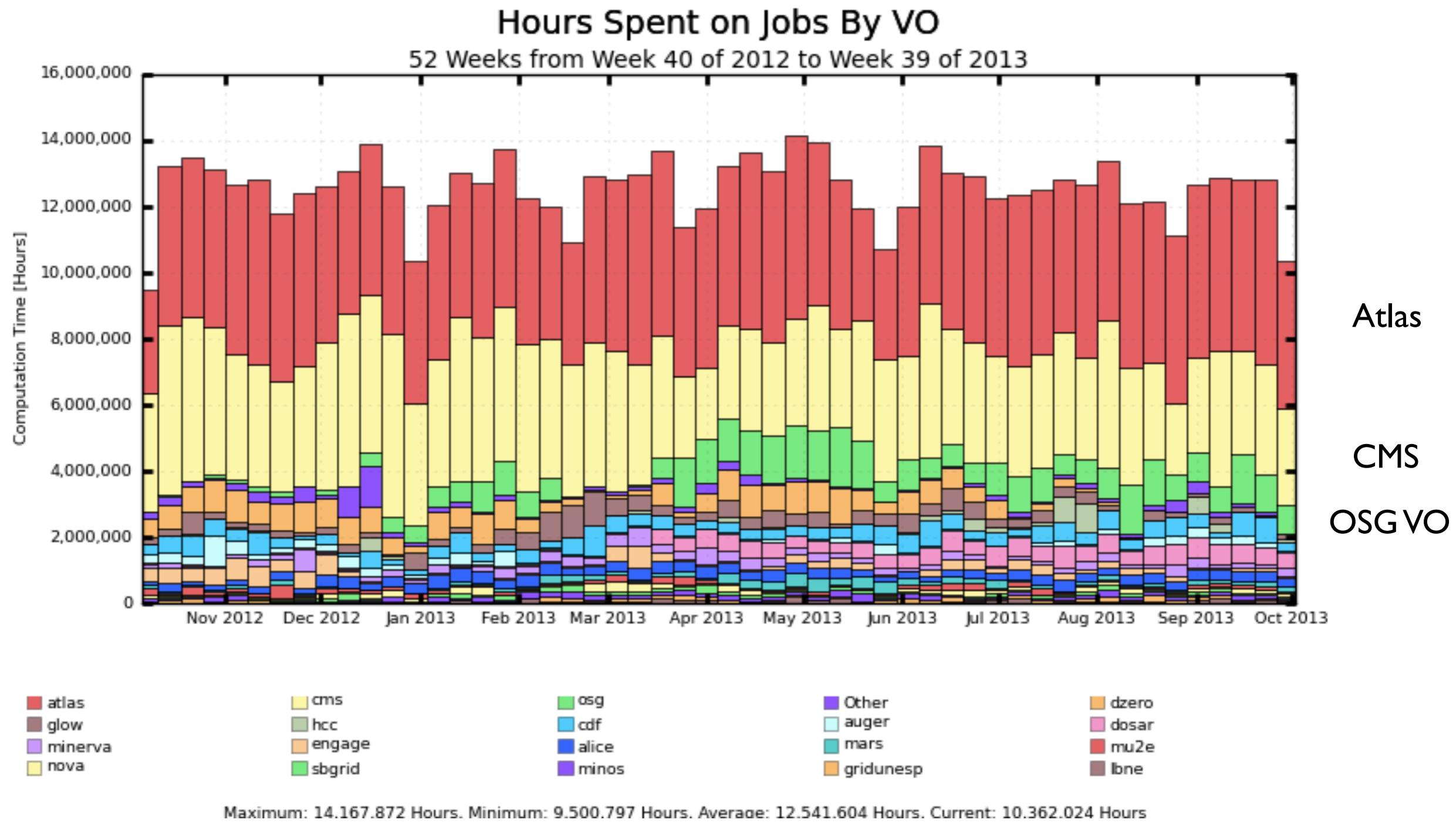
- ◆ Rob Quick/Operations Manager running Production meeting
- ★ defined roles in operations group, interfacing to software through release manager at UW
- ★ a number of important transitions: DigiCert, SHA-2,





Open Science Grid

Resource Use across OSG ~Constant during the year



- ◆ LHC use somewhat down, given the shutdown
- ◆ non-LHC use, in particular OSG-XSEDE and others, are up!



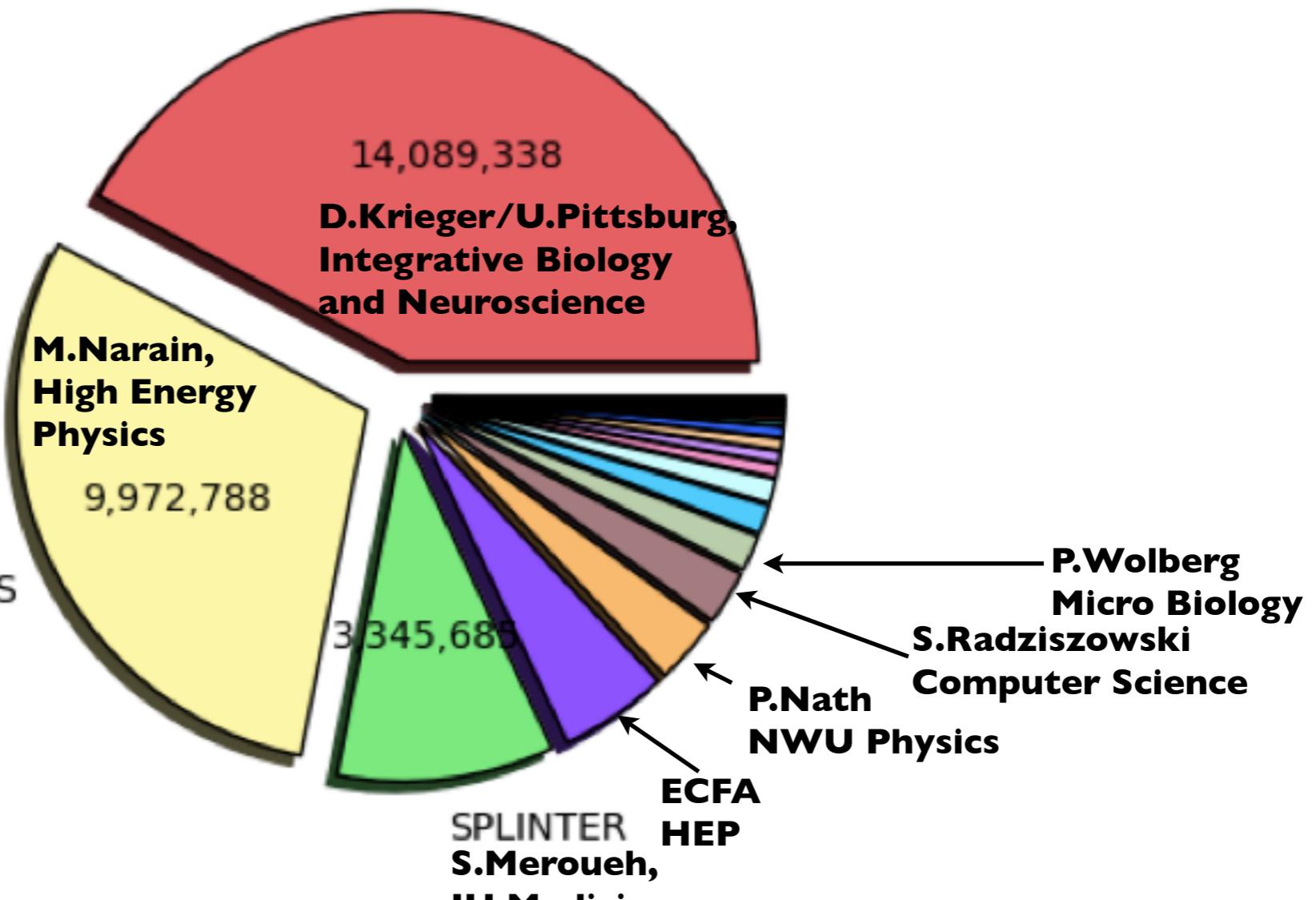
Open Science Grid

Opportunistic Resource Delivered To Individual Science Projects in 2013

Wall Hours by VO (Sum: 33,432,599 Hours)

39 Weeks from Week 00 of 2013 to Week 39 of 2013
TG-IBN130001

See Tanya's presentation tomorrow



■ TG-IBN130001 (14,089,338)
■ TG-PHY110015 (1,004,429)
■ DUKE-QGP (371,289)
■ Other (162,274)
■ TG-PHY120014 (59,429)

■ SNOWMASS (9,972,789)
■ RIT (836,827)
■ DETECTORDESIGN (202,613)
■ TG-ATM130015 (77,169)
■ CON-PSHMC (53,272)

■ SPLINTER (3,345,686)
■ UMICH (553,786)
■ TG-DMR130036 (183,874)
■ TG-DMS120024 (68,908)
■ TG-ATM130009 (51,222)

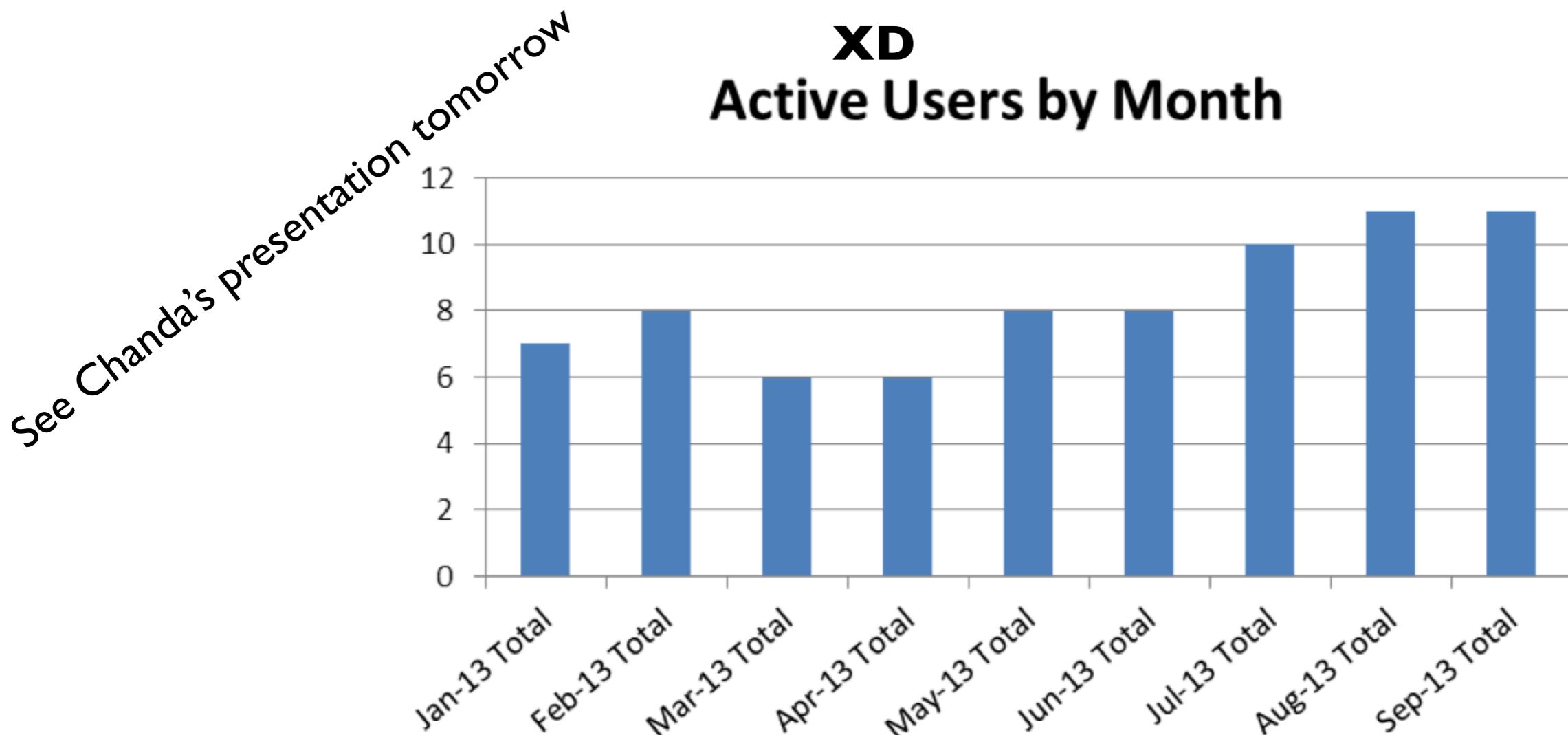
■ ECFA (1,669,414)
■ TG-TRA100004 (444,375)
■ EIC (175,818)
■ TG-DMR120085 (60,625)
■ OSG-STAFF (49,464)



User Support and XD SP

Open Science Grid

- ◆ Support of VOs, Campuses, Applications
- ◆ Support of individual researchers
- ◆ Reaching science groups as Service Provider in XD
 - ★ number of active users now at 11 individual groups/month
 - ◆ also looking at other metrics like “number of researchers touched by OSG”





Other “Production” Topics

Open Science Grid

♦ Software Release Area

- ★ led by Tim Theisen/UW

♦ Networking Area

- ★ goal: push deployment of perfSonar instrumentation beyond LHC sites
 - ◆ RobQ to take “ownership” of deployment push, as an overall production issue
- ★ still issues with perfSonar, software issues slow down deployment
 - ◆ and OSG might have to take some leadership role in improving the situation, maybe partnering with ESnet/Internet2 to help with release process and problem follow-up?
 - ◆ also future of dashboard and next steps towards a “perfSonar dashboard service”

Also across
WLCG...

♦ Security: developing new models for authentication and authorization

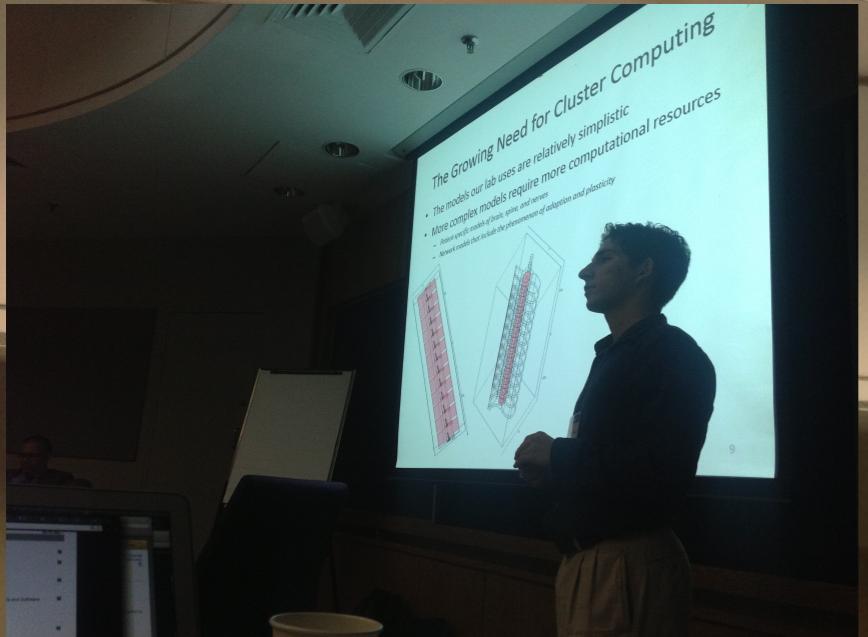
- ★ “overlay” pilot-based execution environment:
 - payload jobs get executed by OSG “on behalf” of science users
- ★ exploiting the existing trust relationships and traceability capabilities
 - ◆ first step: at Fermi got security team acceptance of traceability in lieu of the previous PKI-based authentication requirements
- ★ also, OSG-connect links campus ID and GlobusOnline ID with OSG job



OSG Providing User Support

Open Science Grid

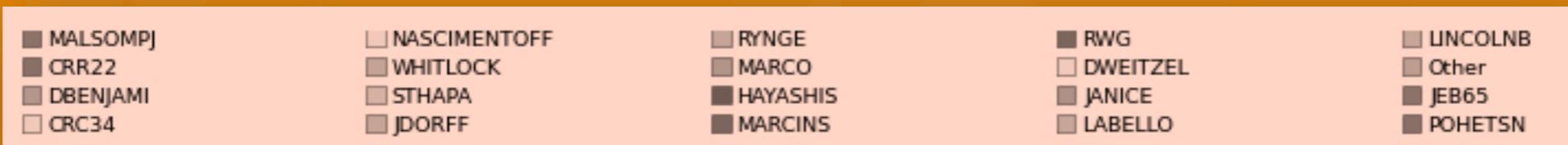
- ◆ support for individual researchers
 - ★ BOSCO as the product for the researcher to move her application to HTC
 - ★ XD Allocation, run from OSG-XD host
 - ★ Campus Researcher Club, run at a number of “franchising” sites
- ◆ OSG support for applications
 - ★ BOSCO to support R
 - ★ Galaxy, what are the demands on OSG project? Software/TI?
 - ★ MathLab at particular campuses
 - ★ etc
- ◆ support for VOs
 - ★ strengthen opportunistic use: User Support, hiring a “production analyst”
 - ★ target "flagship" VOs relevant to DOE: IF experiments, LHC Opportunistic



Campus Infrastructure Community Workshop at Duke

Duke CIC Workshop by Numbers

- ✿ 65 registered participants (5 OSG staff, 2 UChicago ATLAS admin staff, 1 UChicago CI staff)
- ✿ 5 presentations by Duke science and IT
- ✿ 20 technical OSG modules presented
- ✿ 29 users filled out pre-workshop survey
- ✿ 4 projects established at workshop
- ✿ 60 activated OSG Connect accounts
- ✿ Most users successfully submitted $O(1000)$ jobs





Focus and consolidate the OSG Campus Grids Efforts

♦ Existing Efforts

- ★ “Campus Grids” Area, led by Dan Fraser until June
 - ◆ main deliverable: BOSCO
 - ◆ BOSCO team subcritical, leadership left
- ★ CIC, led by Rob Gardner
- ★ “Campus Researchers Club” between UNL, UW, UCSD, UC
- ★ User Support helping VOs, applications, individual users
- ★ OSG-XSEDE, joint between User Support and Operations
- ★ XD Service Provider, led by Chander Sehgal
- ★ Opportunistic Use, Production Analyst, tbh

♦ OSG Staff Retreat in June

- ★ first discussion of the OSG-Connect idea, also in June Council meeting
- ★ appointed Rob Gardner to start the OSG Connect service

♦ OSG ET Retreat in September

- ★ discuss the strategic goals and plans for campus area
- ★ discuss clouds



OSG Strategy For Campus Grids

♦ OSG taking a three-prong approach:

- ★ support the individual research with the “myHTC”/BOSCO product
 - ◆ BOSCO is a great tool, currently mostly used by “system experts”
 - ◆ most of functionality is “HTcondor with ssh” —> HTcondor project
 - ◆ start a new thrust to fully “productize” the “myHTC” product
 - ◆ UW, User Support
 - ◆ goal: a downloadable self-installable and self-configuring robust tool to allow researchers HTC on whatever resources are easiest available to her
- ★ support campus grids with OSG Connect service
 - ◆ provide campuses with a OSG connect service so they don’t have to build their own
 - ◆ RobG has put together the initial set of services, presented at the Duke CIC mtg
 - ◆ discussed staffing of this area, how to integrate the new services with OSG areas
- ★ support VOs with OSG proving access to opportunistic resources
 - ◆ work to make this more dependable, robust, accountable — improve elasticity
 - ◆ add new provisioning and scheduling capabilities, like dynamic resources provisioning, allocation-based provisioning, clouds etc

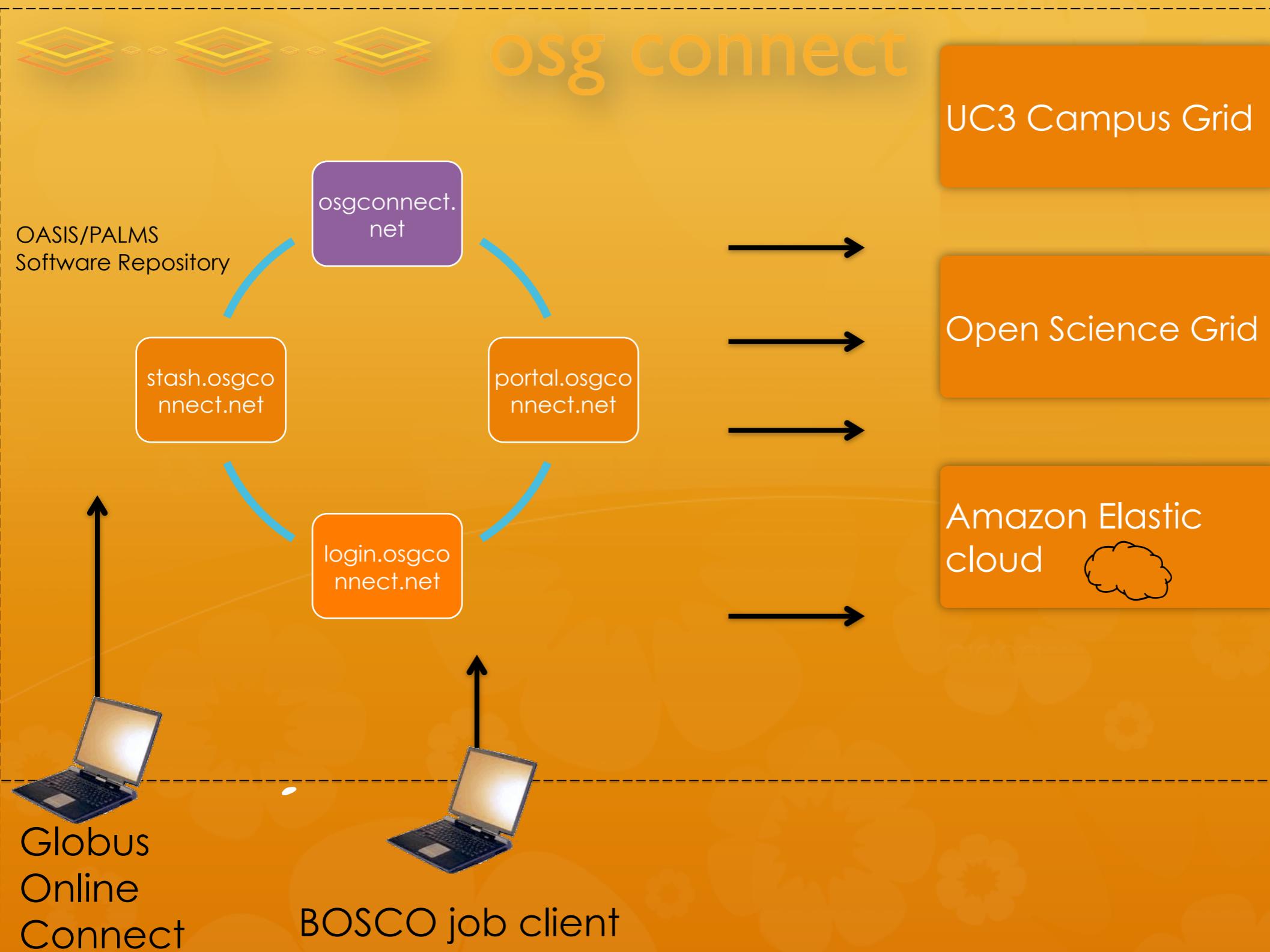


OSG Connect

Open Science Grid

- ◆ Basic idea: OSG to provide Campus Grids “as a Service”
 - ★ OSG-connect provides range of service to individual users, department/facility, or campus
 - ★ enabling them to move their applications to use HTC
- ◆ Provide value to
 - ★ The end-user science communities
 - ★ The departmental cluster or discipline facility
 - ★ The campus research center
 - ★ in a first step, target additional campuses: IndianaU, DukeU, others
- ◆ OSG-connect already starts with a useful set of services, to be further developed
 - ★ Support user communities, common dHTC templates, community software and service frameworks (e.g. Galaxy)
 - ★ Support campus instance deployments (e.g. “osgconnect.duke.edu”, “go-connect.umich.edu”, etc)
 - ★ Integrate with OSG-wide services (software & data & workload mgt)
 - ★ Develop materials for periodic campus visits
- ◆ Provides a platform to innovate and develop the OSG capabilities and services
 - ★ e.g. in the space of data, ID management, etc

CI Connect Services Deployed as OSG Connect Instance





osg connect

Support Resources OSG Connect Log In or Register

Welcome to OSG Connect

Efficiently connect your science to cycles and data.

Connecting Science to Cycles

OSG Connect offers investigators simple and efficient access to distributed high throughput computing resources required by many of today's most challenging problems in science, engineering and the humanities. Whether accessing your [campus grid](#), science cloud or the opportunistic cycles available in the national-scale [Open Science Grid](#), you need a simple environment to bring resources together in order to accelerate your science goals. OSG Connect offers several capabilities that will fit into your vision of scientific computation.

Connect Your Campus or Your Project

While OSG Connect offers immediate user access to opportunistic cycles from participating sites and campus grids of the Open Science Grid, it can also be configured to implement institution-specific connection environments. For example, resources of a campus grid (or cluster) can be virtually extended using an OSG Connect service, providing a computing center with a high throughput "overflow" service, or it can implement bridged campus arrangements as may be required by multi-institutional collaborative science partnerships.

Cycles and Data Virtually in One Place

OSG Connect has [Stash](#), a fast, high-volume file storage service that connects your data anywhere you need it to be. "Stash" your input data for processing and write output from your jobs for later retrieval back to your home institution using services such as [Globus Online](#), Web, or other file transfer and sharing tools.

User Sign Up

All that is needed to get started is what you likely already have: a research problem in a recognized field of science and a network identity issued by your home academic institution. If you are a member of an existing OSG Connect Project established by a principal investigator, the sign-up will be even quicker. Simply [sign in](#) to request access to OSG Connect. Once you're fully enabled, you can begin the [tutorial](#) to learn more about distributed high throughput computing in the OSG Connect environment and start accelerating your science.

[Sign up here.](#)



stash.osgconnect.net

Support Resources OSG Connect Log In or Register

OSG Connect Stash

Stash is a data access solution for distributed high-throughput computing.

Fast

Stash runs on [Ceph](#), a highly-scalable, clustered, POSIX-compliant network filesystem. Stash is connected to the OSG through a 10 Gbit WAN, and can perform parallel writes very quickly.

Available

Stash is mounted on all OSG Connect submit nodes ([login.osgconnect.net](#)) at [/data](#), and is visible to workers via Parrot/Chirp. The portion of your Stash designated public is also available directly on the WWW as [http://stash.osgconnect.net/+yourusername](#). Projects have stashes, too! A project's public stash can be found at [http://stash.osgconnect.net/@projectname](#).

Integrated

Stash is connected to Globus Online for fast, easy, high-volume data transfers. Just by registering for OSG Connect you can perform GO transfers to your Stash datastore. You can begin a transfer above in the [Connect](#) menu.

Sign up

Every fully enabled OSG Connect user has Stash access. If you're not already signed up, [register now!](#)



Support Resources OSG Connect Transfer rwg

My Groups Search

Groups
Use groups to easily manage access to shared data and web resources.

Access Groups
The My Groups tab on the left provides instant access to all of the groups that you belong to and manage. If you manage a number of groups use the My Admin Queue link to get a consolidated view of all your outstanding admin tasks.

Find Groups & Users
Use the Search tab on the left to quickly locate both groups and users.

Create Groups
Easily create new groups and invite users to join them. Use the link at the bottom of the left panel.

Create New Group * • admin action required

© 2013 Open Science Grid and University of Chicago



https://confluence.grid.iu.edu/display/CON/Home

Spaces Browse

Contents

- 1 Distributed High Throughput Fundamentals
- 2 Getting Started on OSG Connect
- 3 What can OSG Connect do?
- 4 OSG Connect Data
- 5 Software Access Fundamentals
- 6 More Application Examples
- 7 Higher Level Tools
- 8 Connecting Your Campus Cycles
- 9 Getting Help and Communities
- 10 References

ConnectBook Home
64 Added by Rob Gardner, last edited by Rob Gardner on Sep 13, 2013 (view change)

Distributed High Throughput Fundamentals

- Properties of High Throughput Computing Applications

Getting Started on OSG Connect

- Registration and login instructions (or go directly to the site)
- QuickStart job submission tutorial
- Job submission to OSG Connect from a laptop
- Start a Project with OSG Connect

What can OSG Connect do?

- Scaling up to more resources: campus grids, the OSG grid, and clouds
- An example: an R (statistical analysis) run on OSG Connect
- Example: use Mathematica on OSG Connect

OSG Connect Data

- Stash: the OSG Connect storage service
- Using Globus Online with your laptop and Stash
- Transferring data with HTCondor
- Access Stash remotely from your job using HTTP
- Access Stash remotely from your job using Parrot and Chirp

More Application Examples

- Using GNU Octave (numerical computations) on OSG Connect
- Using CERN ROOT (high energy physics) on OSG Connect
- Using BLAST (gene sequence similarity) on OSG Connect

Higher Level Tools

- Using Bosco Integrated with R (BoscoR) on OSG Connect
- Using Pegasus to manage workflows on OSG Connect
- Using Swift (a parallel scripting language) on OSG Connect

Connecting Your Campus Cycles

- Submit to your campus cluster from OSG Connect

Getting Help and Communities

- The OSG Connect User Community
- The OSG Campus Researcher Club
- The OSG Campus Infrastructures Community

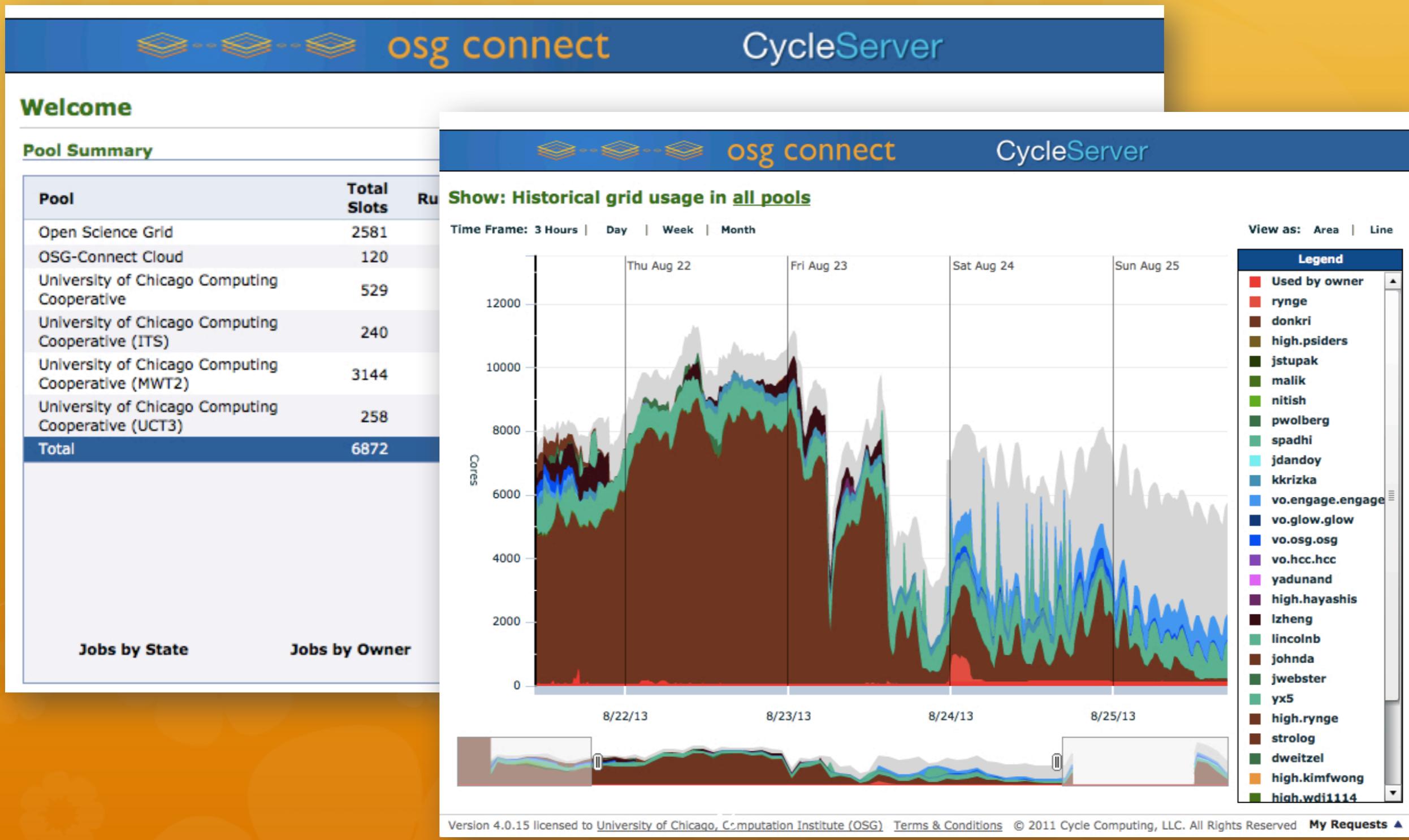
References

- Bosco: submit locally, compute globally
- Globus Online for reliable, high performance data transfer and sharing
- HTCondor for high throughput computing
- The Open Science Grid: distributed high throughput computing for science

© 2013 The Open Science Grid Consortium
Rob Gardner likes this

For help, contact connect-support@opensciencegrid.org

OSG Connect resources: campus grid, opportunistic OSG, & Amazon cloud in one place – integrated with identity and research data management services



OSG Connect Project Summary

The screenshot displays the osg connect web application interface, specifically the 'osgconnect' project summary page. The top navigation bar includes links for Support, Resources, OSG Connect (selected), Transfer, and rwg. On the left, a sidebar lists 'My Groups' (including 'MyOSGgroup' and 'osgconnect' which is selected) and various subgroups like 'osgconnect.AtlasConnect'. A 'Search' button is also present. The main content area shows the 'osgconnect' project details, including tabs for Home, Members, SubGroups, and Settings. The 'Home' tab contains a summary: 'OSG Connect offers users simple access to distributed high throughput computing resources which may be comprised of campus grids, cloud services, and the national-scale resources of the Open Science Grid using only your home campus identity. We provide a login service through which you may create and submit jobs to campus clusters, cloud resources, or to the Open Science Grid. In addition, administrators of campus grids can extend their existing resources by bridging through OSG Connect.' An 'edit' link is available for this text. Below this, a callout box states '16 subgroups → OSG Projects in OIM'. The bottom section shows 'My Projects' with a list of subgroups: AtlasConnect, biostat, CompChem, CompNeuro, ConnectTrain, EvoTheory, glass, KnowledgeSys, nescent, pshmc, rdcep, SouthPoleTelescope, swift, and UChicago-RCC. A success message 'Successfully created a new Project' is visible. A 'Create New Group »' link and an 'admin action required' note are at the bottom.

osg connect

Support ▾ Resources ▾ OSG Connect ▾ Transfer ▾ rwg ▾

My Groups

MyOSGgroup

osgconnect

osgconnect.AtlasConnect

osgconnect.biostat

osgconnect.CompChem

osgconnect.CompNeuro

osgconnect.ConnectTrain

osgconnect.Duke-QGP

osgconnect.evotheory

osgconnect.extenci

osgconnect.glass

osgconnect.KnowledgeSys

osgconnect.nescent

osgconnect.OSG-Staff

osgconnect.pshmc

osgconnect.rdcep

osgconnect.SouthPoleTelescope

osgconnect.swift

osgconnect.UChicago-RCC

UC ATLAS

UC3 Support

Create New Group »

● admin action required

osgconnect

Home Members SubGroups Settings

OSG Connect offers users simple access to distributed high throughput computing resources which may be comprised of campus grids, cloud services, and the national-scale resources of the Open Science Grid using only your home campus identity. We provide a login service through which you may create and submit jobs to campus clusters, cloud resources, or to the Open Science Grid. In addition, administrators of campus grids can extend their existing resources by bridging through OSG Connect.

edit

16 subgroups → OSG Projects in OIM

OIM ▾ Robert Gardner ▾

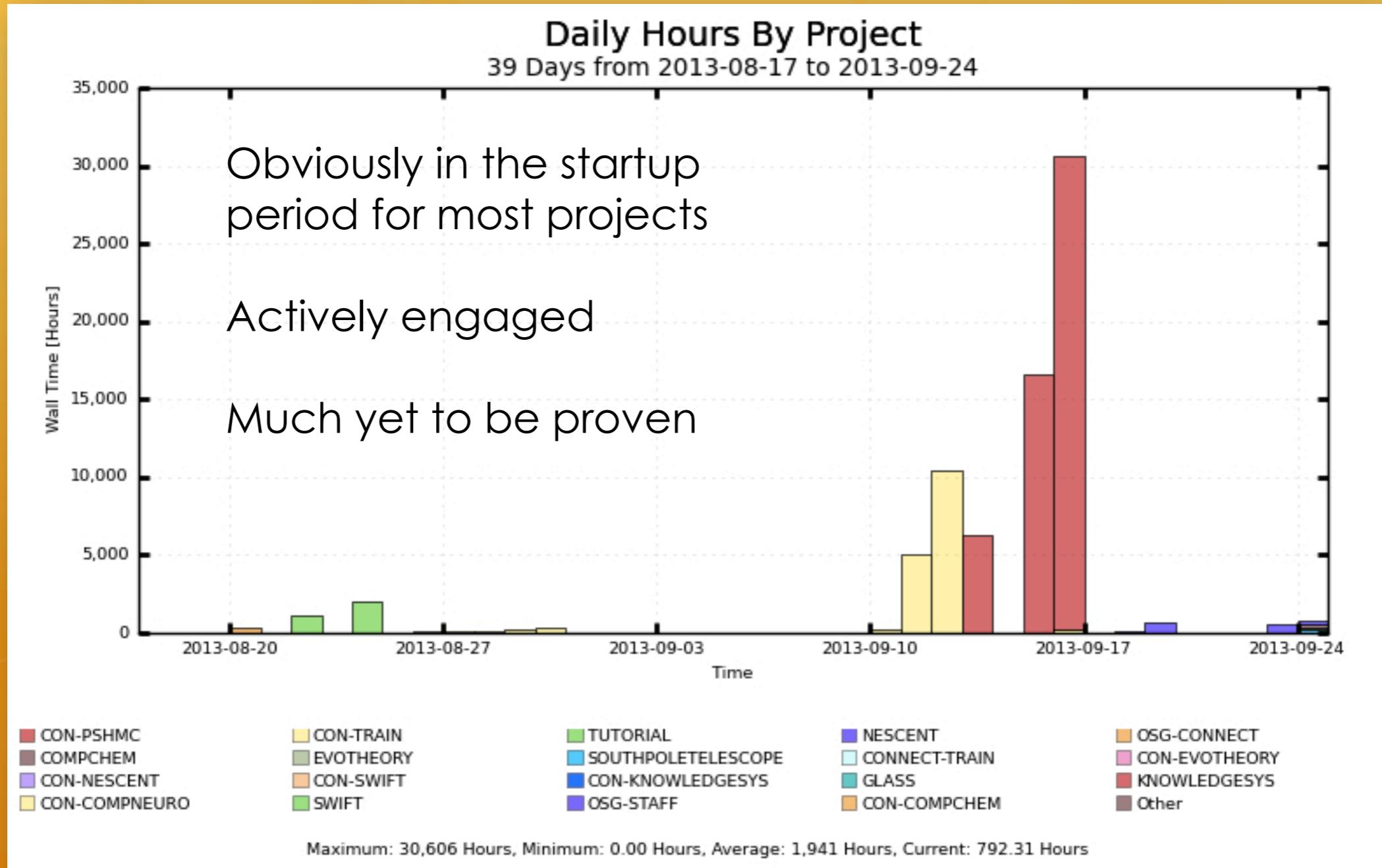
Successfully created a new Project

My Projects

+ Add New Project

AtlasConnect	biostat	CompChem	CompNeuro	ConnectTrain	EvoTheory
glass	KnowledgeSys	nescent	pshmc	rdcep	SouthPoleTelescope
swift	UChicago-RCC				

Gratia Statistics





OSG and Clouds: Summary from ET Retreat

- ◆ Distinguish clouds (a business model) and virtualization
 - ★ clouds: resource provisioning model (focus of our ET discussion)
 - ★ virtualization: where do we use VMs (should also be addressed)
- ◆ OSG is already using a resource provisioning model,
however with quite limited functionality
 - ★ this lack of functionality is clearly affecting important use cases,
like for “overruns” and “opportunistic use”:
 - ◆ “I'd like to run these workflows at Purdue, and if I'm running out, I'd like to send them elsewhere”
 - ◆ “I have an allocation at Texas Ranger, I'd like to run some fraction of my jobs there”
- ◆ ET proposes a program to overcome these limitations
 - ★ would then also enable OSG to extend its resource provisioning into the clouds, e.g. to procure commercial cycles on behalf of an OSG member
- ◆ which raises other important questions:
 - ★ what are the roles we can play in providing resources
 - ★ what are the responsibilities that we can assume



“Clouds” → Resource Provisioning

Open Science Grid

◆ Problem Statement

- ★ Can OSG assist VOs provision resources with different costs?
 - ◆ clouds are a “business model” that Resource Provides use to provide resource
 - ◆ for user, “costs” for using a specific compute resources are not necessarily money
 - ◆ other examples are CPU power, network locality, allocation
- ★ Should OSG act as an intermediary, between VOs and resources?
 - ◆ in this model, instead of OSG going to NERSC to ask for an allocation for it's users, e.g. Atlas has the allocation and OSG helps to manage and provision it to the Atlas workflows
- ★ OSG does not have a “provisioning” capability
 - ◆ although it has been discussed as a natural extension addressing a number of shortcomings
 - ◆ raises new questions of liability, accountability, audit-ability for the “money” spent?
 - ◆ can this model be extended to “enabling acquisition” of commercial cycles o behalf of an OSG member
 - ◆ maybe even be extended to the provisioning of opportunistic resources



Models To Consider

Open Science Grid

- ♦ VO establishes relationship with resource provider (credit card, allocation) and allows OSG to manage resources.
 - ★ What processes need to be in place for this? How do we add protections for both OSG and user?
 - ★ Key issue - “manage resources”. How does the VO want OSG to manage resources on their behalf? How do we express requirements and preferences?
- ♦ OSG establishes relationship with resource provider, manages some resources, and accepts responsibility for “re-selling” the allocation to VOs.
 - ★ This is the model used with NERSC currently.
 - ★ Key issue is also in managing resources.



Issues and non-Issues

Open Science Grid

- ◆ Identity management - how can we interact with the provider on behalf of the VO?
- ◆ Visibility into management - more important with “real money”.
- ◆ How do we view what is being done? How do we validate / audit activities of OSG follow VO policy? How does the VO validate its policy is expressed correctly? **Auditing capabilities.**
- ◆ What rules are needed for doing the provisioning? How do we express requirements and preferences of the VO for the resources they want? **Resource provisioning policies.**
- ◆ We need real improvements in the resource provisioning layer as it currently exists.
- ◆ An important use case is incorporating the provisioning of opportunistic resources.



Starting the BluePrint Process

Open Science Grid

- ◆ Input from OSG Consortium today/tomorrow
- ◆ State our model clearly.
- ◆ Design document of what is needed for new model.
 - ★ Document the risks to OSG.
- ◆ Document the risks to VOs.
- ◆ Build a program of work to build a provisioning service for OSG
 - ★ working with our stakeholders



OSG External Review

Open Science Grid

- ♦ the plan was to help us prepare for the upcoming external DOE/NSF review by running an internal review
 - ★ will have to develop a convincing work plan for years 3-5
 - ★ this is why we used the September slot for the ET review
- ♦ prepare for review:
 - ★ Lali: wants it joint, will start to contact NSF, date in the spring
 - ★ internal review charge: prepare ourselves for DOE/NSF review
 - ★ time scale: early next year -- early Feb?
 - ★ reviewers? comments?