FERMILAB-CONF-07-217-CD

The Open Science Grid

The Open Science Grid Executive Board on behalf of the OSG Consortium:

Ruth Pordes, Don Petravick: Fermi National Accelerator Laboratory

Bill Kramer, Doug Olson: Lawrence Berkeley National Laboratory

Miron Livny, Alain Roy: University of Wisconsin, Madison

Paul Avery: University of Florida

Kent Blackburn: California Institute of Technology

Torre Wenaus: Brookhaven National Laboratory

Frank Würthwein: University of California, San Diego

Ian Foster, Rob Gardner, Mike Wilde: University of Chicago

Alan Blatecky, John McGee: Renaissance Computing Institute

Rob Quick: Indiana University

Corresponding Author: ruth@fnal.gov[1]

Abstract. The Open Science Grid (OSG) provides a distributed facility where the Consortium members provide guaranteed and opportunistic access to shared computing and storage resources. OSG provides support for and evolution of the infrastructure through activities that cover operations, security, software, troubleshooting, addition of new capabilities, and support for existing and engagement with new communities. The OSG SciDAC-2 project provides specific activities to manage and evolve the distributed infrastructure and support its' use. The innovative aspects of the project are the maintenance and performance of a collaborative (shared & common) petascale national facility over tens of autonomous computing sites, for many hundreds of users, transferring terabytes of data a day, executing tens of thousands of jobs a day, and providing robust and usable resources for scientific groups of all types and sizes. More information can be found at the OSG web site: www.opensciencegrid.org.

1. The Open Science Grid

The Open Science Grid Consortium consists of many research organizations working together to maintain and evolve a national petascale distributed facility for science. The OSG distributed facility is providing the backbone for the Large Hadron Collider (LHC), LIGO, STAR, and Tevatron Run II experiments distributed data handling and analysis systems. These applications are driving the scale of the OSG. The OSG enables other science and research groups to make opportunistic use of available

[1] Fermilab is supported by the U.S. Department of Energy under contract No. DE-AC02-76CH03000.

resources that they do not own and to easily add and use their own resources on the common infrastructure. The majority of the science communities participating in the Consortium are global in nature. The OSG is also working on gateways to other distributed infrastructures at the campus and regional level as well as internationally. The OSG Consortium includes: the science communities themselves—from single researcher to several-thousand person collaborations; DOE laboratory and DOE/NSF university computational facilities and projects providing the networking, processing and storage resources; and the middleware communities providing the software technologies used.

The Open Science Grid project is a joint SciDAC-2 and NSF project to deliver specific activities managing and evolving the distributed infrastructure and providing support for its' use. The collaboration across DOE and NSF has grown organically over the past few years between the SciDAC-1 Particle Physics Data Grid (PPDG)[1] and NSF ITR Grid Physics Networks (GriPhyN) [2] and International Virtual Data Grid Laboratory (iVDGL) [3] programs. The OSG project currently consists of sixteen institutions [4](to be expanded to eighteen in the next year) and collaboration with eight "external projects"[5] that the OSG project depends on for software (both middleware and application), base infrastructure (e.g. networks, certificate authorities), and interoperation with other projects.

The Consortium members contribute the resources available to the OSG (see Table 1). The owners of the resources control their use, with an expectation that 10-20% are on average available for opportunistic use, and with policies such that OSG members can use any unused cycles.

	CPU MillionSpecInt2000s (MSI2K)[2]				**Cache Disk in Petabytes:**			
	2006	**2007**	**2008**	**2009**	**2006**	**2007**	**2008**	**2009**
ATLAS	3	5	14	24	1.1	2.6	7.6	11.8
CMS	4	8	16	22	1.0	2.5	4.5	4.9
LIGO	4	5	6[3]	6	0.2	TBA	TBA	TBA
STAR	2	3	6	12	0.04	0.06	0.1	0.2
DOE facilities: Brookhaven, Fermilab & NERSC; + University groups	10	13	17	22	1.0	1.0	1.4	1.9
Total	**23**	**34**	**59**	**86**	**3.3**	**6.1**	**13.6**	**18.8**

Table 1: Resources Committed to the OSG

Thus the existence of the OSG does not obviate the need for the purchase of hardware and building of computational facilities by and for each science community. The scope of OSG is to:

- Enable scientists to use a greater % of the available compute and storage cycles.
- Help scientists to use distributed systems and software with less effort.
- Enable more sharing and reuse of software and reduce duplication through providing effort in integration and extensions.
- Establish an "open-source" community working together to communicate knowledge and experience and reduce overheads for new participants.

The benefits come from reducing risk in and sharing support for large, complex systems, which must be run for many years with a short lifetime workforce. And also from leveraging the expertise

[2] Conversion from MSI2K is 1500-1700 SI2K/CPU in 2008.

[3] LIGO will have 3,000 CPUs in 2008. In 2008 we estimate: 53 MSI2K = 26,000 CPUs; 74 MSI2K = 37,000 CPUs;

and support for such systems to enable new communities to more easily participate in distributed science including:

- Savings in effort for integration, system and software support,
- Opportunity and flexibility to distribute load and address peak needs.
- Maintenance of an experienced workforce in a common system.
- Lowering the cost of entry to new contributors.
- Enabling of new computational opportunities to communities that would not otherwise have access to such resources.

To be clear, however, the deliverables and milestones of OSG are driven directly by the needs of the current set of scientific stakeholders and evolve through balancing of their needs and those of the new communities to the available effort. This has worked well to date because of the contributions and commitment of the stakeholders themselves, the fact that the initial stakeholders needs are large and thus the impact of adding small new communities is low, and that the OSG organization has managed to maintain an energetic and engaged collaboration across all its participants towards the common goals.

The OSG Grid Operations Center at Indiana University provides front line support for all areas of the OSG and the OSG web site [6], document repository and @Work Twiki site give a lot of information about all aspects of the facility. We will only touch on a few representative areas of activity here.

2. Use of the OSG

There are more than forty active computational sites on the OSG. The infrastructure supports job throughput of more than a hundred thousand CPUhours a day and supports several hundred users. About twenty of the sites are part of the US LHC distributed data handling and analysis systems (Brookhaven and Fermilab Tier-1, University Tier-2s). These sites are supporting ongoing data distribution at aggregate of tens of terabytes a day (see Figure 1)

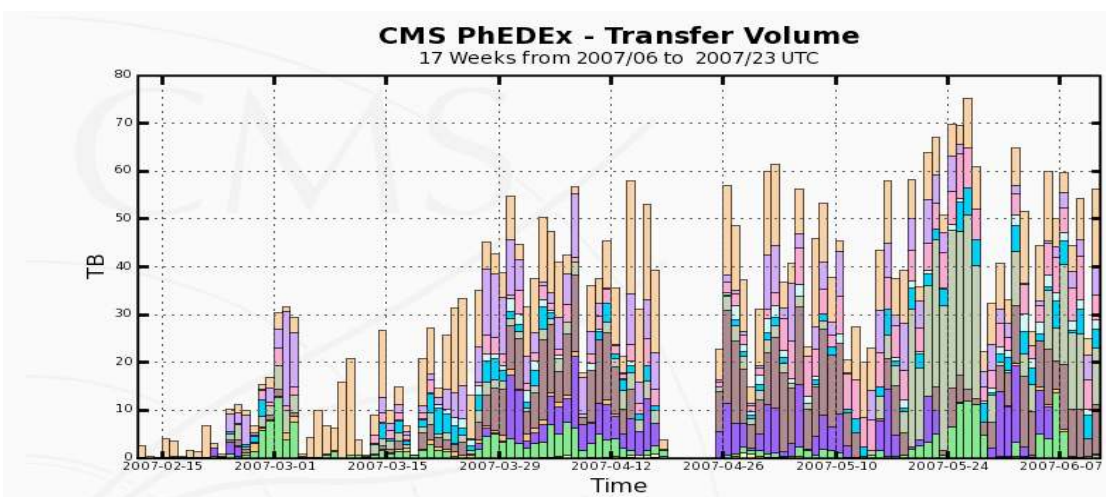

Figure 1: CMS Data Transfer on OSG (the software was being upgraded during the April downtime).

2.1. Initial Users of the OSG

Four sites are owned by LIGO and are being used for transitioning analysis codes from the existing LIGO data grid to full production on the common infrastructure (see Figure 2)

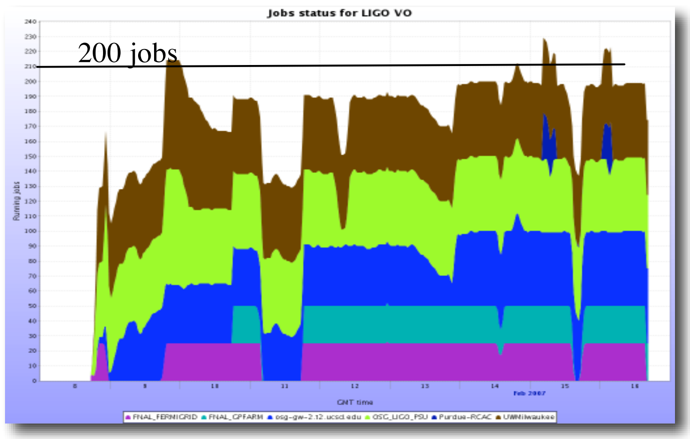

Figure 2: LIGO Inspiral Analysis on OSG, Feb 2007

Four sites are owned (or partially owned) by STAR and are being used to bring STAR data distribution, simulation and production codes. The Tevatron experiments are making good opportunistic use of the OSG: CDF runs monte-carlo simulations opportunistically across six OSG sites and analysis across four OSG sites locally at Fermilab; And over the past six months D0 has requested OSG support for re-processing a 500 million event dataset for the summer conferences in 2007 – about which more later.

2.2. Bringing new users onto the OSG

The Open Science Grid (OSG) engagement activity works closely with new user communities over periods of several months to bring them to production running. These activities include: providing an understanding of how to use the distributed infrastructure; adapting applications to run effectively on OSG sites; engaging in the deployment of community owned distributed infrastructures; working with the OSG Facility to ensure the needs of the new community are met; providing common tools and services in support of the new communities; and working directly with and in support of the new end users with the goal to have them transition to be full contributing members of the OSG. In the nine months since the start of the OSG project engagement activities have resulted in the following use of OSG[7]:

- Adaptation and production running opportunistically using more than a hundred thousand CPUhours of the Rosetta [8] application from the Kuhlman Laboratory in North Carolina across more than thirteen OSG sites which has resulted in structure predictions for more than ten proteins. We have so far tested the robustness of the system to the submission of up to about three thousand jobs simultaneously (see Figure 3)

Figure 3:Rosetta Job Submissions that then run over >13 sites on the OSG

- Production runs of the Weather Research and Forecast (WRF)[9] application using more than one hundred and fifty thousand CPUhours on the NERSC OSG site at Lawrence Berkeley National Laboratory
- Improvement of the performance of the nanoWire application from the nanoHUB [10] project on sites on the OSG and TeraGrid, such that stable running of batches of five hundred jobs across more than five sites is routine. Work was also done in support of nanoHUB scientists to use OSG resources to run BioMoca simulation jobs last year and the first couple months of this year.
- Production running using more than twenty thousand CPU hours of the CHARMM [11] molecular dynamic simulation to the problem of water penetration in staphylococcal nuclease using the ATLAS workload management system, PANDA [12], and opportunistically available resources across more than ten OSG sites (see Figure 4).

2.3. Training

The OSG trains new entrants to distributed computing through hands-on workshops, online tutorials and integrating and publishing training material and courseware. These schools build on the successful iVDGL summer grid workshops that were held annually for the past four years. The OSG training schools continue to be very well received. Two training schools [13] were held in Chicago and Texas in the first nine months of the project. We have also provided teaching materials as well as instructors to international schools in Argentina and a workshop in Taiwan. There are several more schools planned for the remainder of 2007.

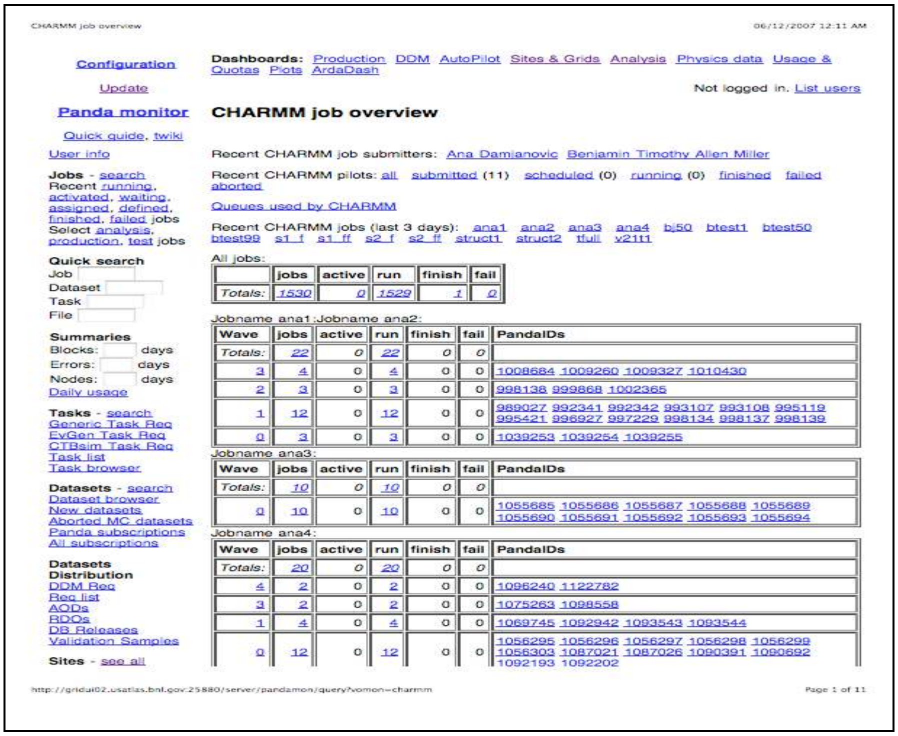

Figure 4: PANDA CHARMM monitoring

2.4. Problems Encountered

There are several problems we have run into with the current users of such a heterogeneous multi-site facility, in particular:

- *Ensuring that users can run across many sites*. Each site is autonomously managed and has a configuration tailored to the local needs. When a user first starts to use a site we often find uniqueness in the specific libraries, environment variables, shared file systems that need to be “tweaked”. We extend our test programs each time we run into these problems – and we are gradually seeing improvement.
- *End-to-end communication of and response to downtimes, failures and errors:* These problems can range from a site saying it is going down for a few days for maintenance and the user not realising this until several thousand jobs have been submitted and failed to the “black hole”, to a completely generic error code from some layer of the middleware that could indicate a problem in any one (or multiple) of a large set of software layers and services.

3. Implementation Architecture

3.1. Sites and VOs

A Site is a set of processing and/or storage resources and/or services co-located and centrally administered. A Virtual Organization (VO) is an organization that includes people using the resources (users, developers, administrators, and managers), the services needed by the organization and the resources owned by the organization. The OSG architecture [14] defines interfaces between sites and VOs to the common infrastructure. The OSG provides an integrated and tested reference set of software the Virtual Data Toolkit (VDT) [15] for the sites and the VOs to use to interface to the OSG distributed facility. Both sites and VOs have responsibility and authority over the resources, software and services that they own and they control and mange their use.

The OSG implementation architecture is cognizant that any resource may be supporting use through multiple interfaces—from local submission and access, from the OSG, and from other similar infrastructures such as Campus Grids (e.g. FermiGrid [16]) or other national infrastructures such as TeraGrid [17]. Similarly, the implementation architecture is cognizant that any VO may be using multiple infrastructures simultaneously and may have a deep set of (sometimes complex) shared software and services that are specific to the VO and operate across these infrastructures. These are additional drivers to the model that sites retain local control and management for all use, services and processes, and that VOs have control and management over their internal processes, priorities and use. In addition to levels of service and resource use being agreed between resource owners and users, the site and VO processes are implemented to support sharing and opportunistic use of the resources accessible to the OSG.

3.2. OSG Services

The OSG provides common services across the distributed facility in support of VOs and Sites: monitoring, validation and information about the full infrastructure; tracking of any and all problems and ensuring they are resolved; the Virtual Data Toolkit software packaging and support; integration and testing facilities; security; troubleshooting of the end-to-end system; and support for existing and new user communities. The OSG also provides effort to bring new services and software into the facility and to collaborate with the external projects, as well as documentation and training of site and VO administrators and users.

3.3. Issues with the Architecture

Issues we are facing with the current architecture are related to:

- The immaturity of the services and components to defend themselves against overload and unintentional misuse by the application software.
- The impossibility of testing all configurations in such a heterogeneous environment before putting new software into production.
- The lack of software to support dynamic sharing of storage on a distributed set of resources.
- The lack of the infrastructure to support (iteratively) “sub-VOs” (sub-groups within a community) in a manner equivalent to VOs. This is necessary to allow management of the resources within a community (where one of these communities is the OSG itself).

4. D0 Re-processing: an example of Opportunistic Use.

D0 is one of the two currently running experiments at the Tevatron in Batavia, Illinois. D0 currently has more than two and a half petabytes of raw and processed data on archival tape. D0’s own resources are committed to the processing of newly acquired data and analysis of these datasets as they are processed. In November 2006 D0 asked to use fifteen hundred to two thousand CPUs for up to four months for re-processing of an existing dataset (~500 million events) to deliver science results for the summer conferences in July 2007. The OSG Executive Board estimated there were currently sufficient opportunistically available resources on OSG to meet the request; we also reviewed that the local storage and I/O needs could be met. The Council members agreed to contribute resources to meet this request. D0 had several months of smooth production running using more than a thousand CPUs and met their goal by the end of May.

4.1. D0 Production Running

The steps to achieving this included: D0 testing of the integrated software system from December to February; OSG staff and D0 working closely together as a team to reach the needed throughput goals during February and March; working through detailed problems at more than twelve OSG sites as well as sites on the EGEE, Canada and UK grids; support for sustaining the throughput during April and May. The goal was achieved by the end of May. OSG contributed over half of the total events processed (286 million events) using more than two million CPUhours (see Figure 5) of opportunistically available cycles by three hundred thousand jobs, the average job length being between six and eight hours. During the reprocessing 48TeraBytes of data were moved from the central archive at Fermilab to the OSG sites and 22TeraBytes was moved from the OSG reprocessing sites back to the central archive.

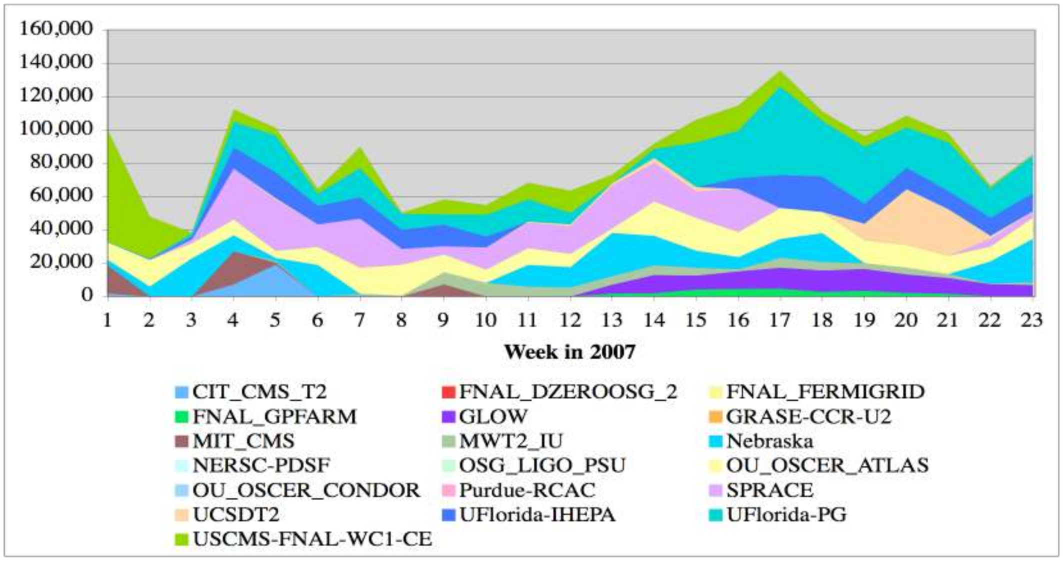

Figure 5: D0 Re-processing on OSG: WallClock Hours/week

4.2. Lessons Learned

This production activity was a very instructive experience in learning how OSG as an organization, as well as an operating infrastructure, can handle such requests. The ramp up time , as well as the unique problems experienced at each site as the VO's application was commissioned, can clearly benefit from improvement. Additional problems and overheads were experienced scaling the whole system up to the throughput needed. As explained above we faced two main problems: Overall root cause analysis and troubleshooting of the quite different hardware configurations at the sites—especially data storage architectures—was a challenge; and during the steady state operations phase we found inefficiencies in our processes for notification and responding to site downtimes and problems.

While the activity was a success there are several issues that we are taking under consideration for the future:

- Sites do provide substantial sharing of their resources and VOs can rely on significant effective throughput from resources they do not own.
- Ongoing teamwork between the application and infrastructure groups is essential in such a complex software and hardware environment.
- We need to start preparing now for the future where the resources on OSG are oversubscribed and prioritization across the infrastructure is needed.

5. Benefits from a Common, Integrated Software Stack

OSG software releases consist of the collection of software integrated and distributed as the Virtual Data Toolkit (VDT) (see Figure 6) with a thin layer of additional OSG specific configuration scripts. Modules in the VDT are included at the request of the stakeholders. The Condor [18] and Globus [19] software provide the base technologies. VDT includes about thirty additional modules, including components from other computer science groups, the Enabling Grids for EScience (EGEE) [20], DOE Laboratory facilities (Brookhaven, Fermilab and LBNL), and the application communities themselves, as well as standard open source software components such as Apache and Tomcat. OSG also supports the VDT for external projects. For example, VDT is used by the EGEE and Australian distributed computing infrastructures. Additionally, in support of interoperability across their infrastructures, the OSG and TeraGrid software stacks include aligned versions of the Condor and Globus software.

The VDT provides a reference software stack for use by OSG sites. Once the software is installed a site supports remote job submission, shared storage at a site, data transfer between sites, has services to manage priorities and access between VOs, and can participate in the OSG monitoring, validation and accounting services. OSG supports use of the reference software. Sites must provide the common interfaces to OSG services but the actual implementation is not dictated. The VDT also provides client libraries and tools for the applications to use to access OSG resources and services.

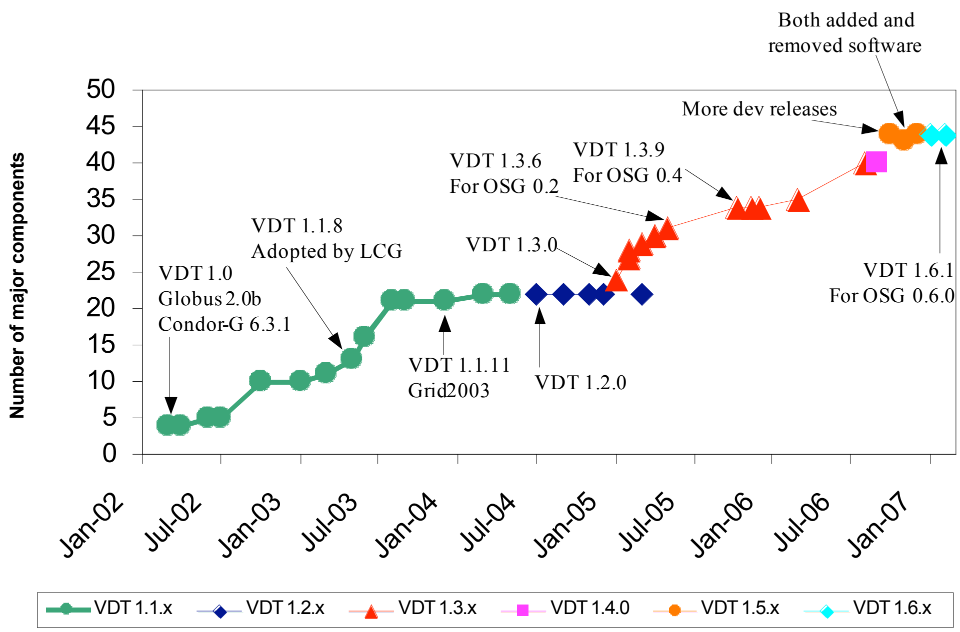

Figure 6: Timeline of VDT Releases

5.1. Software Release Process

To ensure that the operating infrastructure remains stable and robust, a careful process is followed to deploy new versions of the OSG software stack on the production grid. Before a new release of the VDT, all components are built up to ten Linux platforms (as well as some limited builds on AIX and MacOSX) using the NSF Middleware Initiative build and test facility [21] and, by re-using builds, the VDT tested on up to fifteen platforms. The validity of the integrated release is checked—for example many components use Apache web server and we check that usage is compatible—and validated on a small number of well-controlled sites (the Validation TestBed, VTB). It is then tested on a many (~15) site infrastructure, the Integration Testbed (ITB). The work of the integration activity is to check not only the individual services and software components but also that the system and VO applications work at scale.

Many of the VOs have application specific software and services layered above the OSG software stack. Most VOs install their software using Grid methods and some do so dynamically when they are submitting their jobs. The OSG does not constrain the VO software and services except if they have negative security or performance impact on the operation of the OSG itself. The integration activity is responsible for testing the complete matrix of horizontally integrated OSG software and vertically integrated VO software.

5.2. Software Patching and Auditing

The OSG also provides tools for incremental installation of new releases and patches, verification of installed configurations, and for functional testing of the sites. The OSG operates Grid-wide services for monitoring, information and accounting. We are currently developing an extended system of functional tests which will enable site administrators to more easily monitor and manage their own installations, as well as alert them when problems occur.

Security and robustness of the software as deployed is a significant concern and an ongoing area of activity on the production infrastructure. We are developing tools to audit the use of and probe the software interfaces to determine overload and "denial of service" conditions.

5.3. Challenges with Software Evolution and Support

The following needs for software support are challenging in the OSG environment:

- Quick and efficient patches of the OSG stack and redeployment in response to security notifications.
- Balancing the amount and effort spent on testing with the need to get new services "to the user" quickly. The stability of the resulting infrastructure becomes at risk since testing invariably does not cover the full set of usage patterns.
- Adding functionality (driven by the user communities) balanced with the need to make the software stack minimal and low impact.
- Integration of diverse software components from multiple software suppliers with different levels of development maturity and different release cycles.

6. Operational Security and the Security Infrastructure

OSG is well aware of the essential and integrated nature of security operations and management. We have comprehensive risk analysis and security plans [22]. We respond to software security notifications by a prompt analysis of the problem and assigning high priority to patches and fixes. Over the past year we have had about ten such notifications which have resulted in new software, downloads within between a day and a week or two. The VDT, Condor, Globus, and EGEE software teams communicate security risks as soon as they are identified and work together on patches and solutions. The collaborative nature of the OSG means that communication is natural and happening all the time and security is part of the day-to-day normal processes. OSG has mechanisms to deny user's access to sites and resources. The grouping of users into VOs gives us a small number of well-identified responsible managers who control user entry to the infrastructure. This leads us to a model of trust between sites, VOs and the OSG, with delegated trust between the VO and the end users.

6.1. OSG Security Infrastructure

The OSG security infrastructure is based on: X509 user, host and service certificates gained through one of the International Grid Trust Federation accredited Certificate Authorities; user identity proxy certificates obtained through the VO Management Service (EGEE VOMS) which provides checking of the user as part of a VO; management of extended certificate attributes to assign "roles" to a particular access by a user; flexible definition of mapping of user certificates to accounts and ACLs (access control lists) on a site; and policy (including blacklist) enforcement points at the site (processing and storage) services themselves. The specific implementations are shown in Figure 7.

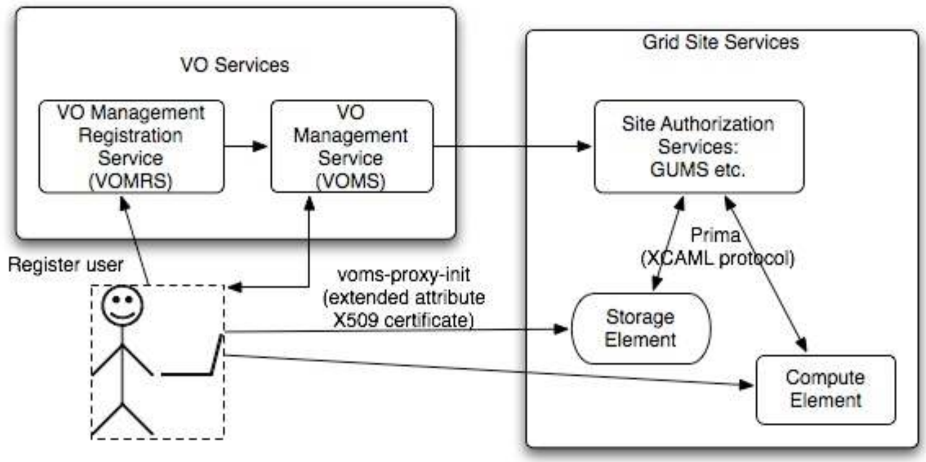

Figure 7: The OSG Authorization Infrastructure [23]. The VO maintains user membership information, while Sites control access and privileges to resources.

6.2. Problem Areas

- Not all sites yet implement these authorization components. This restricts the options available for VOs to manage their policies.
- There are gaps in the infrastructure which prevent smooth communication of VO policies to all sites. (And sites are not required to execute the VO requests).
- The creation and maintenance of accounts is a manual process for the site administrators.
- The validation and trust infrastructure between services and software components is lacking.
- Certificates with multiple attributes are not handled consistently. This will be solved in the next version of the OSG software release in August 2007.

7. Job Management and Execution

OSG sites present interfaces allowing remotely submitted jobs to be accepted, queued and executed locally. The priority and policies of execution are controlled both by the VO and the site itself. VO policies are defined through "roles" given to the user through the VOMS service. Site policies and priorities are defined through mapping the user and their roles to specific accounts used to submit the job to the batch queue. OSG supports the Condor-G job submission client which interfaces to either the pre-web service or web services GRAM Globus interface at the executing site. Job managers at the backend of the GRAM gatekeeper support job execution by local Condor, LSF, PBS, or SGE batch systems.

7.1. Resource Selection

Stand-alone Condor match-making, the generalized D0 Resource Selection Service (ReSS) [24], and the EGEE resource broker (RB) provide alternate ways for VOs to automatically select which site jobs are dispatched to based on user criteria. Both RESS and the RB depend on the OSG site information services which present information about the resources using the Glue Schema [25] attributes and providers and optionally converting the information to Condor Classads.

The largest science communities on OSG use "pilot job" mechanisms to increase the VOs ability to manage the prioritization of jobs across their organizations in a manner transparent to the site, and thus increase the overall throughput. Pilot jobs are submitted through the normal Condor-G mechanism. When the VO pilot job starts execution, using a standard batch slot, it interacts with its partner "VO job scheduler" to download the application executable to be run. The VO job scheduler controls the scheduling of jobs between the users in the VO and schedules jobs to run only on those resources that are immediately ready to execute them. These user jobs execute under the identity of the pilot job submitter—which can break the policy that sites must be able to identify the end user of their resources. OSG has integrated an Apache suexec [26] derivative, the EGEE glexec module, that enables the pilot to run jobs under the identity of the originating user.

7.2. Job Scalability

During large-scale use of the OSG infrastructure we have encountered several instances of overload and also the inability of the job management and execution infrastructure to respond appropriately to overload conditions. We are working to find the root causes and working with the software providers to fix the problems. We are also improving the configurations that we recommend for OSG sites.

8. Data Transport, Storage and Access

Many of the OSG physics user communities have large file based data transport and application level high data I/O needs. The data transport, access, and storage implementations on OSG take account of these needs. OSG relies on GridFTP protocol for the raw transport of the data – using Globus GridFTP in all cases except where interfaces to storage management systems (rather than file systems) dictate individual implementations. The community has been heavily involved in the early testing of new versions of Globus GridFTP as well as defining needed changes in the GridFTP protocol.

8.1. Storage Resource Management

OSG supports the Storage Resource Management (SRM) [27] interface to storage resources to enable management of space and data transfers to prevent unexpected errors due to running out of space, to prevent overload of the GridFTP services, and to provide capabilities for pre-staging, pinning and retention of the data files. OSG currently provides reference implementations of two storage systems the LBNL Disk Resource Manager and dCache [28].

In addition, because functionalities to support space reservation and sharing are not yet available through grid interfaces, OSG defines a set of environmental variables that a site must implement and a VO can rely on to point them to available space, space shared between all nodes on a compute cluster, and for the use of high-performance I/O disk caches.

8.2. Challenges in Data Storage and Access

The major challenges for data storage and access are to provide implementations of managed storage areas that:

- Are easy to install, configure and support;
- Support the full range of disk storage resource sizes from several to hundreds of TeraBytes;
- Provide guaranteed and opportunistic sharing of large amounts of space between and within VOs;
- Deliver hundred terabyte data sets on-time;
- Provide applications high throughput access to data at a local or remote site.

We plan to improve this situation during the next year by new releases of software and a focus on configuration, deployment and tuning of data management on and across OSG sites.

9. Gateways to other Facilities and Grids

We are seeing a rapid growth in the interest and deployment of shared computational infrastructures at the local and regional level. We are also seeing a rapid growth in research communities' needing to move data and jobs between heterogeneous facilities and build integrated community computational systems across high performance computing (HPC) facilities and more traditional computing clusters.

OSG can provide interfaces to these HPC facilities to support these use cases. For example, the NERSC facility is becoming a site on the OSG providing MPI based resources for OSG VOs.

The OSG includes software and support for campus and regional organizations to form their own locally shared facilities and distributed infrastructures which then gateway to the OSG. Three examples participating in OSG are the Fermilab Facility Grid (FermiGrid), the Grid Laboratory of

Wisconsin (GLOW) [29] and the Purdue University campus-wide infrastructure. Each local organization contributes to and manages the gateway between the local facility and the OSG. The implementation principles of the gateways for the first two differs:

- On FermiGrid each VO interfaces their own resources directly to the OSG and then accesses them through their regular remote access mechanisms. FermiGrid interfaces a gateway to the OSG which dispatches jobs from all other OSG VOs automatically across the complete set of resources (see Figure 8).
- GLOW provides a single interface to the OSG. Local job and data submissions are handled through the existing local mechanisms and then automatically "uploaded" to OSG resources under control of the GLOW gateway.

The campus infrastructures provide multiple access mechanisms to the processing nodes so that computing cluster appear as local Condor pools as well as being accessible by OSG and, in the case of Purdue, TeraGrid.

The OSG also federates with other large infrastructures—notably the TeraGrid and EGEE—by providing gateways between them and OSG, and supporting groups to submit jobs across and move data between them. For example, the OSG collects information from the resources and publishes them in the format needed by the EGEE. The CMS VO "Resource Broker: job dispatcher then submits jobs transparently across EGEE and OSG resources.

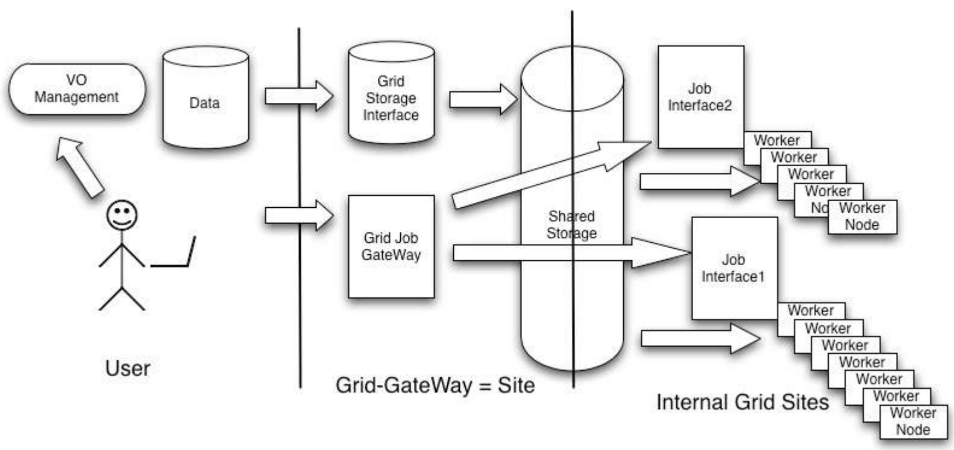

Figure 8: Accessing the FermiGrid Campus Grid from OSG

10. Participating in the OSG

New organizations contribute to OSG by providing resources, using the infrastructure, working with the communities to provide software, participating in training and documentation activities, and/or participating in the security, troubleshooting, or other OSG activity areas. The overhead to participation is low and the benefit is matched to the principle that "what you get out depends on what you put in". An organization registers with the Grid Operations Center and provides contact and planned usage information. The OSG staff then helps to interface the resources to the OSG and provides support for the VOs usage.

For this model to scale as OSG gets larger new communities are expected to ramp up to contribute in some aspect of the collaboration to match the benefits they receive. Thus the large science collaborations all make contributions to OSG from within their projects, partner with OSG on joint activities, and embrace collaboration with OSG as part of their roadmap.

10.1. Other SciDAC-2 Projects

OSG works closely with the SciDAC-2 Center for Enabling Distributed Petascale Science (CEPDS) [31] in all three areas of its work. The managed movement of data is clearly a key challenge for existing and future OSG communities. Members of the OSG management team are also part of the CEPDS data area. We expect that CEDPS deliverables such as the Managed Object Placement System (MOPS) will be used on OSG in the next year or two. Troubleshooting (in the CEDPS scoped to forensics not the real time auditing for security problems) is already working closely within the OSG troubleshooting and security activities. These OSG activities are providing requirements and the OSG integration activity is providing the testbed for deployment and evaluation. CEDPS scalable services area is working with STAR to provide dynamic workspaces that will provide a VO scoped environment over multiple OSG resources.

OSG supports the Earth System Grid (ESG) [31] in components of their software, and is working with the project to understand whether reliance on VDT will provide benefit and efficiencies in their environment.

11. Challenges in our Future

The Open Science Grid is charting new territories in several areas: Community driven production support and evolution of a large distributed infrastructure with coordination but no control over the "end-points"; Security and operations across autonomously owned and managed facilities which scale from small university department clusters to large leadership class high performance computing facilities; Maintaining full production use while updating software and adding services and ensuring robustness to failure, overload and (unintentional and intentional) misuse; Developing a sustainable economic model for usage and growth of the infrastructure.

OSG is also well aware of the challenges and complexity of how to measure success. Success is not just the number of jobs executed and the amount of data transferred and stored. Success has to be measured by the impact on scientific productivity and maturity of computation as a cornerstone (with experimentation and simulation) of the research portfolio. We are gradually putting in place measurements of many parameters to help us determine the impact, but we do not understand yet how to translate and analyze this information to quantify value and benefit.

The turn-on of the LHC and the LIGO and STAR upgrades will result in a ten-fold increase in the data and job throughput required on the OSG over the next three years. These will challenge the scalability, performance and robustness of the infrastructure. In our experience they will also prove a challenge to the functionality and software provided as new needs of the collaborations emerge.

In addition, our challenge will be to succeed in our commitments to our balance to bring new (non-physics) communities to production use at least every six months, provide access to new high performance and local computing, and be welcoming to participants and contributors from university, laboratory, national and international groups.

Acknowledgments

This work was supported by the Office of Science, U.S. Department of Energy, SciDAC program under Contract DE-FC02-06ER41436 and the National Science Foundation Cooperative Agreement, PHY-0621704.

References

[1] PPDG web site: www.ppdg.net

[2] GriPhyN web site: www.griphyn.org

[3] iVDGL web site: www.ivdgl.org

[4] OSG institutions: Boston University, Brookhaven National Laboratory, Caltech, Cornell University, Columbia University, Fermilab, Indiana University, ISI, Lawrence Berkeley National Laboratory, Renaissance Computing Institute, SLAC, University of California San Diego, University of Chicago, University of Florida, University of Iowa, University of Wisconsin Madison

[5] OSG External Projects: Community Driven Improvement of Globus Software (CDIGS), Center for Enabling Distributed Petascale Science (CEDPS), Condor, dCache, Data Intensive

Science University Network (DISUN), DRM, Energy Sciences Network (ESnet), Internet2, Laser Interferometer Gravitational-Wave Observatory (LIGO) Physics and the Information Frontier, Fermilab Grid accounting and security projects, UltraLight/advanced networks, US representation of the WorldWide LHC Computing Grid Collaboration.
[6] OSG web site www.opensciencegrid.org
[7] Engagement Report 2007 http://osg-docdb.opensciencegrid.org/0006/000632/001/OSGEngagementReport200706-V2.pdf
[8] Rosetta web site: https://wl.rosettacommons.org/
Kuhlman Laboratory at the University of North Carolina, Chapel Hill: http://www.unc.edu/kuhlmanpg/
[9] WRF at RENCI web site: www.renci.org/disaster/hydromet.php
[10] NanoHUB web site: http://www.nanohub.org/
[11] CHARMM Chemistry at Harvard Macromolecular Dynamics: http://www.charmm.org/
[12] ATLAS PANDA software: https://twiki.cern.ch/twiki/bin/view/Atlas/Panda
[13] OSG grid schools: https://twiki.grid.iu.edu/twiki/bin/view/Education/WebHome#Upcoming_Grid_Schools
[14] OSG Blueprint Document: http://osg-docdb.opensciencegrid.org/0000/000018/005/OSG-Blueprintv0.10.pdf
[15] The Virtual Data Toolkit web site: http://vdt.cs.wisc.edu
[16] FermiGrid web site: http://fermigrid.fnal.gov
[17] TeraGrid web site: www.teragrid.org
[18] Condor ™ project web site: www.cs.wisc.edu/condor
[19] Globus Tookit™ web site: www.globus.org/toolkit
[20] Enabling Grids for EsciencE (EGEE) web site: http://public.eu-egee.org/
[21] Metronome (formerly NMI build and test): http://nmi.cs.wisc.edu/
[22] OSG Security plan: http://osg-docdb.opensciencegrid.org/0003/000389/017/OSGSecurityPlan.doc
[23] VO Services project web site: www.fnal.gov/docs/products/voprivilege/
[24] Resource Selection Service: https://twiki.grid.iu.edu/twiki/bin/view/ResourceSelection/WebHome
[25] Glue Scheme Open Grid Forum Working Group: http://forge.gridforum.org/sf/projects/glue-wg
[26] Apache suexec: http://httpd.apache.org/docs/1.3/suexec.html
[27] Storage Resource Management Collaboration: http://sdm.lbl.gov/indexproj.php?ProjectID=SRM
[28] dCache Consortium web site: www.dcache.org
[29] Grid Laboratory of Wisconsin web site: www.cs.wisc.edu/condor/glow
[30] Center for Enabling Distributed Petascale Science (CEDPS): http://www.cedps.net/wiki/index.php/Main_Page
[31] Earth System Grid (ESG): www.earthsystemgrid.org