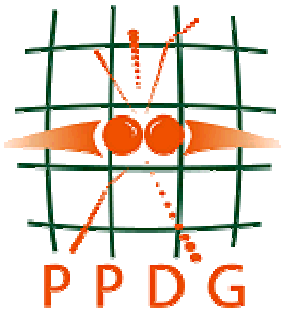
Particle Physics Data Grid: From Fabric to Physics

Semi-annual Status Report of the Steering Committee,

October 2005 – March 2006

15 May 2006



1 Project Overview

The primary focus of the project during the last six months continued to be on the Open Science Grid; hardening and extending the core services, operating the production virtual facility, running and enabling applications' usage, and planning the future of OSG. The main reason this report covers two quarters rather than one is the effort that went into developing the OSG Program of Work[1] which formed the basis of a five year proposal submitted to the NSF and the DOE SciDAC-2 program. The PPDG Common Project, described in Section 2 continues to deliver functionality for OSG services in the 0.4 and 0.4.1 releases during this period. There were several presentations at SuperComputing 2005 related to OSG (doc #s 304, 305, 306, 307, 308, 309) and many presentations at the Computing in High-Energy and Nuclear Physics Conference (CHEP 2006) in February (listed in the Appendix). The experiments, ATLAS, CMS,

[1] OSG Doc 310, http://osg-docdb.opensciencegrid.org/0003/000310/001/OSGProgramofWork.pdf

CDF, D0, STAR, continued to expand their use of OSG in resource utilization and/or functionality. Daily reports of all applications usage of OSG are available at http://grid02.uits.indiana.edu:8080/reports/daily/.

There has been increased collaboration with the EGEE in delivering to the WLCG and LHC experiments: CEMON is used in a common component for the resource selection service (section 5.1.4 below) and the PPDG use has helped harden this service for the EGEE itself; The US Grid use of VOMS continues to increase; The EGEE continutes to rely on the VDT; Work continues to provide interoperation between the TeraGrid and OSG. CDF,STAR , ATLAS and CMS jobs can be submitted to a selected TeraGrid sites. Collaborative work with the ESnet DOEGrids PKI is enhancing the usability of the grid certificate infrastructure and preparing for the OSG Registration Authority.

1.1 Documents and Presentations

OSG documents and presentations published and updated since Oct. 2005 by PPDG participants.

131-v5	OSG Accounting System Project Charter	Philippe Canal	20 Apr 2006
130-v3	Accounting System Requirements Version 2.0	Matteo Melani	20 Apr 2006
378-v2	OSG Use of PKI	Doug Olson	04 Apr 2006
379-v1	Compute Element for OSG 0.4.1	Frank Wuerthwein	31 Mar 2006
377-v1	Site Policy Template for Open Science Grid	Vikram Reddy Andem	23 Mar 2006
376-v1	VO Policy Template for Open Science Grid	Vikram Reddy Andem	23 Mar 2006
374-v1	OSG Security Issues and Activities	Bob Cowles	08 Mar 2006
373-v1	SDSC/CIP	Frank Wuerthwein	08 Mar 2006
372-v1	Distributed CMS Analysis on the Open Science Grid	Gutsche, Oliver	06 Mar 2006
371-v1	OSG-CAF - A single point of submission for CDF to the Open Science Grid	Neubauer, Mark	06 Mar 2006
370-v1	Grid Deployment Experiences: The interoperations activity between OSG and LCG.	Grundhoefer, Leigh	06 Mar 2006
367-v1	GRATIA, a resource accounting system for OSG	Philippe Canal	06 Mar 2006
366-v1	CMS Software Distribution on the LCG and OSG Grids	Michael Thomas	06 Mar 2006
365-v1	gPLAZMA: Introducing RBAC Security in dCache	Jon Bakken *et. al.*	24 Feb 2006
364-v1	An Edge Services Framework (ESF) for EGEE, LCG, and OSG	Rob Gardner *et. al.*	24 Feb 2006
363-v1	IWLSC, Kolkata India 2006	Jerome Lauret	15 Feb 2006
86-v8	OSG User Acceptable Use Policy	Security Technical Group	09 Feb 2006
340-v1	Application Requirements for OSG 0.6	Frank Wuerthwein	07 Feb 2006
332-v1	What Applications need on OSG 0.4	Frank Wuerthwein	07 Feb 2006
359-v0	Complete 0.4.1 Deployment Plans		22 Jan 2006
314-v1	OSG Management Plan	Ruth Pordes	19 Jan 2006
312-v1	OSG Grid Operations	Leigh Grundhoefer	03 Jan 2006
311-v2	Open Science Grid consortium voting procedures	Jerome Lauret *et. al.*	01 Jan 2006
300-v3	OSG Function Set 0.4 Definition	Razvan Popescu *et. al.*	12 Dec 2005
262-v3	Requirements for Grid Service Interoperability between the Open Science Grid and the TeraGrid	Stuart Martin *et. al.*	27 Nov 2005
307-v1	DASH: Database Access for Secure Hyperinfrastructure	Alexandre Vaniachine	23 Nov 2005

306-v2	ESF: Edge Services Framework for Computational Grids	Alexandre Vaniachine	23 Nov 2005
310-v1	OSG Proposed Program of Work	Ruth Pordes	21 Nov 2005
309-v1	GLOW - A Campus Grid within OSG	Miron Livny	16 Nov 2005
305-v1	How to Participate in OSG	Dane Skow	07 Nov 2005
18-v5	A Blueprint for the Open Science Grid	Blueprint	06 Nov 2005
304-v1	DZero Production Computing		04 Nov 2005
302-v4	Open Science Grid: Beyond the Honeymoon	Dane Skow	25 Oct 2005
193-v1	Edge Services	Frank Wuerthwein	11 Oct 2005

More than 30 presentations were made at the Computing in High-Energy and Nuclear Physics Conference, February 13-17, 2006 in Mumbai, India from groups participating in PPDG are shown at http://www.ppdg.net/docs/chep06-presentations.html, and listed in the Appendix.

2 The Common Project

The common project effort focused activities on integration and testing of Release 0.4 and 0.4.1 of the OSG software toolkit. This Release includes several updates to previous PPDG Common Project contributions and a few new services.

The Privilege Project products, PRIMA and GUMS, were upgraded for the 0.4 Release. Security presentations were made at the MiddleWare Security Group meeting at CERN and a plan developed for consolidating current levels of interoperability while working on new features/functions. This work was presented at GGF and SuperComputing and used as a baseline contribution to the nascent MultiGrid interoperability working group on Identity Management and Authorization.

PRIMA has been ported to and tested on 64bit linux platforms. The PRIMA modules were integrated into the GT4 Web Services GRAM framework. The software has been packaged and is distributed via VDT.

GUMS is in production throughout OSG for grid identity -> local UNIX account mapping. An updated release of GUMS was prepared to address some items uncovered under load in previous deployments. The main features have been implemented although a list of requests for additional functionality is being maintained. Communication with EGEE developed in response to their inquiries about interoperation or even the use of GUMS.

gPlazma continues to be in active development. The gPlazma and Prima java code was modified to work with the new GUMS server and more recent opensaml and axis libraries. Testing was successful for gridFTP doors and srm. gPlazma is now deployed on all gridFTP doors in the FNAL USCMS dCache and srm will follow soon.

The SRM-Tester harness (SRM functional test) has been integrated into the OSG operations toolkit and is now providing status information into the GridCat service catalog. The SRM-TESTER for SRM v2.1.2 was developed. The package is a tester program to test any SRMs based on the SRM v2.1.2 protocol. SRM-TESTER has been tested against Berkeley SRM and SRM/dCache at FNAL. Validation of Storage Elements (SEs) that use SRMs can be accomplished through the use of the SRM-TESTER in VDT. The SRM-TESTER checks not only whether the SRM is alive, but also checks getting and putting files from/to the SRM. Additional elements have been identified that need to be added to the GLUE schema of an SE for this purpose. In addition, the SRM-TESTER can check the interoperability between another SRM and the SRM being tested, by issuing an srmCopy call. SRM-TESTER for SRM v1.1 will be used. This functionality will be used by the OSG-ITB.

CERN EGEE people expressed interest in the SRM-Tester source code, and have been given access to it. The GGF-GIN (Grid Interoperation Now) group plans to use SRM-TESTER for SE interoperability testing.

Common Project effort continues to provide a leadership role in the OSG Accounting Activity. FNAL and CMS are contributing large fractions of the effort devoted to this activity. An agreement has been made

with all of the OSG Partner grids to standardize on the GGF Usage Record format for accounting records. This should facilitate interoperability and possible reuse of components between the various efforts. Work has been proceeding toward a v0.0 distribution which will be released in May. The Collector and Reporting prototype are in place and more reporting will be added once the Condor probe is finalized. The PSACCT probe will be available at the beginning of May. Work is being done to extend the installation base to each of the Condor farms at Fermilab. Work is now also on-going on the beta version of the server and on the analysis of the data being collected.

Resource Selection Service work continues focusing on a collaboration with our EGEE partners on use of the CEMon tools for publishing resource characteristics. Much work was done on the development, integration, and deployment of the Resource Selection Service for OSG. The project is ready for deployment on the OSG-ITB and is preparing the service to be integrated within the DZero job management infrastructure (SAM-Grid).

The Condor team continues to operate the OSG-ITB Grid Exerciser (GridEx) on behalf of the OSG-ITB. In this quarter, the GridEx team was able to post the first error-free OSG-ITB daily report. This was achieved by working with sites to update their local configuration, and by incorporating OSG-ITB lessons learned into the Condor Grid Universe (Condor-G) client. Additionally, the GridEx served as a validation tool for the OSG-0.3.6 release.

The OSG-ITB Clarens based Discovery Service was moved from ultralight.caltech.edu to a dedicated machine at osgitb-discovery.caltech.edu. A Discovery Service publisher was configured to push the contents of the vo-0.4.0 package describing the VOMS servers in the OSG. This has been running for over a month with no noticeable problems. The Software Discovery service continued to be used by the CMS Tier2 sites as a catalog of the CMS software available on the Tier2 clusters. There has been one feature request to allow the listing of both the gatekeeper DNS name as well as the shorter site name.

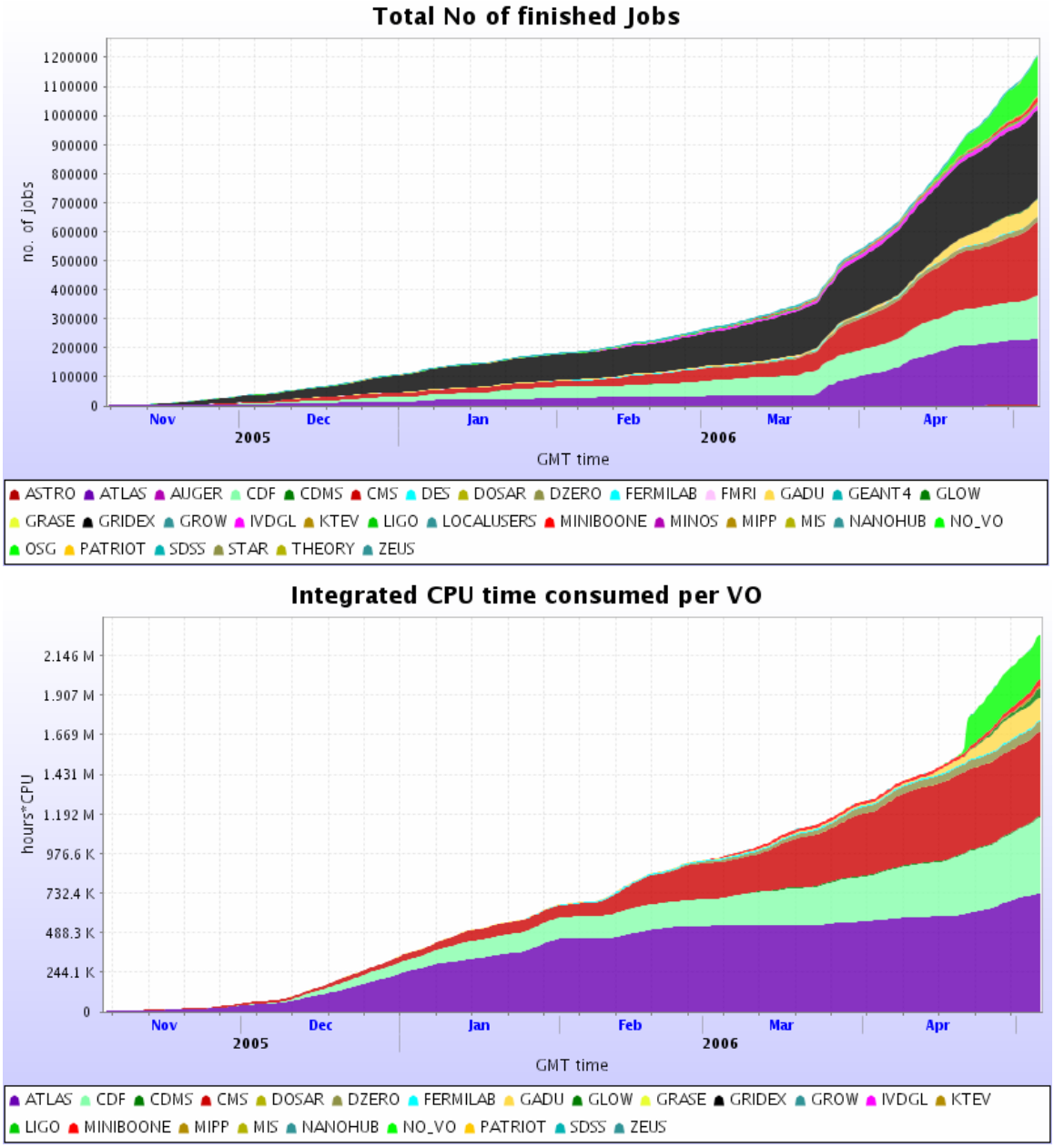
3 Open Science Grid

3.1 Current OSG Activities

The OSG had a major software release (0.4.0) in January and carried out the main share of integration work for a subsequent minor release (0.4.1) which is available in May.

The organization of the OSG Consortium advanced significantly in developing a management plan (OSG Doc 314) and filling the leadership positions. Ruth Pordes was elected Executive Director of OSG and has appointed an Executive Team that includes coordinators for the Facility, Operations, Resources, Applications, Education and Outreach. Bill Kramer was appointed chair of the OSG Council, replacing Frank Wuerthwein.

The usage of OSG connected resources is shown in the two graphs below, and show an increasing usage by PPDG experiments, and others, so that today the PPDG experiments are running > 2000 jobs concurrently.



3.2 Proposed Program of Work

A proposal to support the core OSG effort and extend the functionality to that required for a data intensive open scientific computational grid was submitted to NSF and DOE, based upon the OSG Program of Work, Document 310. The proposed work includes operational support for the OSG Facility, integration and deployment support for new software releases and services, a security plan and processes, extensions for additional (and needed) functionality, education and outreach, and also management of the execution of the program. The Title page of the SciDAC proposal is given in Appendix B, showing the participants and abstract. It is the opinion of the PPDG Steering Committee that the OSG program is the essential next phase in the evolution of their scientific computing functionality and necessary in order to capitalize upon the investments made in the PPDG project.

The following excerpt from the proposal Program of Work section givers a concise summary of the plans and goals:

Operating and maintaining the OSG Facility requires dedicated effort throughout its lifetime. The facility thrust of the program of work sustains a robust, usable infrastructure with increasing quality of service through operational and support activities; ongoing attention to security and troubleshooting; packaging, maintenance and testing of new software stacks as technologies evolve; and engagement with new communities of users.

In addition to the efforts needed to sustain a flightworthy cyberinfrastructure, the program of work attends to the needs of our users for end-to-end capabilities. Developing and deploying such capabilities requires close interactions with the domain-specific applications as well as advances in the functionality of the OSG software stack. The science driven extensions thrust of the program is structured as a set of well-defined projects that include external partners. When ready for production, the software tools developed by these projects are integrated to the VDT, tested by the integration team and deployed within the OSG Facility. Thus the program of work includes:

- *Management, operation and evolution of the distributed facility* including software integration configuration and deployment and well defined and documented middleware releases. Facility activities include system wide performance and availability monitoring and analysis, comprehensive functional testing, and in depth engagement in support of users and system administration.
- *Engagement and training programs* to actively help new entrants make their computing and storage accessible to the common infrastructure, to help new researchers use the common environment, and provide hands-on workshops and published materials for training and dissemination.
- *Provision and maintenance of an at-scale integration and validation testbed* providing a heterogeneous platform for vertical and horizontal system testing of new releases, new technologies, new capabilities, and new applications, to facilitate smooth transition into the production environment.
- *Development and integration of extensions to* develop job and workflow management, security and policy services, integration of advanced network fabric capabilities needed to meet the scientific requirements and schedules of the stakeholders.
- *Interoperation of* the OSG infrastructure with other distributed environments from local campus infrastructures to the national and international grids that are forming the transparent worldwide cyberinfrastructure.

Moreover, the success and progress of the Facility also depends on continuous growth in and contributions to compute, storage and networking capacity as well as continued enhancement in the cyberinfrastructure capabilities in the US, including:

- A comprehensive program for *Grid security research and development* to ensure the integrity and defense of the open infrastructure on which our science depends.
- External funding of *local facilities*, for the purchase and support of the compute, storage and network hardware of the facilities made accessible by the OSG infrastructure.
- An aggressive program of *network and middleware research* to ensure the necessary increases in scale and performance of the infrastructure.
- A sustained program of *middleware software development, packaging and support* to ensure the continued availability, and further development, of the core software on which OSG builds.
- The development of *applications* that can be adapted to run on a distributed, loosely coupled infrastructure.
- A sufficient *cyber security infrastructure* supporting single sign-on to access grid resources, engagement of existing computing facility cyber security staff, risk based mitigation measures.
- *Commitments from the stakeholders* to enable policy driven sharing of resources and when possible the use of common services.

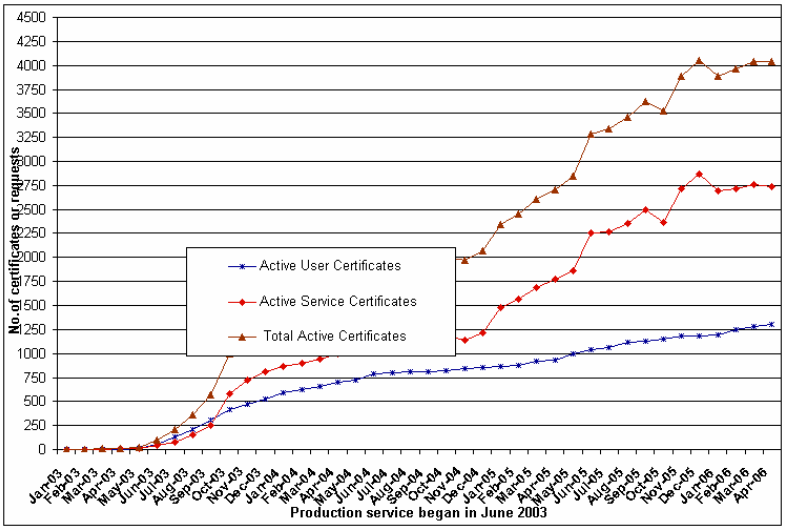
4 Collaborations

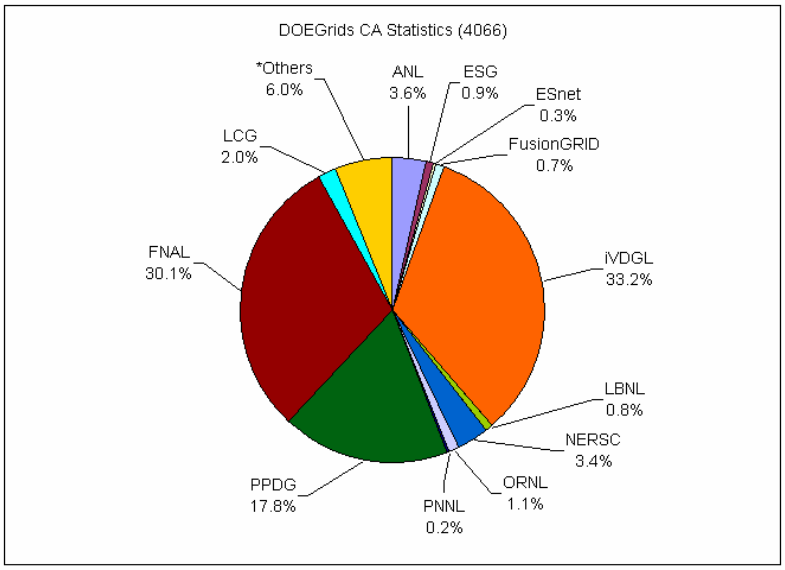
4.1 EGEE and WLCG

The Run II, Babar and LHC experiments run their applications on the EGEE as well as the grid infrastructures in the US. They ensure that the services provided on OSG and EGEE interoperate and we work with them to make the boundary between the two infrastructures as seamless as possible. OSG sites can advertise through the OSG Information System and selectively appear in the WLCG information system such that jobs using the EGEE Resource Broker are dispatched to sites on either infrastructure.

4.2 DOEGrids PKI

The DOEGrids PKI operated by ESnet continues to be an integral aspect of the security infrastructure enabling PPDG and Open Science Grid participants to access and utilize the distributed computing resources of OSG. The charts[2] below show the number of active certificates for people and for computer services, now with more than 1200 active user certificates and about 2750 service certificates. Since the certificate lifetime is limited to 12 months (and therefore no longer active) the continued growth in the number of people and computers "certified" is apparent. The second graph shows the breakdown of number of certificates by the Registration Authority (RA). Most of the certificates issued for the CDF, D0 and US CMS groups participating in OSG are issued by the FNAL RA and the remaining groups in PPDG have certificates issued by the PPDG RA. The PPDG, FNAL and iVDGL RAs today issued most of the certificates to participants in Open Science Grid, amounting to about 80% of the total DOEGrids certificates.





In anticipation of the termination of both the PPDG and iVDGL grid projects there is work in progress to establish an OSG RA with a scope to cover any participant in OSG. This RA is planned to be operational by July1, the end date for PPDG. There are some improvements in the details of the certification

[2] Graphs from http://www.doegrids.org/pages/DOEGridsCAStats.html

procedures and infrastructure associated with this in order to make the request processing more efficient and timely. These improvements and the OSG RA plans are described on an OSG web page[3].

Besides the improvements for OSG there is continued progress on the PPDG addition for usability improvements and OTP pilot study using MyProxy. Two myproxy servers have been installed, one for "production" use to support roaming access in initially, and another to be used for testing proxy renewal scenarios. The "production" server will be called myproxy.doegrids.org when the configuration is finalized but is available for use now under a temporary name. Following the testing of renewal scenarios, a renewal configuration will be supported by the production server and the second server will be re-configured to be a failover backup of the production service. In addition to the myproxy service, there are improvements to the scripting interface to the CA. The most urgent addition was an interface to the gridadmin service, whereby a pre-authorized user can request and receive service certificates without requiring additional interaction. This was added to the VDT in December 2005 (VDT 1.3.10). There is also a script developed to perform certificate revocation as well as modifications to the request script to match changes to the web interface, to include the specification of the VO of the user requesting a certificate. These improvements should in included in VDT in time for the OSG 0.6 release. There has been some delay in progress on the OTP pilot aspect of the myproxy project due primarily to the need to prepare the OSG RA but it should be possible to proceed with this aspect of the project during the next quarter using the NERSC myproxy-OTP infrastructure.

4.3 TeraGrid

5 Single Team Reports

5.1 Experiments

5.1.1 ATLAS

ATLAS released several versions of the new production simulation system, PANDA. PANDA jobs are running on several ATLAS owned sites as well as on CMS sites on the Open Science Grid. Daily usage statistics are provided with a typical 24 hours report given below:

From: wenaus@bnl.gov Subject: Panda usage summary 2006-05-11

Panda usage summary (midnight to midnight central time) for 2006-05-11

All_sites: Walltime: 10656 hours

Site: ANALY_BNL_ATLAS_1 Walltime: 81 hours

Site: BNL_ATLAS_1 Walltime: 4105 hours

Site: BNL_ATLAS_test Walltime: 4 hours

Site: BU_ATLAS_Tier2 Walltime: 811 hours

Site: IU_ATLAS_Tier2 Walltime: 440 hours

Site: OU_OCHEP_SWT2 Walltime: 1833 hours

Site: PROD_SLAC Walltime: 0 hours

Site: UC_ATLAS_MWT2 Walltime: 1423 hours

Site: UTA-DPCC Walltime: 1960 hours

[3] http://osg.ivdgl.org/twiki/bin/view/Security/OsgRaPlanning

5.1.2 BaBar

5.1.3 CMS

Michael Thomas:

Summary

Q4-2005 saw another release of jClarens which was added to the VDT as well as a gruelling SuperComputing 2005 Bandwidth Challenge and an important live GAE demonstration using a new Clarens UI. Most of Q1-2006 activities revolved around maintaining and upgrading the Caltech CMS Tier2 computing cluster. Additionally, some jClarens improvements were made and the Discovery Service got a new home.

JClarens

JClarens 0.6.1 was released on Nov. 2. Files from this release can be found at [WWW] http://sourceforge.net/project/showfiles.php?group_id=53073&package_id=114981 This release included an upgrade to the internal Axis library as requested by the LambdaStation project, support for using Safari as a web client, and a handful of internal database improvements. This release was sent to the VDT team for inclusion in VDT 1.3.9.

The earlier VOMS service publication scripts were also modified to make them easier to use. All of this was documented in the OSG ITB twiki at:
http://osg.ivdgl.org/twiki/bin/view/Integration/DiscoveryServicePublication
http://osg.ivdgl.org/twiki/bin/view/Integration/DiscoveryServiceInstallation

The LambdaStation developers discovered a subtle bug related to file handling in JClarens. A fix was made in CVS and a workaround was provided to the LS team until the next JClarens release.

The gLite security libraries in JClarens were updated to support kerberized x509 certificates. This was done at the request of the LambdaStation team. I assisted the gLite team in making and testing the changes.

JClarens packaging has been refactored so that the web application files are contained in a separate package from the preconfigured Tomcat server. This reduces the size of JClarens updates by 60%.

I spent some time optimizing the database code in JClarens. With the help of the Netbeans java profiler I was able to identify several parts of the code that were leaking database resources, not well optimized, and contained potential security leaks. After fixing most of these problems I wrote up a "JDBC best practices" document for JClarens developers to help prevent these bad coding practices from reoccurring:
http://ultralight.caltech.edu/gaewiki/JclarensDatabaseBestPractices

We continued to work with Ahsan Ikram from NUST on the new Webstart UI to Clarens. We identified some core architectural changes that would be necessary to implement as the project moves forward. We also enlisted the help of a Caltech undergraduate, Jordan Carlson, to help out with the build system and the migration of some of the existing javascript UIs to the new Webstart interface.

I continued to work with the two visiting NUST students, Atif and Adeel, as they developed the Steering and Estimation services for the GAE. Their hard work helped with the successful GAE demonstration presented at CHEP06 in February.

Software Discovery

The Software Discovery Service got a big improvement with the addition of a local database to cache the discovered information. This allows the service to remember the software publications even if the server restarts.

I assisted Brian Bockelman at UNL with the setup of their Software Discovery service.

I worked with Dave Evans and Bockjoo Kim to get the Software Discovery service publication integrated with the CMS software installation process at the various CMS Tier2 sites. Currently only two sites (Caltech and UNL) are publishing to the service, but I continue to work with Bockjoo to activate more sites.

The ITB Discovery Service was moved from ultralight.caltech.edu to a dedicated machine at osgitb-discovery.caltech.edu. In order to jump-start the use of the Discovery Service, I configured a Discovery Service publisher to push the contents of the vo-0.4.0 package describing the VOMS servers in the OSG. This has been running for over a month with no noticable problems.

The Software Discovery service continued to be used by the CMS Tier2 sites as a catalog of the CMS software available on the Tier2 clusters. There has been one feature request to allow the listing of both the gatekeeper DNS name as well as the shorter site name.

Webstart UI

Due to the numerous problems that we've had with maintaining a Javascript interface to Clarens, we finally started replacing much of the functionality with a new Java-based WebStart UI in collaboration with Ahsan Ikram from NUST, Pakistan. This new UI framework helped drive a very successful demonstration in Dec. to the visiting Rector from NUST. This UI is quickly becoming an important GAE component that should get more visibility in the future.

Tier2 Management

As a small task I helped Suresh Singh, our CMS T2 admin, install the VDT on some of our 64-bit nodes. There were some annoying issues with using 64-bit perl with the VDT that was resolved by 'downgrading' to 32-bit perl.

My role as a Tier2 admin was expanded this last quarter as the Caltech Tier2 cluster has grown. I was involved with setting up two single-node ITB clusters for OSG ITB testing, merging the two production clusters into a single large cluster, tracking down various high-load problems on the gatekeeper, and maintaining the shared GUMS server.

Conferences

An Ultralight workshop was held at Caltech, Oct. 24-28, where we got status reports from the various Ultralight sites as well as the various Ultralight working groups. I gave an update on jClarens and how it is used in the Ultralight Applications group.

Supercomputing 2005 in Seattle, Nov. 14-18. I assisted with the planning and preparations of the involvement of a new CACR computing cluster known as the SHC. The SHC is a 80 node dual opteron cluster paid for by several Caltech research groups. CACR was kind enough to let us use this cluster as part of the SC05 Bandwidth Challenge (BWC). I was responsible for the kernel and software installation and configuration of all 80 nodes as well as running and monitoring the BWC contribution from the SHC. The BWC contribution from the SHC involved running bbcp between two sets of nodes, with half acting as senders and half acting as receivers. The data transferred by bbcp was stored in memory in ramfs on both ends. Results from the SHC's contribution to the BWC showing 7.5 Gbs of aggregate throughput each direction can be found here: http://ultralight.caltech.edu/~wart/fast-even-annotated.png http://ultralight.caltech.edu/~wart/fast-odd-annotated.png

In addition to the 80 SHC nodes, I was also responsible for the management of 8 machines at our LA POP that were also participating in the BWC. I also helped with the network configuration on many of the showfloor machines.

During SC05 we were generating new Linux kernels with changes to the FAST module as well as some other monitoring packages. These had to be installed on all of the computers involved in the BWC. To assist with this, I created a yum rpm repository on ultralight.caltech.edu. This greatly simplified the process to install new packages and new kernels on the various machines.

During all of this prepration I made a number of changes to the 'crashall' tool [http://ultralight.caltech.edu/~wart/crashall.html] to assist with the configuration and management of the servers and the many data transfer streams at the SHC.

OSG ITB Workshop, Nov. 29-Dec. 1. I attended the OSG Integration workshop at Fermilab to represent the Discovery Service. I gave a short presentation on how it can help solve some of the recent VOMS url issues that have come up recently. It seemed attractive to most people, yet I was only able to convince two VOMS admins to actually use it. During the workshop I made some significant improvements to the

Discovery Service documentation on the OSG twiki. I also worked with the VDT folks to iron out some packaging wrinkles of jClarens and the Discovery Service. By the end of the meeting we had one VOMS service and one SRM/DRM service validated with the latest Discovery Service code.

NUST Rector visit, Dec. 12-14. The NUST Rector and Prof. Zaidi visited Caltech on Dec. 12-14 to discuss the Caltech-NUST relationship. I worked with the 4 visiting students to help them develop their presentation as well as put together an enticing demonstration of the work that they've done so far. The demonstration showed how the GAE components developed by the NUST students can be used to run an analysis session. The agenda for this meeting can be found at [WWW] http://ultralight.caltech.edu/agenda/fullAgenda.php?ida=a0523

CHEP06 in Mumbai, India, Feb. 13-18. While I did not present any papers at CHEP this year, I was involved in setting up high-profile demonstrations of the GAE Webstart UI, MonALISA, and EVO. These were shown to the President of India at the end of the week.

Tier2 Networking Workshop at Caltech, March 3. Helped organize and run the Tier2 networking workshop held at Caltech.

DOSAR Workshop at UT Arlington. I gave a remote presentation on the new CMS distributed production system and CRAB at this DOSAR workshop.

Education and Outreach

I continued to work with our visiting students from NUST, Pakistan. The Execution and Shell services were cleaned up and published with the jClarens 0.6.1 release. Ahsan's work on the WebStart UI proved to be an important improvement over the existing Clarens Javascript UI.

In November two more students arrived: Atif Mehmood who works on various estimation services, and Adeel Zafar, who works on a job Steering service. Both students worked hard to integrate their work with Ahsan's WebStart UI for a successful demonstration to their NUST university's Rector in December.

I gave the students a CVS tutorial to help them integrate their work with the core jClarens code repository.

Suresh Singh:

This quarter I accomplished many important improvement to our Tier2 cluster.

1. As decided in our Caltech Tier2 biweekly meeting, I became very much involved in merging our two production clusters into one big cluster, thus minimizing lots of system management and data operation overhead. I found a solution and fixed the Rocks 3.3 distribution to add all the Xeon nodes of old cluster (CIT_CMS_PG) into new Opteron cluster (CIT_CMS_T2). Altogether there are 66 nodes in the cluster. The nodes are composed like following:

- cluster frontend --> 1
- dCache PNFS/Admin --> 4
- cluster NAS Appliance --> 1
- workers (also dCache pool) --> 60

We used a 10 Gigabit link to connect 2 48-port Foundry switches to make 1 single braodcast domain for the cluster private network.

2. All the nodes excluding the NAS appliance have been connected to Ultralight network for external Internet connection.

3. I used one dual single core 2.2 GHz Opteron node for PNFS server for the dCache. We increased its memory to 4 GB.

4. I used another similar Opteron node for other dCache services. The memory in this node was also increased to 4GB.

5. All other remaining nodes are configured as dCache pool nodes as well as GridFTP door nodes.

6. After having done that we are able achieve very excellent throughput (123 MB/s) through PHeDEx. For this, I tried various combination of srmcp options like number of batch-files, number of jobs per batch file. The number that I got above was with 2 batch files and 8 jobs in them.

7. I configured our dCache for CERN based "Heartbeat Transfer" monitor.

5.1.4 D0

Gabriele Garzoglio:

Working on the development, integration, and deployment of the Resource Selection Service for OSG. At this time, we are ready for deployment on the ITB. Preparing the service to be integrated within the DZero job management infrastructure (SAM-Grid).

http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/

Working on the Resource Selection Service Project (ReSS) to allow the use of OSG resources from the SAM-GRid for DZero (http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/).

Collaborating with EGEE on the development of an OSG plugin to CEMon, in order to publish resource characteristics in old classad format using the GLUE schema (http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/CEMonCollaboration)

Prepared a test bed to demonstrate resource selection on OSG using the condor-g framework (http://osg.ivdgl.org/twiki/bin/view/ResourceSelection/TestBedPrototype)

5.1.5 STAR

consolidating our infrastructure, deploying and supporting SRM to more sites and deployment, development and consolidation of highly scalable Storage Element access solution for local site using our BNL resources as testbed. While our participation to the OSG council meetings have continued (J. Lauret) as well as on the OSG executive board (D. Olson), our participation in OSG was enhanced by W. Betts , E. Hjort and I. Sakrejda presence in grid operation center, OSG integration, OSG storage

5.1.5.1 Operation and support

General

Support for remote site continued with a focus on making all sites operational within an OpenScience Grid software stack deployment context. Efforts from Elisabeth Atems (Wayne State) and Alexander P. Suaide (São Paulo) have leaded to exploitable sites although several issues still remain to be solved. Especially, multiple bugs were found in conjunctions with the use of the SGE job-manager which seem to be lightly supported or used. Problems included

- Truncated logs - problem has not been resolved yet
- Improper reporting of job state – this problem has not been understood nor yet fixed
- Repeated log information during polling: cleanup of temporary files was not done resulting in multiple append mode IO. This was fixed in STAR by J.Lauret
- Improper treatment of output channels: the sge jobmanager treats all IOto any channel (files or special device) equally and use temporary files via a mechanism described in globus ticket 1287. This results in an attempt to write to /dev/null.real for special device access. Problem resolved by J.Lauret.
- Improper definition and assignment of SGE_ARCH problem communicated to Iosif Legrand by Elisabeth Atems.

We hope most problems are now fixed and it is still to e seen if patches will make into the next OSG software stack release.

BNL Infrastructure

Wayne Betts has continued his support for the infrastructure at BNL. Of particular interest, he traced and demonstrated a problem with the GridFTP server: in the OSG 0.4.0 release: globus-url-copy could not be used to access a directory not owned by the mapped account primary group. It appears the community have "resolved" this in the interim by opening the protection to world writeable, not an option for our BNL security perimeter. W. Allcock was very prompt to fix this unix group membership problem ; we tested his patch and release and confirmed the problem as fixed. The patch was further deployed on all STAR sites.

At our BNL site, we also had to re-discuss with the facility personnel the Grid approach. While generic accounts were created and allowed for the Atlas VO, it appeared that this use was limited and un-practical (un-usable) while accessing a STAR gatekeeper. Since the two set of resources are connected, their physical network layout is such that using pool accounts from Atlas on STAR resources would not be practical. Especially, home directories would cross boundaries of a few firewalls and network. Furthermore, since there is a possibility for jobs submitted to one set of resources to be dispatched on the other set of resources (i.e. jobs submitted to STAR gatekeeper could end up running on Atlas node and vice versa depending on opportunistic resource availability), having two separate groups of home directory in LDAP although considered as initial solution would not be workable. We reshaped this and agreed to have the pool accounts split into two pseudo-arbitrary subset: even number would use Atlas home directories and odd numbers STAR storage resident home directories. In practice, users in a given VO allowed to access STAR resources and using those preferentially would be mapped to an account using STAR resources. Ideally, we would like to migrate to a truly common shared storage but the current lack of a support model for present and future resources does not allow engaging in such direction.

Condor level resource allocation was also reshaped to allow non-STAR VO to access STAR resources with a chance to run without being constantly evicted. We opened our resources to both LIGO and GADU as a VO to VO agreement. This mechanism needs to be tested and monitored as per the job throttle other VO would be able to gain.

We would like to acknowledge the help from and thanks John R. Hover, Morris Strongson and Alexander Withers from the RHIC Computing Facility for their assistance in those matters.

PSDF Infrastructure

Eric Hjort and Iwona Sakrejda worked deploying and validating a Storage Element at PDSF and registering it in Gridcat. Eric further provided documentation which was later used for a similar deployment at WSU. The documentation has been judged adequate and easy to follow. The OSG 0.4.0 was deployed which matched the version at BNL and SPU. Only WSU still needs an upgrade at this stage.

Accounting and Monitoring

Much distraction arise from account and monitoring in general, a topic for which STAR is in principle not primarily involved as per its Grid activities but usually relies on products from others. First, it appeared we had inconsistent records through MonaLisa: no short jobs are reported. There also seem to be periods where we have no jobs reported at all while we are carrying at the same time scalability tests and simulation production. At PDSF, the `run-mis-ci.sh` could take up to four hours to execute (this was reported by Iwona Sakrejda to the operation center but no fix not answer has been provided to date) jeopardizing monitoring as well as stable gatekeeper. We consider at this stage monitoring as an area which would deserve bottom to top re-structuring.

This leaded to the development of our own monitoring tools, as seen from the submitter site. A few ways have been developed:

- The first is as seen submitted by SUMS (STAR Unified Meta-Scheduler). A jsp was provided by Levente Hajdu: it displayed the job dispatched per cluster but do not consider the success/failure of the jobs dispatched not does it considers jobs possibly killed just after submission.
- The second sets of monitoring graphs were provided by Lidia Didenko who developed two tools. The Grid Job site status is available as a CGI: it simply list all jobs and their detailed status as it stands including actual job state, input and output transfer status, WAN transfer status, SRM status if used. All information is kept within an historical timeline. The Grid job efficiency monitoring is

also available as a cgi and has two forms: for a given site, it reports detailed efficiency for the diverse stages and issues we have observed (transfer of logs, globus failures, input/output transfers and our own reconstruction efficiency). Efficiency is simply here a ration of expected versus delivered, the global efficiency do not fold in the standard error channel log return as if the output from the job are complete, we would consider the job done for a more "relaxed" criteria . Our current observation is that the overall efficiency is dominated by either the log file return, showing an un-expected weakness in Condor-G or the output transfer. We will discuss this last controllable issue in the next section

- Doug Olson regularly provides accounting information based on parsing the content of the `$GLOBUS_LOCATION/var/accounting.log` file. The metrics and accounting information are publicly available.

Within this first attempt to bring consistency in our accounting, we have not spent much time yet to understand all the details and especially, consistency between the diverse way to look at the problem would need to be sorted out within the next few months: at the end, all number should match and success/failure, cancelled jobs, accounted jobs, CPU time and Wall times seen from the submission site or at the execution site must be consistent. It is unclear however how "far" we should go into this activity which is not our primary area. For now, we consider this effort as pragmatic and STAR interest focused in the absence (but awaited) of a general and stable tool from the OSG providing similar functionalities.

5.1.5.2 Storage Resource Management

This activity is mainly supported by Eric Hjort (LBNL/STAR) and Alex Sim (LBNL/SDM) but extended across STAR's resource providers with noticeable and welcomed efforts from Wayne State University.

Production level data transfer

We continued our production level data movement and had the equivalent of 10.2 TB moved in the first quarter (60TB and 500k files for FY05) all done using SRM+RRS. We will assume in future this to be a closed chapter and not report further this activity. We would like to thank once again our collaborators from the SDM center for their support within the scope of the PPDG project. This activity has truly been a success story with net benefits to our experiment.

SRM use in testbed

In the last report, we mentioned the use of SRM+RRS in test job as a replacement for a gsiftp approach. We pursued this activity and did proceed with a scalability test of 200 production grid jobs submitted daily. The jobs use SUMS along with SRM file transfers for both input (~170MB) and output files (~700MB). SRM performance has been generally reliable while used between PDSF and BNL although we have encountered a known globus problem (`pthread_mutex_destroy` error) which relates resource exhaustion on a gatekeeper. This condition is reached due to other activities putting a large load on the gatekeeper.

In the second quarter, we extended the usage of the SRM+RRS scheme to a new site: Wayne State University. Submission and tracking problems on multiple sites at a time revealed, in the long term, several site-specific issues which have made the exercise difficult, perhaps due to the lack (again) of clear messaging and tracing capabilities of the grid infrastructure itself when various outages occurs, sometimes after upgrades as new problems are introduced (see previous section for relation with job scheduling on remote site). Overall, we have nonetheless overcome the issues and extended our STAR SRM jobs to a new site. The first SRM specific problem we encountered at WSU was that SRM clients would crash. It was determined that this only occurred on remotely submitted jobs and by working with the SRM collaborators, a bug was found and fixed: memory was being allocated for a user's proxy but a delegated proxy is somewhat larger and the memory allocation was insufficient. The most likely explanation for this issue is that we did not previously observe it due to the use of direct proxy rather than delegated ones (it remains unclear as some tests used delegation but did not exhibit the crash). This is now fixed in the latest SRM release. Our testing and use was severely hindered by the `sge-jobmanager` issue with multiple append of output already mentioned. While fixed, we had not observed this behavior either in previous OSG releases.

5.1.5.3 SUMS and grid data production

SUMS developments

This activity is mainly supported by Levente Hajdu as part of his main responsibility as a funded member of PPDG.

The first major development addition was the addition of JDL level to concrete implementation of a SandBox concept allowing users to bundle into an arbitrary archive-plugin or repository of files and directory tries. Such SandBox can then be retrieved on remote sites and unpacked for execution in an environment similar to the user's initial code layout structure. The implementation allows many different flavors of sandboxes, the differences between them are the methods for copying and packaging the file. We currently have three flavors of SandBox:

- SandBox-ing using soft-links: this first implementation was really a proof of principle for local site and allows for consistency between U-JDL declaration of SandBox in regards of dispatcher: local and Grid dispatching using a U-JDL declaring SandBox would work with equal transparency which is a key feature of SUMS
- Condor-G "dumb" copy with no packaging – this is used (and useful) only when little input/output is needed and only practical to the extent that all files could be enumerated
- ZIP based SandBox: implement a full archiving mechanism with persistent name tagging (assembled and unpacked once for a give specified tag). . In grid context, the full SandBox is transferred via Condor-G one input and handling within the job csh wrapper, the resulting tree is deployed in OSG $APPS. This seemed to work reasonably well providing remote availability of ZIP archive (case on Linux). In local scheduler context, this mechanism works equally well but deploys the code in the local tree (which may be a SCARTCH space defined per job). The ZIP SandBox in both cases is limited by its implementation, one per job.
- PACMAN SandBox: this is pending as we encountered a few PACMAN issues on how to handle this mode of operation within a user based SandBox approach. The intent is to supply a DAG DAG mechanism where a single PACMAN code deployment will be initiated within a workflow, palliating for the previous ZIP mode limitations.

While the first mode (Soft-Link and local) is still the main mode of operation, SandBox is allowing us to migrate user to the next step of virtualization where their jobs are completely and abstractly defined. This brings us close to a user-on-Grid concept (one dispatching policy away) as far as users do not use in their main command block hardwired path. We will need in future to enhance tutorials to change this natural tendency in user's habits.

Other noticeable changes are the addition of a file transfer mode based on SRM rather than the previous `globus-url-copy` approach. This was an important step to support grid jobs production as the previous schemes, relying on `globus-url-copy` or native condor-G, revealed themselves as non-scalable and locking computing and storage resources together. Finally, we added a tracking and statistic plug-and-play module for reporting usage of the Catalog module. In STAR, this allows tracking usage of the Meta-Data most used by users, helping in spreading or replicating on the fly most used datasets as one of the many practical usage examples.

A multi site policy was added to the scheduler in the most recent release. The policy distributes user jobs between multiple sites i.e. BNL, WSU, PDSF and São Paulo in a pre-fixed ratio appropriate for each sites throughput. A function to recover the state of the site (up or down) was also added allowing to automatically disable submission to sites or queues where jobs may not run and reapportions load to other viable candidates. The test will aid Grid job submission as detection of up/down sites is now automated.

Deployment and exploitation of SUMS

SUMS was deployed at WSU which completed its availability at all sites currently part of the resource provider for the STAR VO. Elisabeth Atems was the main contact, tester and now the current maintainer of the site policy.

We also had a very interesting experience of implementing a new dispatcher for Mac OS-X for our MIT colleagues: equipped of multiple weakly connected Apple PowerBook G4 nodes and expanded to an Apple XGrid cluster at MIT, SUMS turned out to be the only tool allowing them to fully exploit all nodes providing a transparent front end to XGid. Test and deployment were successful and once again, SUMS have demonstrated its versatile and adaptive nature (as we did not initially intend to support Mac OS-X XGrid environment). We consider this as a major achievement in fact.

We started the integration of information coming from GridCat into SUMS for Grid job submission. GridCat is the first (and only to date) implementation of a generic module which purpose is to recover simple information such as storage path (APP, DATA, TMP etc …) and resource in general. We believe that a combination of actual GridCat information (or similar) further supplemented by published (to an information service) knowledge about queue characteristics as virtualized in SUMS is a key to the dynamic, self-adaptive and efficient exploitation of resources spanning many disjoint Grids. We believe that the exploitation of resources composed of combination of like XGrid, OSG or other resources will require higher level knowledge and availability of such information.

All recent changes in SUMS, coupled with internal monitoring of our efficiency in exploiting grid resources are part of a focused effort to make the best of the OSG infrastructure. Our success varies as a function of weeks, software stack releases and the number of jobs we have been able to stably sustain is not terribly impressive to date: for simulation, we can ignore failures (which can creep up to 30% at times) and resubmit as much as necessary but such numbers is incompatible with the notion of allowing users to exploit Grid resources. We are far, as a community, from the final ideal and goal. In general, it is disconcerting to imagine we can utilize with greater stability an infrastructure based on and XGrid cluster rather than an infrastructure which is common to our every day work environment. This statement is true without considering scalability and large test testing (100 jobs a day).

5.1.5.4 Other activities

Valeri Fine has continued support for PACMAN including work with Levente for integration with SUMS. This activity is pending at this stage.

Pavel Jakl joined us in March in support of the Xrootd activities. We hope to be able to deliver an SRM based Xrootd implementation but engaged now in a deployment and consolidation phase of the product. Pavel has found many shortcoming and miss-features now corrected. We have deployed a fully functional production mode setup and are equipped of a parallel deployment version for further development. Current issues revolve around round robin support, basically not working in a large cluster and MSS plugin. The later is problematic if direct connection to HPSS for example is used i.e. the sum of all un-coordinated requests breaks IO performance and the system at the end is brought down to a stall. We naturally used the DataCarousel, a tool developed in house to coordinate requests from MSS as a layer between Xrootd and its file mechanism retrieval. Our current throughput remains poor however which needs further investigation. In all issues we have found, assistance and help from the Xrootd team has been outstanding.

Due to the use of SUMS in STAR, the transition from rootd to Xrootd required no major change from our users and in fact, one letter change in their U-JDL. This is not a trivial statement as rootd requires direct access to a physical name space while xrootd demands a connection to a redirector with logical file name namespace.

Finally, we were intrigued by Sun's Sun Grid Utility Computing and attracted the OSG Council's attention to this endeavor (perhaps an avenue to better understand a different community needs with different problems). We invited Dr. Ari Shamash, Director of Sun Grid Engineering team from Sun Microsystems for a BNL technology meeting and had a rather interesting conversation as per the project's objectives, current simple access model (perhaps too simple) but an extremely strong network and cross communication requirement and therefore, data security requirement. We exchange ideas on possible cross-bridging avenues and how we could engage in mutually beneficial partnership.

5.1.5.5 Documentation, policies and presentations

E. Hjort presented an overview of the Storage Element use in STAR at OSG consortium meeting in January. At the time, STAR was one of the two VO having a SE registered in GridCat.

Several presentations of our work and results were made at the CHEP 06 conference. We list them as follow:

- Fine, Valeri: A generic approach to job tracking for distributed computing: the STAR approach
- Hjort, Eric: Data and Computational Grid decoupling in STAR – An Analysis Scenario using SRM Technology (presented by J.Lauret)
- Hajdu, Levente: Meta-configuration for dynamic resource brokering: the SUMS approach
- Jakl, Pavel: From rootd to Xrootd, from physical to logical files: experience on accessing and managing distributed data

Before CHEP06, J. Lauret gave a general presentation on Building a US Based Infrastructure for Open Science. This was made at the International Workshop on Large Scale Computing in Kolkata, India which was mainly dominated by Indian institute for computer science with participation of the STAR and Alice collaborations. The audience was in general interested by Grid developments with perhaps, a preference toward EU based Grid from which they receive major support (including frequent on site visit and training we cannot afford to provide at this stage).

Finally, J. Lauret and D. Olson were the main contributors of The Open Science Grid consortium voting procedures they developed and wrote.

We would like to thank the US agencies for allowing us to be fairly represented at CHEP and IWLSC conferences, important to our virtual organization as we have a strong involvement of Indian colleagues in our collaboration. We would like to further thank the IRP and GA funding agencies of the Czech Republic as well as Prof. Michal Sumbera for their generous support of Pavel Jakl's work on distributed data management and Xrootd.

5.1.6 ALICE

5.2 Facilities

5.2.1 Jlab

After the fall SRM collaboration meeting at Jefferson Lab, I focused on evolving the SRM operational specification in a way that maintains its functionality while improving its ease of implementation. As the specification moves toward version 3, I offered the perspective of the smaller site that may be interested in implementing or using the SRM, but must do so with limited resources. This software engineering perspective values the organization of the SRM code into manageable modules that interact in well-defined ways.

I presented a potential operational specification document that follows this philosophy. In it I differentiate the SRM functionality from its operational specification. The existing functional specification addresses the semantics and high-level services of the SRM. The operational specification concerns itself with issues of software engineering and implementation that are needed to bind the functional needs to a real implementation.

I also engaged in an effort to create a requirements document defining the simplest set of requirements that an SRM need provide. I circulated this to the SRM developers' mailing list for comment. I also have analyzed various proposed advanced features from this perspective and provide this feedback on the mailing lists and in conference calls.

In the January-March 2006 time frame I participated in the bi-weekly SRM conference calls which included FNAL and LBL. My particular focus in these phone calls was with the separation of the functionality of an SRM from the operational spec, i.e. the wire protocol. From a software developer's perspective, it is necessary to have a well-defined low level definition, but while trying to define

functionality in the collaboration it is helpful to talk at a higher level. With this in mind, I provided feedback to the SRM developer's mailing list on the latest revision of the SRM v3 functional specification to help with clarity.

I also upgraded the code base of the JLab's SRM v3 prototype. This included bringing it up to date with our latest code libraries (database packages, configuration packages, etc.) and restructuring the build environment to conform to the JLab Scientific Computing build standards. This will allow us to more easily keep the SRM prototype working as Jasmine (JLab's internal Mass Storage System) is upgraded and as the v3 spec is further tweaked.

5.2.2 BNL RCF/ACF

Gabriele Carcassi: I continued my work on the Privilege Project infrastructure, which included:

* leading the effort in providing an implementation for a GT4 AuthZ module that allows integration of the privilege system with GT4. A preliminary version is being added to VDT and is available for testing.

* participation at the joint security workshop (OSG-EGEE) at CERN.

* slides and demos for SC2005

Among other things, I have contributed to the area of discovery services by reviewing the VOMS registration scripts and setting them up on the ATLAS and STAR VOMS.

5.2.3 FNAL

Fermilab increased the number of clusters on the common grid infrastructure with the addition of the new CDF farm to FermiGrid. CDF jobs are now running on several sites on the OSG and FermiGrid using the "Glidein" technology from the Condor Project for just in time scheduling for job execution.

5.2.4 NERSC/PDSF

PDSF upgraded the gatekeeper (pdsfgrid2.nersc.gov) for the OSG production grid to the OSG 0.4.0 release. A gatekeeper (pdsfgrid1) was installed on the ITB and used for testing the drm.

5.2.5 SLAC

5.2.6 Collaboration with IEPM, Network Performance Monitoring

This section covers Jan. – Mar. 2006. For the Oct. – Dec. 2005 period see http://www-iepm.slac.stanford.edu/about/status/ppdg-2005-12.html.

5.2.6.1 Bandwidth/Throughput Monitoring

As part of the Terapaths project we collaborated with ESnet to make measurements of jitter with OSCARS. The results showed that OSCARS worked well both to limit throughput and to reduce jitter.

Many improvments were made to the IEPM-BW output to add more scatter plots, to provide a zoom capability, extend the use of pathload and work with the author to add timeouts, to improve the detection of events using the Holt-Winters method.

At the requests of the SLAC accelerator operations folks we made a case study of an Oracle application's performance between SLAC and ORNL. The problem was resolved to be in the application.

5.2.6.2 Passive Monitoring

We put together a paper on the effectiveness of passive monitoring for forecasting for CHEP06. The results are promising but do not remove the need for active monitoring.

5.2.6.3 PingER and Developing Region Monitoring

Based on a paper by Bamba Gueye and others on locating Internet hosts, we are developing a web based Java host location tool called TULIP. This uses ping Round Trip Times (RTTs) from landmark sites to the host whose location is desired to provide triangulation information to determine its location. A problem with Gueye's method was its lack of landmarks outside Europe and N. America, so we developed a reverse ping server that is being deployed at IEPM PingER and BW sites to provide landmark facilities. Besides N. America and Europe we now have landmarks in Brazil, S. Africa, India, Pakistan, Taiwan, and Japan and are actively working to add others. In particular we are working with the NLANR/AMP measurement team to install the landmark code at critical AMP sites in New Zealand, Australia, China, Malaysia, Phillippines, Chile and N. Africa. Besides providing a general purpose tool to help locate hosts for Internet users, it will also be valuable to validate the locations of over 700 PingER sites around the world. Since we supposedly know the locations ofthe PingER sites. It will enable us to evaluate the accuracy of the method.

We made progress on automating PingER management by providing scripts to identify and tabulate non-responding hosts.

We finished up the yearly ICFA/SCIC review on network performance worldwide. We then used this as a basis for a detailed study of performance for S. Asia and Africa that was presented at CHEP06. To simplify producing the performance report figures we are developing a Java application to provide long term performance trend plots.

5.2.6.4 Testbeds

We completed a Memorandum of Understanding with Microsoft to evaluate and report on the performance of the new TCP stack that will be released with the next version of Microsoft Windows (Vista/Longhorn). Following this we installed Windows Vista and Longhorn and completed the first phase of the evaluation using production high-speed networks within the US and trans-Atlantic.

5.2.6.5 Admin, visits, papers, presentations, proposals etc.

The IEPM group hosted a successful two week vist by two staff members from the Pakistan National University of Sciences and Technology (NUST) Institute of Information Technology (NIIT). The visit was to discuss further collaboration and projects. As a result we put together projects to:

- Integrate Smokeping and PingER
- Evaluate pathneck
- Improve PingER visualization
- Integrate PingER and AMP
- Provide automated diagnosis of events

Les Cottrell visited NUST/NIIT in Rawalpindi/Pakistan for a week to work with students and faculty and to present findings on Pakistani Academic and Research network performance to various senior ministers and leaders in Pakistan. We also got agreement from NTC and COMSATS to install PingER monitoring hosts at their PoPs in Pakistan.

We submitted the following proposals:

- SciDAC: *SOCRATES: Series of Circuits Rapidly Allocated for Efficient Sharing* together with GA Tech, U.and Ga, U. Va, Colorado State
- SciDAC: *CANTIS: Center for Application-Network Total-Integration for SciDAC* with ORNL, BNL, FNAL, GA Tech, PNNL, UCD.
- SciDAC: *Large Hadron Collider - Monitoring Infrastructure* with Caltech and U. Mich.
- NSF MPS/Physics: *PLaNetS: Physics Lambda Network System* with Caltech, U. Fl, Haystack/MIT, UCAID, FNAL, BNL, U. Mich.

Publications:

- *Evaluation of Techniques to Detect Significant Network Performance Problems using End-to-end Active Network Measurements* R. L. Cottrell, M. Chhaparia, F. Haro, F. Nazir, M. Sandford, NOMS 2006, April 2006
- *Quantifying the Digital Divide: A Scientific Overview of the Connectivity of South Asian and African Countries* A. Rehmatullah, R. L. Cottrell, J. Williams, A. Ali, CHEP06.
- *Terapaths: A QOS-Enabled Collaborative Data Sharing Infrastructure for Peta-Scale Computing Research*, S. Bradley, F. Burstein, B. Gibbard, D. Katramatos, R. Popescu, D. Stampf, D. Yu, R. L. Cottrell, Y. Li, S. McKee, CHEP06.
- *A Better Internet* ComputerWorld News Story, February 13, 2006.
- *January 2006 Report of the ICFA-SCIC Monitoring Working Group* edited by Les Cottrell for the ICFA SCIC Monitoring Working Group.

We made the following presentations:

- *IEPM-BW Deployment Experiences* Presented by Connie Logg to the Joint Techs Workshop February 2006, Albuquerque, New Mexico.
- *IEPM-BW: Bandwidth Change Detection and Traceroute Analysis and Visualization* Presented by Connie Logg to the Joint Techs Workshop February 2006, Albuquerque, New Mexico.
- *Stanford University, SLAC, NIIT, the Digital Divide and Bandwidth Challenge* Presented by Les Cottrell to the NUST/NIIT Faculty,Islamabad, Feb 2006.
- *Using Netflow for Forecasting* Presented by Les Cottrell at CHEP06.
- Quantifying the Digital Divide: A scientific overview of the connectivity of South Asian and African Countries presented by Les Cottrell at CHEP06.
- *SC|05 Presentation to ESnet* (pdf) presented by Yee-Ting Li at the ESCC Meeting, University of New Mexico 02/09/06.

5.3 Computer Science & Middleware

5.3.1 Condor

The Condor Project produced four 6.7 Developer's series releases, and one 6.6 Stable series release for the first quarter of 2006:

Developer's series 6.7.15:

This is a development series release with a number of new features and bug fixes. This version includes a more efficient negotiation protocol to speed up matchmaking cycles in the negotiator. Condor can also automatically start a gridftp server to handle file transfers for gt4 jobs. In addition to platforms that were previously supported, Condor now provides a clipped port to PowerPC under YellowDog Linux, and it is known to be fully functional on Intel x86 under Red Hat Enterprise Linux 4 and Fedora Core 4. Under Windows, the Condor installer is now MSI compliant.

Developer's series 6.7.16:

This release provides the same functionality as 6.7.15, with several bug fixes that we found early enough in the release cycle to warrant a quick release of 6.7.16. There is one small additional feature: the Windows version now supports "personal Condor" mode. Normally, you would still want the usual full-featured

setup in which Condor runs as a Windows service using the System account, but this new capability gives you the option to run Condor directly in any account, using only the privileges of that account.

Developer's series 6.7.17:

This is a development series release with a number of new features and bug fixes. Condor now include direct support for running backfill computations on unused nodes. Parallel job scheduling can now be grouped to avoid spanning a parallel job across different switches. Parallel jobs can now be managed with DAGMan. There are a variety of Condor-G scalability improvements. Condor provides more flexibility in setting port ranges for Condor to work through, easing configuration when working with a firewall.

Developer's series 6.7.18:

This release features important security fixes and many other improvements, including Kerberos support for Windows enabling cross-platform authentication, the ability to run jobs as the submitting user on Windows, and a port to HPUX11. We strongly recommend sites running earlier versions of Condor in the 6.7 development series upgrade to 6.7.18 as soon as feasible.

Stable series 6.6.11:

Condor 6.6.11 has been released. This release contains important security fixes. We expect 6.6.11 to be the final release of the 6.6.x series. We strongly recommend sites running earlier versions of the 6.6.x series to upgrade to 6.6.11.

In this quarter, the VDT team produced Developer's series release 1.3.10, which includes Condor 6.7.18 and critical security updates. The new Condor managed fork job manager feature is now available to manage the load of fork resource jobs upon the gatekeeper host. The Virtual Data Toolkit (VDT) team continues to work aggressively with the OSG to develop the OSG 0.6.0 release, by adding and upgrading packages and patches needed by the OSG. The VDT is a major component of each OSG release.

The Condor team continued to operate the OSG-ITB Grid Exerciser (GridEx) on behalf of the OSG-ITB. In this quarter, the GridEx team was able to post the first error-free OSG-ITB daily report. This was achieved by working with sites to update their local configuration, and by incorporating OSG-ITB lessons learned into the Condor Grid Universe (Condor-G) client. Additionally, the GridEx served as a validation tool for the OSG-0.3.6 release.

Condor's accounting support has seen increased development, with enhancements to allow sites to increase the resolution of user-based resource utilization reports. User accounting capability has been added to the CondorView monitor tool. Development has begun for a new job-based accounting system.

5.3.2 Globus – ANL

Effort from both the PPDG and CMS DISUN projects is being used to test scalability of the Globus GT4 web service GRAM for CMS data analysis jobs.We are currently iterating between GLOBUS and US CMS to test up to 3500 jobs and fix any problems found in Globus.

5.3.3 SRM – LBNL

Participants: Alex Sim, Junmin Gu, Viji Natarajan, Alex Romosan, Arie Shoshani

The following activities took place during the last quarter:

OSG activities

- Reporting active information about SRM status and resources into JClatrens
  The Integration Test Bed (ITB) group of OSG has requested participants to publish periodically into JClarens the operational status, and the resources available. We have responded by adding this capability on to SRM-DRM. This is now part of the version in VDT. We report the following storage information: total space that the SRM-DRM manages, and current available space. The available space includes all space occupied by released files and/or files whose lifetime expired.
- Extending SRM-TESTER to provide status information to GridCAT
  The SRM-TESTER is a new service component that was developed during the previous period. The package is a tester program to test any SRMs. This test program can test the following

functions: Get, Put, Copy, Ping, GetRequestStatus, and getProtocols. During this period a capability was added that lets GridCAT request the status of an SRM by calling the SRM-TESTER.

- Extending SRM-TESTER to provide status information to Web site
  The same capability as above was extended to allow publishing to a web site. The web site has to include a cgi-scripts we have developed for this to work.

- SRM-DRM was ported to Fedora Core 4 at the request of LIGO
  This was performed by using one of their gate-node machine. This ported version will be distributed with the next VDT release.

- Alex Sim participated at the OSG-ITB meeting at Fermilab (11/30-12/1)
  During this meeting Alex has learned how use the NMI facility at UWisc to automate the complilation of future version of SRM-DRM for different operating systems. Alex already set up the NMI framework for taking advantage of this capability in OSG.
- Provided support to PDSF staff for ITB nodes
  We helped Iwona from PDSF to install SRM-DRM from VDT through pacman. We also helped in having the SRM-DRM being displayed on GridCAT.
- Plans to use the SRM-TESTER by The Integration Test Bed (ITB) group
  The Integration Test Bed (ITB) group of OSG needs to have a comprehensive way of validating Storage Elements (SEs) that use SRMs. We have proposed to them to use our SRM-TESTER which is already in VDT. Since the SRM-TESTER checks not only whether the SRM is alive, but also checks getting and putting files from/to the SRM, the information on the storage location to get/put files needs to be made publicly available. We have identified the additional elements that need to be added to the GLUE schema of an SE for this purpose. In addition, our plan is to have the SRM-TESTER check the interoperability between another SRM and the SRM being tested, by issuing an srmCopy call. SRM-TESTER for SRM v1.1 will be used.
- SRM-TESTER for SRM v2.1.2 for PPDG Common Project
  The SRM-TESTER for SRM v2.1.2 was developed during this quarter. The package is a tester program to test any SRMs based on SRM v2.1.2. This test program can test the following functions: srmPreapreToGet, srmPrepareToPut, srmCopy, srmLs, srmReserveSpace, and all release functions. We have tested SRM-TESTER against Berkeley SRM and SRM/dCache at FNAL. CERN EGEE people were interested in the source codes, and we provided them.
- SRM-TESTER usage in GGF-GIN interop demo

  The GGF-GIN (Grid Interoperation Now) group plans to use SRM-TESTER for SE interop testing. The information for the SRM-TESTER mentioned in the previous section will be collected from SRM administrators for this testing rather than published in the GLUE schema (since there is no common discovery service across multiple grid domains at this time). We plan to use SRM-TESTER for SRM v1.1 as well as SRM v2.1. The current institution and representative participants are:

  EGEE (Erwin.Laure@cern.ch;),

  OSG (fkw@fnal.gov, chadwick@fnal.gov),

  APAC (glenn@physics.unimelb.edu.au),

  ARC (oxana.smirnova@hep.lu.se)
  Grid.it (riccardo.zappi@cnaf.infn.it, luca.magnoni@cnaf.infn.it)

<u>Continued development and support</u>

- SRM-DRM was ported to Scientific Linux 3 at the request of STAR/PDSF
  PDSF has decided to change all of their Grid machine to Scientific Linux 3. The grid machines use our SRM/DRM as well as SRM/HRM (which access HPSS) for the STAR replication and Analysis Scenario. Therefore, a port for Scientific Linux 3 was made, and will be distributed with the next VDT release.

- SRM/DRM is being used in Spanish ATLAS Tier-2 center for the LHC Grid Project
  The PI of the ATLAS Tier-2 center in Spain, Andreu Pacheco, contacted us in order to use a light weight SRM that manages a single file system. SRM/DRM was downloaded from our web-distribution site and installed successfully. We provided help in advising them with the appropriate configuration parameters. The PI plans to use SRM/DRM in the next LCG Service Challenge 4 for transfers between Tier-1 and one of the Tier-2.
- Support to the STAR analysis scenario
  SRM-DRM support for the STAR analysis scenario was developed during the last quarter. It consists of the following four steps: 1) files are requested through DRM on the PDSF cluster to get files from BNL's HRM; 2) the files are analyzed on worker nodes at PDSF through posix I/O; 3) the output files go to the DRM cache; and 4) the output files are pushed to BNL's HPSS through HRM.
  A new requirement was made to get files through SRM-DRM into a user-controlled space, and have the users identity associated with the files. This required the feature of calling third party gridFTP. Now, if the target URL is a gsiFTP URL, the SRM-DRM performs the third party transfer to maintain the users' identity with the files.
  The STAR analysis scenario involves several sites that include clusters at BNL, at a site in Brazil, and Wayne State University. Recently, testing with the job manager which calls SRMs for space management and file transfers was failing. This was traced to a problem in the DRM client program with exit codes. This problem was fixed.
- Adapting the SRM-HPSS on STAR to use gsi-enabled PFTP
  The STAR project was using SRM-HPSS with simple login and password PFTP. They have requested to switch to the gsi-enabled PFTP. We have helped in this process, including running multiple tests.

xrootd-srm development

In the previous quarter we have planned a joint activity to have xrootd use the Disk Resource Managers (DRMs) to manage the space of server nodes, and to use the Hierarchical Resource Managers (HRMs) to access HPSS. During this quarter we have implemented two items in support of this activity. 1) We have developed a C++ client API that can be used by the xrootd software to make calls to get files from the archive site (HPSS) and put files into the DRM's disk cache. 2) We implemented a put capability that pushes the file to the archive site. This "push" feature was essential in order to avoid having to install a gridFTP server on every worker node. Instead only the gridFTP client that is bundled with the DRM software is needed.

In this joint activity with the xrootd developer (Andy Hanushevsky) we intend to have xrootd use the Disk Resource Managers (DRMs) to manage the cache space of server nodes, and to use the Hierarchical Resource Managers (HRMs) to access HPSS. The development during this quarter was mainly aimed at making xrootd SRM-aware.

In order for xrootd to use the DRM, we extended the functionality of two xrootd components, called the OFS (Open File System) and the OSS (Open Storage System). We derived from the respective base classes, new classes which implement SRM specific functionality.

The functionality currently available includes putting/getting files into/from an HPSS by each local DRM at an xrootd server node invoking a remote HRM. Files that go into the DRM cache are pinned for a default lifetime on Open( ) file commands and released on Close( ) file commands. To implement these capabilities the OFS libraries had to be modified.

These new classes need only be specified in the config file to enable xrootd to access DRMs in all nodes. This required a few changes to the parsing of the config file.

SC2005 demonstrations

- Large-Scale File Replication using DataMover Technology
  Eric Hjort initiated a large scale data replication using the Data Mover and the SRM-HPSS at BNL and NERSC. We used the File Monitoring Tool (FMT) to display the progress of file

replication. The files were registered to the STAR file catalog at NERSC automatically using our current version of the Replica Registration Service (RRS).

- Berkeley SRM: LBNL implementation of v2.1.1
  The new SRM version from LBNL was demonstrated. This version includes new features: space reservation and directory management. We used the SRM-client recently developed to demonstrate the functionally.
- STAR Analysis Scenario using SRM-DRM in Open Science Grid
  The STAR Analysis Scenario described in the OSG section above was demonstrated by Eric Hjort. He initiated long-running tasks, and demonstrated their correct automatic completion.
- SRM-TESTER as part of Open Science Grid
  Five sites had SRM-DRM and SRM-dCache running, as well as a gsiFTP site. The SRM-TESTER was used to check conformity to the SRM protocols. In addition, the interoperability of the SRMs was checked by copying files from one SRM to another.

5.3.4 Caltech

Conrad Steenberg:

Technical Development

The prototype 64 bit Itanium release of the Clarens server was developed further and deployed at both CACR and the SDSC TeraGrid node. In particular changes were made for a more robust and flexible "shell" service that allows authenticated execution of shell commands on the server side. This allowed further development of an National Virtual Observatory (NVO) "Astronomy portal" or web interface where users are able to submit jobs directly to the scheduler (PBS) on the TeraGrid node. The prototype portal page can be viewed at http://nvo.cacr.caltech.edu:8080/cromlech/

The Virtual Data Toolkit (VDT) release of Clarens was also adapted in response to developer feedback, increasing the security of the default installation, and improving ease of installation.

An effort to integrate Clarens and the JobMon service in the the CMS software distribution was started in response to a need to use the JobMon service in installations that do not use the VDT.

Progress was made on developing a lightweight version of the Clarens server that doesn't require the Apache web server, and is therefor easier to deploy.

Collaboration

I attended the OSG Consortium meeting at the University of Florida, January 23rd to 26th, as well as the CMS week held at CERN from March 13th to 17th.

I attended the Computing in High Energy Physics 2006 (CHEP) conference in Mumbai, India. A presentation on the JobMon service was given, and a paper submitted for publication in the proceedings.

I acted as chair of the weekly OSG Monitoring and Information Systems meetings held to coordinate monitoring systems developments in OSG, and attended meetings of the Grist project mentioned above to aid in the development of the Montage web portal and the HotGrid graduated security implementation.

Technical Development

During this quarter a Clarens server release 0.7.2.1 and web interface release 0.7.2 was made.

This release contained mainly small improvements and some bug fixes, including the use of a new Discovery Service schema to publish parameters to MonALISA, as well as changes to the shell, proxy, and file services. A full changelog is available at
http://clarens.sourceforge.net/index.php?intro+release_notes/server_0_7_2

The prototype 64 bit Itanium release of the Clarens server was updated with more modern dependent packages and deployed on the remaining Caltech CACR TeraGrid systems, as well as on the TeraGrid at the San Diego Supercomputing Center. These deployments are used to host an astronomical image compositing application called Montage as part of the Grist project. The web service "portal" and

associated web interface to this application also contains support for a differentiated trust model, including so called HotGrid users that only have a valid e-mail address as a requirement.

Collaboration

The SuperComputing 2005 conference in Seattle, WA, held during November was used as a venue to demonstrate a distributed High Energy Physics analysis application as part of the annual bandwidth challenge. The Caltech-led team was the overall winner in this challenge event.

Clarens was also used in the Open Science Grid booth to demonstrate the web browser and web service interface to the CMS Monte Carlo Processing Service (MCPS).

I gave a presentation on Clarens and the HotGrid security model at the TeraGrid booth. I also attended the CMS Computing, Physics and Tridas (CPT) week held at CERN during December.

I acted as chair of the weekly OSG Monitoring and Information Systems meetings held to coordinate monitoring systems developments in OSG, and attended meetings of the Grist project mentioned above to aid in the development of the Montage web portal.

5.3.5 SRB

The Global Grid Forum Interoperability working group is planning demonstrations at GGF17 on data access between data grids. As part of this effort, SDSC has established a SRB federation hub in San Diego. The SRB data grids that are linking to this data grid federation include:

- BaBar

- Australian Partnership for Advanced Computing

- DEISA data grid

- Worldwide Universities Network data grid

- Teragrid

The federation is using version 3.4.0 of the SRB. Controlled user access to data in a remote data grid will be demonstrated.

Asssistance has been provided to the BaBar high energy physics experiment team, as they make use of the SRB software for federation of two independent SRB data grids between Stanford and Lyons, France. For the French team in particular, extensions were made to the SRB java admin tool to make it easier and more convenient to connect to additional SRB servers. See bugzilla item 201 for more information (http://srb.npaci.edu/bugzilla/show_bug.cgi?id=201).

Our SRB home page/documentation system has been completely reworked using a MediaWiki infrastructure (as used by Wikipedia) to significantly improve the quality and accessibility of our documentation. In addition, this makes it much easier for the SRB team to update and revise the documentation and, much like wikipedia, allows our user community to also participate in the extension and refinement of the information available. The current system includes many key pages from the old SRB web site, most of our "readme" documents from the distribution, the Scommand man pages, plus some new pages such as a glossary and organization pages. The information is presented in a clean and direct manner, and is significantly hyper-linked, and immediately searchable. This system assists the various SRB communities, including the BaBar researchers, and many others. See our new home page (Main Page) at: http://www.sdsc.edu/srb

The MediaWiki SRB pages also assist with the Federation of SRB systems (Zones) by consolidating and integrating our Zone documentation. The federation of zones is semi-automated as it involves the operation of multiple scripts. Having the zone documentation together in one place, linked together, and updated, assists the administrators to quickly and easily find accurate information. We plan to continue to improve this documentation.

Plans for providing an SRM interface to SRB are being considered in the context of user needs, in particular in collaboration with the Academia Sinica Grid Computing Centre in Taiwan. SRB functionality

is largely orthogonal to that of SRM, as SRB includes both transport (like GridFTP) and archival storage capabilities, and a different manner of cache management. The details of exactly how one wants to integrate SRM and SRB, and how to handle the various levels of authentication, become involved. Thus a careful analysis of requirements if necessary.

We have received a copy of the LBNL Java SRM code (much thanks to Arie Shoshani and Alex Sim), have studied it a bit, and this has lead to the questions mentioned above. In the SRM context, would one want to treat SRB like a archive (like MSS), even tho it is actually a datagrid itself, or as a transport (like GridFTP), or both. A key complication in any of these approaches is the handling of authentication, at the system and client/server levels, and in passing it between components.

Reagan Moore:

The SRB data grid is being more fully integrated with grid services. Interest has been expressed in:

- GSI authentication, which is fully integrated with the SRB.

- GridFTP access, which uses the GridFTP Data Storage Interface to interact with a SRB server. This is running on the NSF Teragrid.

- GridFTP driver, which enables a SRB server to access a remote storage system using GridFTP. This is running on the NSF Teragrid

- GMCAT, which provides a mapping between SRB filespaces and other Grid tools via the Giggle framework. This is running on the UK e-Science data grid.

- SRM interface to the SRB. This is being discussed with Arie Shoshani for version 2 of the SRM.

- VOM interface to the SRB for validation of users as members of groups. This is waiting on the final version of the VOM standard.

A goal of the SRB team is to show that SRB collections can be used with grid computing environments.

Wayne Schroeder:

Asssistance has been provided to the BaBar high energy physics experiment team, as they make use of the SRB software for federation of two independent SRB data grids between Stanford and Lyons, France.

Plans for providing an SRM interface to SRB are being developed. SRB staff (Reagan Moore, Lucas Gilbert, and Wayne Schroeder) met with LBNL SRM staff (Arie Shoshani and Alex Sim) to discuss various options. The most promising approach appears to be to use LBNL's Java implementation of SRM and interface it to SRB via the SRB Java API, Jargon. We plan to evaluate and test this approach in the coming quarter.

SRB 3.4.0 was released October 31st. The continuous automatic testing system (http://www.sdsc.edu/srb/tinderbox.html) found and helped SRB staff resolve various problems during the release preparation period. This system was substantially extended.

The automatic tests on the tinderbox QA hosts can now run non-MCAT-enabled servers and remote (LAN) resources. This is the completion of some development started a while ago, and is a major improvement in the testing environment, as it performs typical cross-network transfers, and exercises some additional SRB logic. This involved some extensions to the autotest script and tinderbox configuration, and some debugging.

A port to using gcc on Solaris (instead of the Solaris compilers) was completed, and this is now running as part of the Tinderbox system. This revealed a new problem (not closing certain descriptors) due to the host's lower limits, and this has now been fixed.

The tinderbox client perl script was modified to stop if an error occurs, rather then cleaning up and starting over as it normally does. This made the system more useful for us, as we can then login and investigate problems on the running system.

The scripts were also modified to install and run multiple SRB Zones (federated but separate SRB instances) on the tinderbox systems. This allows us to include some Zone operations in our automatic testing. Patches for two bugs discovered after the released were developed and are now available from the SRB web site.

Appendix A

Presentations made at CHEP 2006, also posted at http://www.ppdg.net/docs/chep06-presentations.html.

ADAMS, David, DIAL: Distributed Interactive Analysis of Large Datasets

GARCIA-ABIA, Pablo CMS Monte Carlo Production in the Open Science and LHC Computing Grids

GUTSCHE, Oliver, Distributed CMS Analysis on the Open Science Grid

DE, Kaushik, Panda: Production and Distributed Analysis System for ATLAS

NORMAN, Matthew, OSG-CAF - A single point of submission for CDF to the Open Science Grid

FINE, Valeri, A generic approach to job tracking for distributed computing: the STAR approach

HJORT, Eric, Data and Computational Grid decoupling in STAR – An Analysis Scenario using SRM Technology

SHANK, James, Lessons from ATLAS DC2 and Rome Production on Grid3

PORDES, Ruth, The Open Science Grid

YU, Dantong, BNL Wide Area Data Transfer for RHIC and ATLAS: Experience and Plan

RANA, Abhishek, An Edge Services Framework (ESF) for EGEE, LCG, and OSG

RANA, Abhishek, Implementing Finer Grained Authorization on the Open Science Grid

CANAL, Philippe, GRATIA, a resource accounting system for OSG

LEVSHINA, Tanya, The Virtual Organization Management Registration Service

HAJDU, Levente, Meta-configuration for dynamic resource brokering: the SUMS approach

JAKL, Pavel, From rootd to Xrootd, from physical to logical files: experience on accessing and managing distributed data

RANA, Abhishek, gPLAZMA (grid-aware PLuggable AuthoriZation MAnagement): Introducing RBAC (Role Based Access Control) Security in dCache

RABBERTZ, Klaus, CMS Software Distribution on the LCG and OSG Grids

NOWACK, Andreas, CMS Software Packaging and Distribution Tools

STEENBERG, Conrad, JobMon: A Secure, Scalable, Interactive Grid Job Monitor

VAN LINGEN, Frank, On Demand, Policy Based Monte Carlo Production and Tracking, Leveraging Clarens, MonALISA and RunJob

LEGRAND, Iosif, MonALISA : A Distributed Service for Monitoring, Control and Global Optimization

GABRIELE, Garzoglio, The SAM-Grid / LCG interoperability system: a bridge between two Grids

GABRIELE, Garzoglio, A case for application-aware grid services

SNOW, Joel, DØ Data Reprocessing with SAM-Grid

POULARD, Gilbert Poulard, ATLAS Experience on Large Scale Productions on the Grid

PERELMUTOV, Timur, Enabling Grid features in dCache

VESELI, Sinisa, SAMGrid Web Services

VESELI, Sinisa, SAMGrid Peer-to-Peer Information Service

NEVSKI, Pavel, Large scale data movement on the GRID

PORDES, Ruth, Grids - Collaborations and Gateways

Appendix B

Cover page of the OSG SciDAC proposal:

DOE Office of Science Notice DE-FG02-06ER06-04 and LAB 06-04
Sustaining & Extending the Open Science Grid: Science Innovation on a PetaScale Nationwide Facility
Submitted as a Science Application Partnership to: High Energy and Nuclear Physics with Petabytes.
For the period July 1, 2006 – June 30, 2011

Project Principal Investigator, University:
Miron Livny (CS), 608-262-0586 Fax 608-262-9777, miron@cs.wisc.edu Computer Sciences Department, University of Wisconsin, 1210 W. Dayton St., Madison, WI 53706.
Project Principal Investigator DOE, Laboratory; Co-Principal Investigator NSF:
Ruth Pordes (S) 630- 7840-3921 Fax630-840-2645, ruth@fnal.gov Computing Division, Fermilab, PO Box 500, Batavia, Il 60510
Project Co-Principal Investigators:
Paul Avery (S), avery@phys.ufl.edu 352-392-9264 Fax 352-392-8863, 029 New Physics Bldg, University of Florida, Gainesville, FL 32611-8440.
Ian Foster (CS), foster@mcs.anl.gov 630-252-4619, Fax 630-252-9556, The University of Chicago, 1100 E. 58th Street, Ryerson Hall, Room 155, Chicago, IL 60637
Albert Lazzarini (S), (626) 395-8444, lazz@ligo.caltech.edu, Laser Interferometer Gravitational-Wave Observatory (LIGO) Laboratory, 102 E. Bridge, California Institute of Technology, M/C 18-34, Pasadena, CA 91125.
Senior Personnel : (+ Member of the OSG Council. ϑ Member of the OSG Executive Board.)
Boston University: Jim Shank[+](S); *Brookhaven National Laboratory*: Bruce Gibbard(S), Howard Gordon[+](S), Jerome Lauret[+](S), Razvan Popescu(S), Torre Wenaus[ϑ](S); *California Institute of Technology:* Julian Bunn(S), Kent Blackburn[+](S), Albert Lazzarini(S), Harvey Newman[+ϑ](S); *Columbia University:* Jeremy Dodd(S), Michael Tuts(S); *Cornell University:* Laurence Gibbons (S), David Lifka (CS); *Fermilab:* Lothar Bauerdick[+](S), Amber Boehnlein[+](S), Ian Fisk(S), Don Petravick[+ϑ](CS), Ruth Pordes(S), Vicky White[+](S); *Harvard University:* John Huth[+ϑ](S); *Imperial College London:* Gavin Davies[+](S); *Indiana University:* Leigh Grundhoefer[+ϑ](CS), Fred Luehring(S), Rob Quick(S); *Information Sciences Institute:* Ann Chervernak(CS), Carl Kesselman[+](CS); *Lawrence Berkeley National Laboratory:* William Kramer[+](S), Douglas Olson[ϑ](S), Arie Shoshani(CS); *Renaissance Computing Institute:* Alan Blatecky[ϑ](CS); *Stanford Linear Accelerator Laboratory:* Robert Cowles(CS), Richard Mount[+](S); *Texas Advanced Computer Center:* John Boisseau[+](CS); *Thomas Jefferson National Facility:* Andy Kowalski[+] (S); *University at Buffalo:* Mark Green[+ϑ](CS); *University of California San Diego:* Frank Wuerthwein[ϑ] (S); *University of Chicago:* Ian Foster[+ϑ] (CS), Robert Gardner[ϑ](S), Mike Wilde[+ϑ](CS); *University of Florida:* Paul Avery(S); *University of Iowa:* Shaowen Wang (CS); *University of Wisconsin Madison:* Miron Livny[+ϑ](CS), Alain Roy(CS);

Preamble

The OSG effort spans DOE and NSF. This proposal, with the exception of the final section, is identical to an unsolicited proposal submitted to the NSF on 2/6/2006. High Energy and Nuclear Physics with PetaBytes is the driving science application domain for this proposal. Other science application domains like QCD and Astrophysics are also potential beneficiaries of, or are in the process of establishing links to, the OSG. The success of OSG depends on a strong and effective partnership between the domain scientists and Computer Science projects. The partnership built by the SciDAC-1 PPDG and NSF GriPhyN and iVDGL programs is a cornerstone of OSG. As a Partnership, OSG has dedicated *liaisons* to Centers for Enabling Technology, *leverages* capabilities developed by the NSF cyberinfrastructure projects such as CDIGS and NMI (including Condor, Globus, NMI build and test), and *depends* on the work of cooperating SciDAC SA(P)s and NSF PIFs. *Outreach and education* are an integral part of the OSG Consortium which will use a broad range of mechanisms to disseminate the OSG concepts, capabilities and software.