Open Science Grid (OSG) Introduction for Belle & PNNL Community

January 25, 2012

Chander Sehgal
Gabriele Garzoglio
Burt Holzman

Open Science Grid Ecosystem

Consortium

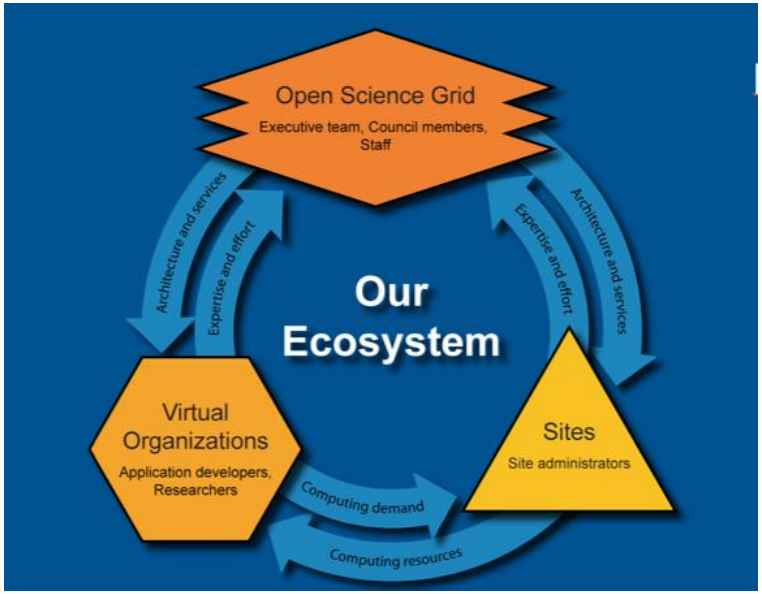
Infrastructures

Project

Satellites

Services:

- Consulting
- Production
- Software



Mission: *The Open Science Grid aims to promote discovery and collaboration in data-intensive research by providing a computing facility and services that integrate distributed, reliable and shared resources to support computation at all scales.*

http://www.opensciencegrid.org/

Introduction to OSG

Resources accessible through the OSG are contributed by the community. Their autonomy is retained.

Resources can be distributed locally as a campus infrastructure.

>30 research communities
>100 sites
>70,000 cores accessible

https://oim.grid.iu.edu/oim/home

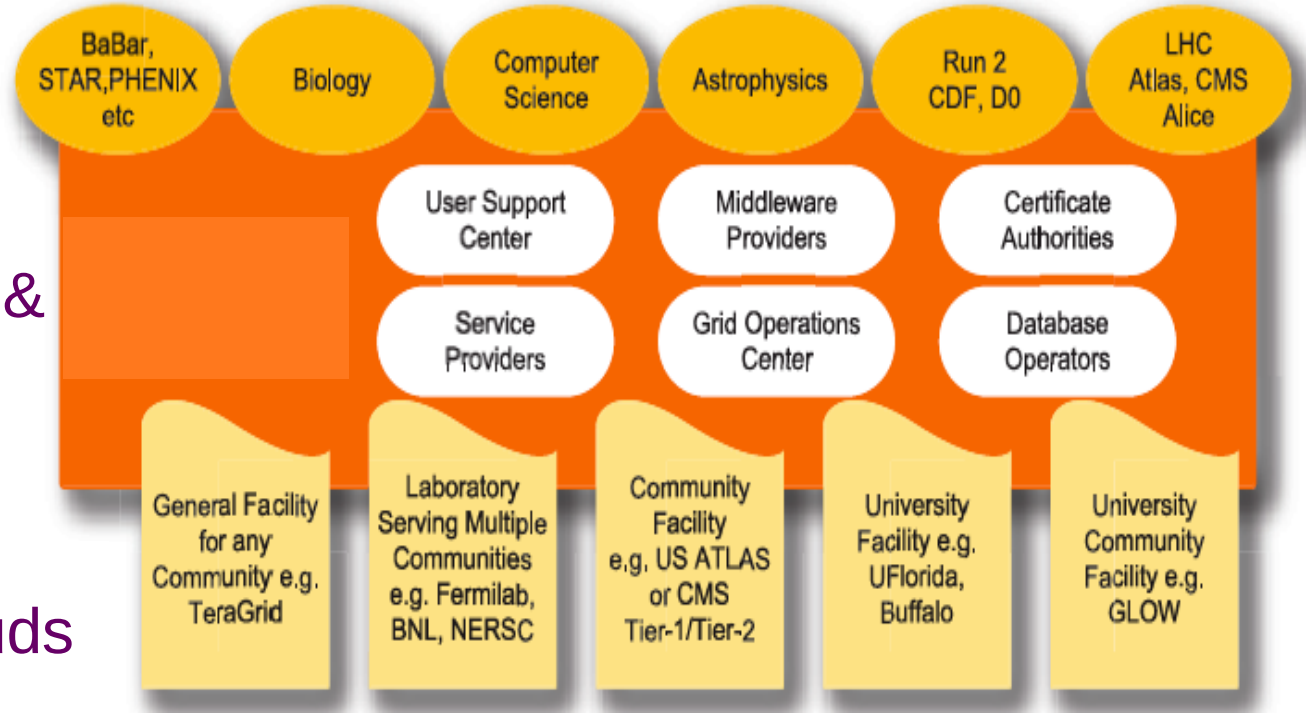
Users communities (aka VOs) and Campus Infrastructures bring:
1) Users, and/or 2) Resources

Basic Architecture

Virtual Organizations

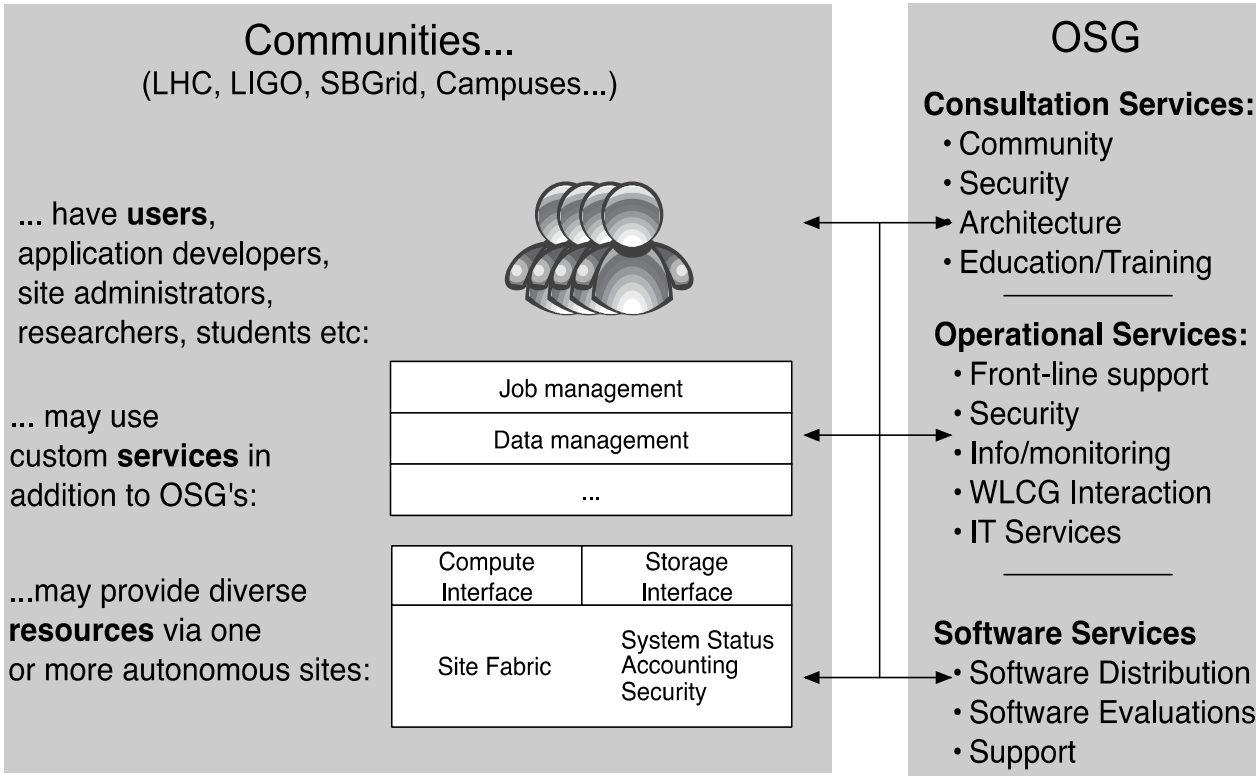
Common Services & Software

Resources: Sites, Campuses, Clouds



OSG Architecture

Architecture is driven by principles of Distributed High Throughput Computing (DHTC) captured in our Blueprint.



There is a sharing of software, operational services, and knowledge between the communities and OSG in each of these areas.

http://osg-docdb.opensciencegrid.org/0000/000018/012/OSG%20Blueprint%20v2.0.pdf

Some Contacts

OSG Staff	Responsibility
Chander Sehgal (Fermilab)	User Support Lead and Project Manager
Gabriele Garzoglio (Fermilab)	User Support – new communities integration
Burt Holzman (Fermilab) Anthony Tiradani (Fermilab)	CMS – reference for inter-operability with EU
Rob Quick (Indiana)	Operations Lead
Dan Fraser (ANL/UofChicago)	Production and Campus Infrastructure lead
Ruth Pordes (Fermilab)	Executive Director
Miron Livny (U Wisconsin)	PI and Technical Director

Introduction to OSG:
http://osg-docdb.opensciencegrid.org/0008/000839/004/OSG%20Intro%20v23.pdf

General OSG User Documentation:
https://twiki.grid.iu.edu/bin/view/Documentation/WebHome

Central ticketing system at the Operations Center http://myosg.grid.iu.edu/about
Email lists for community discussions https://twiki.grid.iu.edu/bin/view/Documentation/ContactsMailingLists

Software – the Virtual Data Toolkit

- OSG Software is packaged, tested, distributed as RPMs through the Virtual Data Toolkit (VDT). http://vdt.opensciencegrid.org/
  - The VDT has more than 40 components.
  - It also provides installation, configuration, administration tools for resources, VOs and Campus Infrastructures.
  - Alain Roy at the University of Wisconsin Madison is the Software Team Lead.
  - Software is added at the request of stakeholder needs.

Virtual Organizations (VOs)

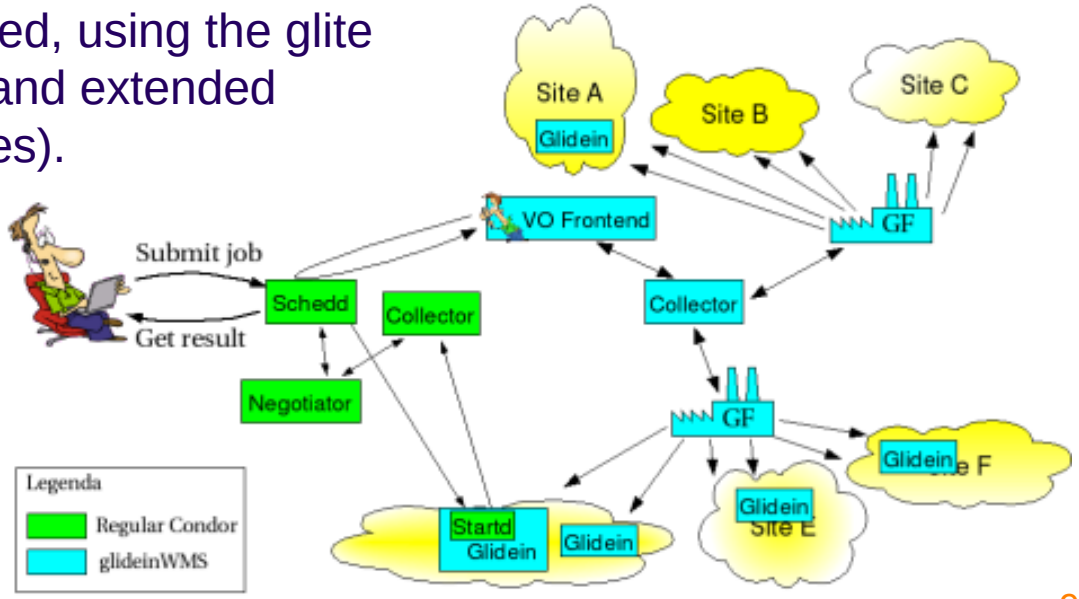
- The OSG environment is VO based.
  - VOs can be science communities (e.g. Belle, CMS) or
  - Multi-disciplinary Campus based [e.g. U-Nebraska(HCC), U-Wisconsin(GLOW)]
- OSG assumes VOs will register with, support and use multiple Grids as needed [e.g. EGI, UK National Grid Service (NGS)]
- Users can be members of multiple VOs
- Site resources are owned by one or more VOs.

Using the OSG

The prevalent OSG job submission mechanism is via Glide-ins which submit Pilot jobs to Gram gatekeepers, where the Pilots then create a Condor overlay environment across the sites and manage the user job queues and execution.

Sites also accept many other job submission mechanisms: gliteWMS*, Globus, PANDA.

Security is X509 PKI certificate based, using the glite VO Management System (VOMS) and extended attributes (including groups and roles).



* European Grid Infrastructure (EGI) Lightweight Middleware for Grid Computing

OSG Site

- An OSG Site (one or more Resources) provides one or more of the following:
  - Access to local computational resources using a batch queue
  - Local storage of and access to data using a distributed file system.
  - Remote access to data through the Storage Resource Management Interface (SRM)
  - Access to users to use external computing resources on the Grid.
  - The ability to transfer large datasets to and from remote sites.

- An OSG site offers access to these resources to grid users.
  - The policies and priority of use are determined locally by the site owners.
  - Typically the owner VO has highest priority, other “like” VOs have middle priority and the rest of the community has a lower “opportunistic” access.

What does a Site Owner Agree to?

- Offer services to at least one VO within OSG.
  - While fair access is encouraged, you need not offer services to all communities within OSG
- To advertise accessible services accurately
- To sign the “OSG Appropriate Usage Policy”
  - Do not knowingly interfere with the operation of other resources
  - Do not breach trust chains, be responsible and responsive on security issues

Requirements: System and Software

- There are no specific hardware requirements but there are recommendations as to the size of memory, local disk, server nodes depending on the expected usage.
- The VDT supports several different Linux flavors of operating system:
  - Scientific Linux5 is currently the most used.
  - We also support some packages in CentOS and Debian
  - Scientific Linux6 is on the horizon (already in use by LIGO, expected to be needed by US LHC soon).
- OSG Software Stack
  - The current software release is RPM based (OSG 3.0).
  - Previous software releases used Pacman for distribution and installation.
- Worker Nodes
  - No specific requirement
  - Expect site to install the OSG Worker Node Client which is in practice needed for most VO applications

Requirements: To support the Users

- Mapping from Grid credentials to local Unix accounts done using GUMS or Gridmapfile:
  - Some VO prefer group accounts, some require pool accounts
  - None are privileged
- Some sites use Glexec* to map the Pilot user to the actual User credentials.
  https://twiki.grid.iu.edu/bin/view/ReleaseDocumentation/InstallConfigureAndManageGUMS
- User and/or groups ids should be mapped to allow consistent file access

*https://www.nikhef.nl/pub/projects/grid/gridwiki/index.php/GLExec

Requirements: Storage

- These requirements are revisited and adjusted periodically by agreement to match current needs of the VOs.
  - Common locations provided using Environment Variables
- Provide persistent storage space for VO applications - $OSG_APP.
- Provide space for User/VO data
  - At least 10 GB per Worker Node
  - One of:
    - shared file system ($OSG_DATA),
    - special file systems ($OSG_READ, $OSG_WRITE),
    - Convenient/local Storage Element ($OSG_DEFAULT_SE)
  - Consistent data access across the resource (gatekeeper and all nodes)
- Provide local scratch space, $OSG_WN_TMP

https://twiki.grid.iu.edu/bin/view/Documentation/Release3/LocalStorageConfiguration

Requirements: Network Interfaces

- Grid Accessible Interfaces and Services
  - Public IP, name in DNS
  - Connection requirements specified in the installation and firewall documents
    - https://twiki.grid.iu.edu/bin/view/Documentation/Release3/FirewallInformation
    - https://twiki.grid.iu.edu/bin/view/Documentation/Release3/InstallRSV
- Worker nodes
  - No specific requirement, but must be able to access updates to the Certificate Revocation Lists.
  - VO are encouraged not to require persistent services on Worker Nodes
  - Most VOs use some outbound connectivity – but OSG supports sites that do not.

Open Science Grid

OSG - EU Interoperability

Federated Grids

- The OSG model is of Federated Infrastructures which have local services, software and control and Interface or Bridges to other Grids.
- OSG has an Interoperability Working Group (part of the Technology area) that tests to ensure continued interoperation works between EGI, WLCG and OSG (and other infrastructures on request).
- We currently have most experience in supporting the LHC and the Tevatron Run II experiments.

Interoperation of Software and Services

- OSG is already capable of interoperation with the EGI software stack.
  - CMS and ATLAS utilize both grids in a “bi-directional” manner
  - Jobs are sent from OSG to WLCG/EGI sites and jobs are sent from WLCG/EGI to OSG sites; Data is transferred in all directions between sites worldwide.
- The OSG and EGI Information Systems are linked “uni-directionally”
  - OSG BDII information is forwarded to WLCG Top Level BDIIs
  - WLCG BDII information is **NOT** forwarded to OSG BDII
  - EGI BDIIs may be different from WLCG ?
- Accounting interoperability is functional and is handled separately.

How to be Interoperable

- Direct submission to OSG Gatekeeper. EGI/gLite Client Tools can send jobs and data directly to OSG end points if you know the host and port.

- Can enable VOs and users using EGI/glite job submission (wms) to send jobs and data to OSG sites
  - Register VO in EGI and OSG. Configure the VO registration to send data to WLCG/EGI BDIIs.
  - Configure the supported VOs within the local site user mapping files.
  - In OSG, the GOC will issue an update for GUMS for OSG sites to use

- Have OSG jobs run on EGI sites
  - In EGI, the VO VOMS server certificate information must be bundled in .lsc format and installed on all sites willing to support the new VO. See: http://operations-portal.egi.eu/vo/registrationWelcome

Additional slides with more detail

$OSG_APP

- can be read-only mounted on the worker nodes in the cluster
- Clusters can allow installation jobs on all nodes, only on the gatekeeper, in a special queue, or not at all
- Only users with software installation privileges in their VO (if supported by the VO) should have write privileges to these directories.
- At least 10 GB of space should be allocated per VO.

Some Differences Between the Grids

- Information locations are different at the worker node level. Examples:
  - VO Software Locations
    - OSG: $OSG_APP
    - EGI: $VO_<vo-name>_SW_DIR
  - Storage Element
    - OSG: $OSG_DEFAULT_SE
    - EGI: $SE_LIST
- In general the software used differs between the grids. Example:
  - OSG: CE uses Globus GRAM5
  - EGI: CE uses CREAM

How a typical GlideinWMS based job submission works

- For every user job slot, a pilot job process starts up.
- The pilot job sends outbound TCP traffic over HTTP to a Factory service (at UCSD or IU) and via HTTP to a host at the VO Frontend (submit site; typically each VO installs its own Frontend).
- The pilot job spawns a condor startd, which spawns a condor starter.
- The startd sends outbound UDP traffic to a single port on the frontend. This is randomly chosen (per pilot) from a range of 200 ports. This can be changed to TCP if necessary.
- The starter sends outbound TCP traffic to a port on the frontend. This is chosen randomly (per pilot) from the frontend's ephemeral ports.

- Example Hosts and ports:
  - VO Frontend glidein.unl.edu (129.93.239.145). Ports: 80, 9618-9820
  - OSG Factory is glidein-1.t2.ucsd.edu. Port 8319

Source: Derek Weitzel et al. HCC VO requirements for ATLAS sites
http://www.usatlas.bnl.gov/twiki/bin/view/Admins/SupportingHCC.htmll