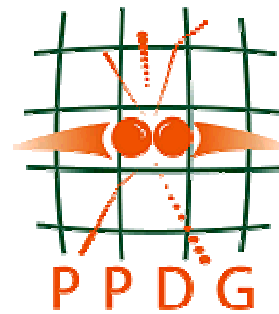


**Particle Physics Data Grid:
From Fabric to Physics**

**Quarterly Status Report of the
Steering Committee,**

April - June 2005

1 August 2005



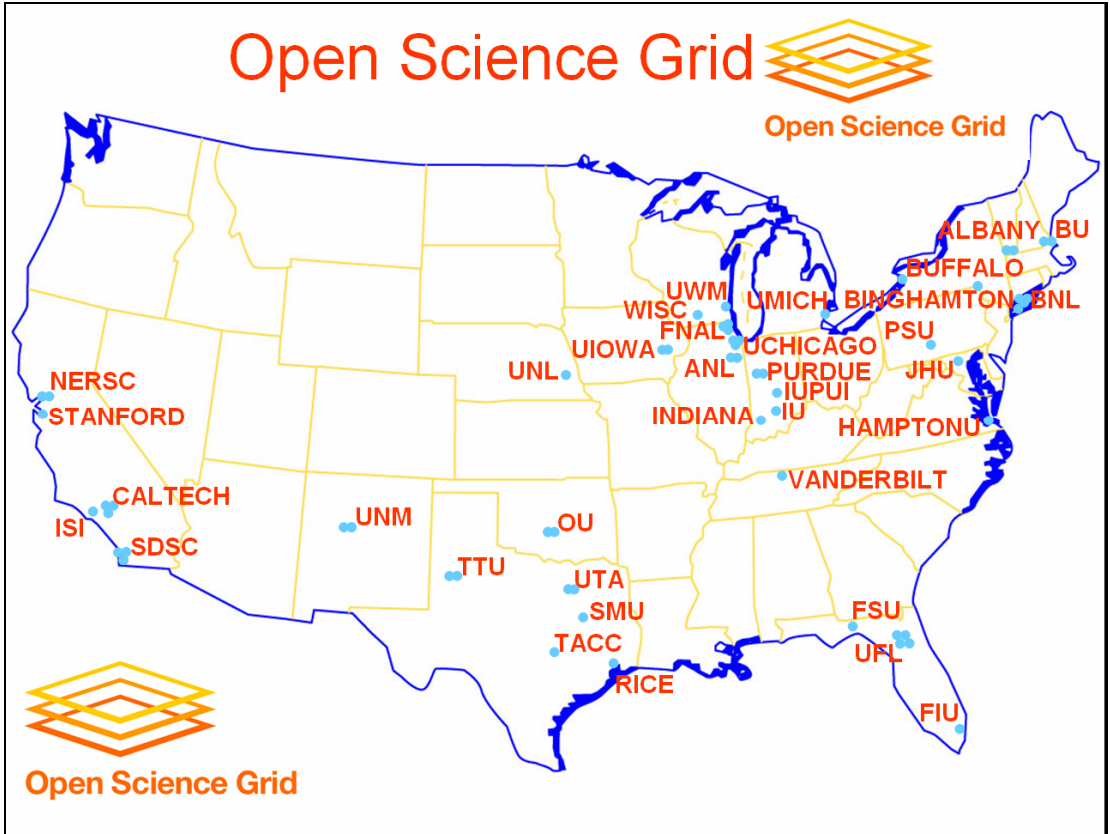
1 Project Overview	1	5.1.5 STAR.....	10
1.1 Papers and Documents.....	2	5.1.6 ALICE	14
2 The Common Project.....	3	5.2 Facilities	14
3 CS-11 Data Analysis Working Group	3	5.2.1 Jlab.....	14
4 Collaborations	4	5.2.2 BNL RCF/ACF.....	14
4.1 Open Science Grid.....	4	5.2.3 FNAL.....	14
4.1.1 Documentation, Education, Outreach..	4	5.2.4 NERSC/PDSF.....	14
4.1.2 Governance & Management	4	5.2.5 SLAC.....	14
4.1.3 Operations & Support	4	5.2.6 Collaboration with IEPM, Network Performance Monitoring.....	14
4.1.4 Security.....	4	5.3 Computer Science & Middleware	17
4.2 EGEE, LCG.....	4	5.3.1 Condor	17
4.3 DOEGrids PKI	5	5.3.2 Globus – ANL	17
4.4 TeraGrid	5	5.3.3 SRM –.....	18
5 Single Team Reports	5	5.3.4 Caltech.....	19
5.1 Experiments.....	5	5.3.5 SRB	21
5.1.1 ATLAS	5	6 Appendix	22
5.1.2 BaBar	5	6.1 Milestones	22
5.1.3 CMS.....	5		
5.1.4 D0	10		

1 Project Overview

The major focus of effort this quarter was on deploying applications, resources and services on Open Science Grid, as illustrated in the map, below, showing the sites participating in OSG for the grand opening ribbon cutting ceremony on July 20.

PPDG also published Science Accomplishments at the time of the June 2005 SciDAC PIs meeting.
<http://www.ppdg.net/docs/may05/index.html>

A live view of the active OSG sites is shown at <http://osg-cat.grid.iu.edu:8080/>.



PPDG effort is contributing leading roles in coordinating the deployment of the OSG infrastructure as well as key services for storage management, security, and discovery.

CPUs accessible from OSG number about 13,000 and there are SRM based mass storage systems available at LBNL, Fermilab and BNL.

1.1 Papers and Documents

Reports, Documents and Papers	Date/Version
OSG-205 OSG Accounting System Requirements	doc (21Jul05), V1
OSG-134 Open Science Grid Authorization Architecture - Role Based Access Control	doc (20Jun05) V1
OSG-47 OSG Operations Model Roadmap	doc (14Mar05) V7
OSG-28 Open Science Grid Deployment Plan	doc (27May05) V6

In addition to these papers many PPDG participants contributed to numerous working documents and presentations for OSG.

PPDG sponsored at OSG applications workshop at SLAC which had representation from ANL GADU, University at Buffalo GRASE, and TeraGrid Nanohub projects, as well as a presentation from the LCG

ARDA distributed analysis project. <http://osg-docdb.opensciencegrid.org/cgi-bin/DisplayMeeting?conferenceid=20>

The 2005 OSG Consortium meeting was hosted by LIGO at the University of Wisconsin-Milwaukee.

2 The Common Project

The common project effort focused its activities for the past quarter on the Provisioning Activity¹ for deploying the Open Science Grid. Dane Skow serves as co-chair of this Activity (with Leigh Grundhoeffer of iVDGL) and there are contributions from many PPDG members. Provisioning in OSG is the process of taking a tested release candidate software stack and turning it into the deployed grid-wide production services, including the documentation, software cache and support infrastructure that sites use to install and run the OSG production infrastructure.

Preparations and mentoring site admins was declared top priority for Common Project participants for this quarter. The results were that almost all major Grid3 site was able to transition to OSG Production prior to the July 20 "Opening" of the Open Science Grid. Feedback from several participants was that this deployment was much simpler than Grid3 startup and that progress has been made.

The Privilege Project products, PRIMA and GUMS, have undergone significant documentation improvement cycles and are now running on many OSG production facilities. Effort continues to be expended with the VDT team and INFN VOMS providers to improve the reliability and usability of this software stack. First steps were taken this quarter to determine what changes will be required with a migration to a GT4 basis of the VDT's included Globus Toolkit.

The Web Service Discovery Service for OSG, based on Clarens, is deployed in OSG Production and is beginning to see first use. The extension of this machinery to a Software Discovery service was presented at the SLAC Applications meeting in June and at least one major VO was keenly interested in this capability.

Development work on a resource accounting system for OSG is being done in an OSG Activity with participants from SLAC and FNAL. A requirements document and initial system design was presented at the Consortium meeting and preparations are underway for a Blueprint architectural design review at Buffalo in August.

The gPLAZMA Authorization framework for storage (a parallel to the PRIMA framework for GRAM gatekeepers) was deployed for testing in dCache and is performing well. It is currently under evaluation for distribution with the dCache deployment planned for later this Summer.

The Grid Exerciser (GridEx) has been instantiated on the OSG Integration Grid and an Activity started to address issues raised in Grid3 and with initial tests. Options for inclusion of a variable payload for GridEx (eg. validation jobs rather than null jobs) are being investigated and this may be one method of addressing an OSG-wide opportunistic scheduler for OSG Release 0.4 (targeted for Nov. 2005).

3 CS-11 Data Analysis Working Group

The activities of the data analysis working group are documented at <http://ultraviolet.caltech.edu/gaeweb/wiki/index.php/ppdg-cs11>. There are two types interactive data analysis being discussed, a pseudo-interactive style that is based upon a rapid-response batch mode processing, and a true-interactive style where there is control from the user interface to all of the associated running processes. A set of requirements for these with a view of issues related to deployment on OSG sites is at <http://ultraviolet.caltech.edu/gaeweb/wiki/index.php/InteractiveAnalysis>.

¹ <http://osg.ivdgl.org/twiki/bin/view/Provisioning/>

The group working on DAGS/JAS² (Tech-X & SLAC) are preparing to deploy DAGS on some sites of the OSG Integration Test Bed (ITB).

4 Collaborations

4.1 Open Science Grid

As explained above PPDG made significant contributions to the OSG deployment. Fermilab FermiGrid and SLAC production systems were both made accessible to the common infrastructure. The Privilege Project components were deployed at many sites to provide both Site and VO control of local accounts and role based access to local resources.

4.1.1 Documentation, Education, Outreach

The new OSG web site was developed for the opening of OSG. <http://www.opensciencegrid.org> and documentation to describe how organizations and sites join OSG was posted.

4.1.2 Governance & Management

The Charter <http://osg-docdb.opensciencegrid.org/0000/000025/004/Charter.pdf> and initial By-Laws <http://osg-docdb.opensciencegrid.org/0000/000070/008/ByLaws-1.0.2.pdf> for the OSG Consortium were ratified. A solicitation for pre-proposals to contribute to the OSG needs and roadmap was sent to all members of the OSG mail list. To date this has resulted in 30 submissions.

4.1.3 Operations & Support

The operations model of distributed support centers is being implemented and the various support centers³ are in progress of learning how to participate and getting their communications procedures set up. At present there are two weekly teleconference meetings to cover issues and questions; a Monday meeting covering both Grid3 and OSG and a Wednesday meeting focusing only on OSG. It is planned to drop the Wednesday meeting in September when Grid3 support is concluded and have the Monday meeting focus solely on the OSG production grid issues.

Three people representing OSG attended the 2nd EGEE/LCG Grid Operations Workshop⁴ in Bologna in May. This was very fruitful in developing plans for interfacing and interoperation at various levels ranging from user support and trouble ticket exchange to resource compatibility for applications that span the EGEE and OSG grid infrastructures (e.g., LHC experiments). It is planned that future grid operations workshop will be jointly organized by OSG and EGEE, and then next one is scheduled for Sept. 27-29, 2005 in the UK.

4.1.4 Security

Continued collaboration with the EGEE and LCG Joint Security and Security Middleware Working Groups have resulted in some changes to the AUP. There is a plan to have a joint meeting with the EGEE middleware working group in September.

4.2 EGEE, LCG

The SAM-Grid / LCG interoperability project is making available to DZero the resources provided by the LHC Computing Grid (LCG) through the SAM-Grid system. In the months between March and June, the collaboration put in place a “proof-of-principle” test bed, which demonstrated the feasibility of the integration.

² <http://grid.txcorp.com/confluence/display/DAGS/DAGS+Home>

³ http://osg.grid.iu.edu/OSG/index.php?option=com_content&task=view&id=37&elMenu=Grid%20Support

⁴ <http://grid-it.cnaf.infn.it/workshop/>

The new Glue Schema V1.2 and Generic Information Providers are included in VDT and optionally installed on OSG. Work on a sensible configuration for deployment on OSG sites was done as part of the OSG Interoperability Activity and Monitoring and Information Systems technical group. US CMS has now tested submission of MOP jobs across both LCG and OSG and expects MOP (a PPDG developed component) to have extended use by CMS for simulation production.

4.3 DOEGrids PKI

There were no major topics or issues that affected PKI service this quarter. The PKI team contributed to a GGF document on OCSP Requirements for Grids⁵.

The Certificate Management System software used for the current CA has been EOL by the vendor (SUN) and a new product and vendor has been chosen as an eventual replacement for the CA service. The interfaces and service details will change in this transition and there is a need to look at community requirements and desires for modifications to the interfaces and services. This will likely take some effort in the next two quarters.

4.4 TeraGrid

The OSG Interoperability Activity tested submission of US ATLAS and US CMS simulation production jobs on TeraGrid resources. Other Interoperability work is starting to better understand what gateways are needed between the Grid infrastructures to allow VOs to support cross-grid job execution and data access.

5 Single Team Reports

5.1 Experiments

5.1.1 ATLAS

Jerry Gieraltowski

The work effort in this quarter was focused on enhancing the performance capabilities of the ATLAS executor software for US-ATLAS. To provide an easier installation and maintenance interface for the user both the client software and the job management software were merged into a single "Capone" product. New (in development) software from the Condor team which supports the submission of multiple dags to a single dag-manager was incorporated into the latest Capone product and new Capone APIs were constructed to support this multidag batching functionality. A performance run was made the last week of June to determine whether a single Capone job submission environment could reliably manage a minimum of 5,000 jobs. Marco Mambeli and I setup and executed this run and were able to effectively manage between 5,000 and 7,000 jobs from the single submit host. Of course, as the number of jobs increased, so did problems at the compute nodes - but that is a separate story!

5.1.2 BaBar

5.1.3 CMS

Abhishek Singh Rana

Common Project areas: Data Management and Security

=====

We focused on familiarizing Open Science Grid participants with security/authorization in CEs and SEs (Role-based Access Control or RBAC). Elaborated on design, functionality, deployment and usage.

⁵ https://forge.gridforum.org/projects/caops-wg/document/OCSP_Requirements_for_Grids/en/5

Attended GriPhyN All-hands Meeting (April 27-29, 2005 at ANL, Argonne) and OSG Applications Workshop (June 1-2, 2005 at SLAC, Menlo Park). This was accompanied with an introduction sheet 'FQANs Distilled' intended to serve as a quick and complete reference for VO management personnel who are expected to declare FQANs for their VOs. [[osg-doc-113](#), [osg-doc-153](#), [osg-doc-154](#)]

Our previous work on gPLAZMA continues. An immediate aim is production-level deployment and usage in the next quarter. This will introduce RBAC, SRM callouts and GridFTP callouts in a few SEs (SRM-dCache with gPLAZMA module). [[osg-doc-174](#)]

Effort is also directed at management of SRM-dCache deployment at UC-San Diego Tier-2 Center. The Center is exploring a sophisticated deployment with migration from a single head-node to a collection of 6-7 head-nodes. This scheme will abate system- and network-level bottlenecks, and provide improved performance of WAN data transfers.

Muhammad Anzar Afaq

In this quarter my focus remained on Dataset Publishing Service and Dataset Book-keeping Service (DBS).

Publishing Service: is a Tool developed by me, based on FNAL-Runjob. Its being used CMS-wide for Publishing MC Datasets after being produced in "Production". These published datasets are then further used to run more production steps in production chain or perform user analysis. For example, at completion of Hit Simulation, a Hit Dataset is published that can be used by a Digitization step and later in a user analysis job. Publish Service interfaces to RefDB at CERN, extract detailed information about a dataset, performs catalog creation and meta data attach operations required by CMS Applications/Framework. And then publishes everything to a central Federation area or if so desired to PubDB (another tool used CMS-wide for running Grid-Analysis). Catalog manipulation and publishing is a cumbersome process in current Event Data Model. Its made very simple by Publish Service. New ideas about catalog organization are explored at CMS and they are all being supported by Publish Service. The catalogs become Grid-available for individual analysis (like on USCMS User Analysis facility), or via wget to wider community, or via other Grid-WorkFlow tools like CRAB and MCPS via direct or indirect access.

DBS: CMS is moving towards a new Event Data Model (EDM) and also changing the way dataset(s) are presented to users. The new set of services that will replace current RefDB based structure include Dataset Book-keeping Service (DBS) and Dataset Location Service (DLS). DBS is going to be the tool that will keep track of various datasets produced by MC-Production/User Analysis or Online farms. User/Applications will need to talk to DBS to get the datasets they need to work upon. I am an active member of DBS development team and Data management team in general, which is a bigger umbrella. I introduced ideas of developing a multi-layered server for DBS with a Query Objects Layer, that combines create objects of interesting data items/tables in the database to be presented to what we call business logic layer. The idea is later discussed in team and groomed to add details. A prototype DBS-Server is being developed in C++ and I am taking active part in its design and development.

Beside these major tasks I still support MOP/McRunjob and CMS Production at USMOP in general.

I also take part in active discussions about MCPS, the next MC-Production/Analysis tool for CMS/USCMS.

Conrad Steenberg

Collaboration

A version of the C++ based Clarens client for the ROOT analysis application was committed to the official ROOT source tree, and made available to all developers as a default component of the system.

I collaborated on the development of the JobMon project that uses the Clarens framework in the context of a large compute farm to allow bi-directional communication with jobs running on worker nodes. This is an important step towards one of the original design goals of Clarens: to enable fully interactive remote analysis. More information on JobMon can be found at <http://jobmon.sourceforge.net/>.

The HotGrid certificate request service was stabilized to the extent that it was usable enough to be used in a Grid security tutorial, as described later. This service as well as the shell command execution service are used as the basis of an Astronomy portal for the GRIST experiment.

I also continued in my role as co-chair of the Open Science Grid Monitoring and Information Services technical group.

Papers

I was a co-author of the following papers during this period:

M. Thomas, C. Steenberg, F. van Lingen, H. Newman, J. Bunn, A. Ali, R. McClatchey, A. Anjum, T. Azim, W. ur Rehman, F. Khan, J. Uk In "JClarens: A Java Framework for Developing and Deploying Web Services for Grid Computing", Submitted to the International Conference on Web Services, Orlando, July 12-15, 2005

F. van Lingen, C. Steenberg, M. Thomas, A. Anjum, T. Azim, F. Khan, H. Newman, A. Ali, J. Bunn, I. Legrand "Collaboration in Grid Environments using Clarens", To be presented at 9th World Multi-Conference on Systemics, Cybernetics and Informatics, Orlando, July 2005

F. van Lingen, J. Bunn, I. Legrand, H. Newman, C. Steenberg, M. Thomas, A. Anjum, T. Azim, The Clarens Web Service Framework for Distributed Scientific Analysis in Grid Projects, Workshop on Web and Grid Services for Scientific Data Analysis (WAGSSDA), Oslo, June 14-17, 2005

Education and Outreach

A multidisciplinary workshop was held from June 8 to 10 by the Computing in High Energy Physics Research and Education Outreach (CHEPREO) unit at Florida International University for students in Computer Science and Physics graduate students. The workshop focused on the computing needs of High Energy Physics, including high speed networking, distributed computing and specific HEP data analysis applications. A description of the workshop can be found at http://ultralight.caltech.edu/portal/workgroups/e_and_o/index.html

As part of the program I presented three tutorials. The first focused on Grid security concepts, including public key encryption, X509 certificates, single sign-on and virtual organizations. As part of the tutorial participants had the opportunity to request and use their own personal certificates using the Clarens HotGrid service.

The second tutorial introduced the Clarens web services framework and its security model, while the third allowed participants to create their own Clarens services and access these services using their new personal certificates.

Suresh Man Singh

Here are two main accomplishments in this quarter.

1. Following the OSG Integration and Operation Provisioning workshop held in University of Chicago (Feb. 15-17, 2005), I am actively involved in testing out various releases of ITB software and reporting the feedback to OSG-ITB. For this, I upgraded our existing six node ITB cluster to RHEL3 from Redhat 7.3 using Rocks 3.3. One of the new features in the ITB software suite is the use of Grid User Management

System (GUMS) authentication mechanism. This has effectively replaced grid-mapfile based authentication. Other middlewares that were extensively tested were Condor batch scheduler; GSIFTP, fork and condor job managers as part of Globus; Gridcat, MonALISA and MIS as part of the Information and Verification infrastructure; Generic Information Provider (GIP) as part of OSG-LCG inter-operability test. In all of these cases, we installed DOEGrids signed host and service certificates. Various VO jobs (USCMS, ATLAS, SDSS, GRASE, iVDGL) were submitted and tested out their functionality in this testbed cluster.

2. As part of the OSG deployment activity, I did necessary preparation to make smooth transition of our 33 node production cluster from Grid3 to OSG. This involved following major upgrades:

(i) we replaced old two Dell PowerConnect switches by very high end Force10 E600 Terascale managed switch. This switch has ample slots for the future growth of cluster and WAN network ports. This has practically eliminated many network related problems like latency, throughput and number of ports.

(ii) This cluster was rebuilt from scratch using Rocks 3.3 cluster management software. Currently, it's running 2.4.21-20.ELsmp kernel but will be upgraded to latest kernel release to circumvent some critical kernel related problems.

(iii) I installed a GUMS server (gums.cacr.caltech.edu) to provide prima callout based authentication service to all the Caltech hosts that participate in OSG.

(iv) I installed OSG-0.2.1 release candidate in the Gatekeeper with the sitename as CIT_CMS_PG. A 4U 1.5 TB NAS appliance has been used for three OSG related directories (\$APP, \$DATA and \$TMP). Condor 6.7.7 has been configured as job scheduler with appropriate jobmanager. Monalisa and Ganglia has been configured to monitor cluster health status information. DOEGrids signed host certificate have been used in the gatekeeper. The cluster is now fully engaged in accepting and running jobs submitted by various OSG VOs.

Michael Thomas

Summary

=====

Work this quarter was once more focused on improving the core JClarens server and its integration in the OSG and Ultralight activities. Additional E&O work maintained this focus.

Ultralight

=====

More progress was made on packaging the Sphinx grid scheduler as a JClarens service. All Sphinx components are now included in a single rpm that can be installed alongside the latest JClarens release. This has allows us to deploy Sphinx on 4 different machines in the Ultralight testbed.

OSG

====

OSG contributions include the continued development and support for the Discover Service as described above. I have also volunteered to run a public instance of the Discovery Service at Caltech for OSG testing purposes. This public Discovery Service instance is already up and running on ultralight.caltech.edu and has been successfully used by the SRM/DRM team and the voms-admin service developers. Plans have been made to help set up a second public Discovery Service instance for the OSG Grid Operations Center.

In the TG-MIG working group I participated in discussions on ways to extend the Discovery Service to handle more types of information than Web Service descriptions. To better understand the OSG requirements for this task, I helped to document the process of configuring the LCG Generic Information

Provider (GIP) for the OSG on the OSG Twiki documentation page, and configured the GIP for the CIT_CMS_OSG cluster at Caltech.

Meetings

=====

During the quarter I attended many meetings and workshops including:

OSG Applications Workshop at SLAC on June 1,2: I presented the Discovery Service and brought Tahir Azim to present the Software Discovery Service to the applications group. There were several interesting side conversations with application developers on how they would like to use the Discovery Service with their applications.

OSG Kickoff meeting at UERJ in Rio de Janeiro, Brazil: I continued our close collaboration with the UERJ team and we were able to get their 100 node cluster installed and validated as part of the OSG integration testbed at this meeting. At the end of the meeting I showed a short demonstration showing that the cluster was being recognized as part of the OSG ITB and was actively running CMS Monte Carlo production jobs.

ICFA conference in Korea. The talk described how one can perform data analysis using the Ultralight testbed. The talk can be found at:

http://ultralight.caltech.edu/gaeweb/portal/presentations/2005/05mike_korea/AnalysisExample_Korea_mt.ppt

CMS_DCT: I attended most of the CMS DCT meetings where I actively promoted the use of the new Software Discovery Service for CMS.

TG-MIG and OSG_ITB: I attended the TG-MIG and OSG-ITB meetings to discuss the use of the Discovery Service in the OSG.

CS11: In the CS11 meetings I have helped to work to define a set of requirements for the OSG for performing interactive data analysis on OSG resources. Interactive data analysis has some special requirements that are not present for unattended batch data processing. I wrote a summary of this ongoing discussion at:

<http://ultralight.caltech.edu/gaeweb/wiki/index.php/InteractiveAnalysis>

Shih-Chieh Hsu

We developed an interactive GRID job monitor system, JobMon, since January 2005. The JobMon system is an open source, secured, allows users to view process status and log files interactively as the job runs in batch system. This facilitates debugging and reduces unnecessary resource consumption due to user error. JobMon is written by Python language and based on a web-service system, Clarens Clarens has the primary responsibility of authenticating the user and implementing the access control that limits users to interacting with jobs for which they have permission. The JobMon system is included into DISUN project and might be used in LCG.

jobMon is available via sourceforge at: URL: <https://sourceforge.net/projects/jobmon/>

Frank van Lingen

Have been working on the Clarens based MCPS Web Service. This will serve as a frontend to enable users to submit jobs by selecting (potentially) complicated workflows on the server side specified in runjob. The aim is to only expose a limited set of parameters to the user and allow users to upload their own pluggable code. The service would typically reside at Tier2s and handles the (complexities) of job submission, job monitoring and CMS software installations. We have investigated the use of existing workflow tools such as Chimera but the amount of work to integrate runjob and Chimera proved to be substantial.

A prototype has been successfully tested with a python client. Work is underway for a browser client and we are currently testing it with real runjob code at UCSD. More information on the progress of this can be found at:

http://ultralight.caltech.edu/gaeweb/portal/proposals/2005/05MCPS_Web_Service/status/

Two weeks ago I attended the ICCP in OSLO to present a paper on Clarens in a Web Service workshop. A summary of this conference can be found at:

http://ultralight.caltech.edu/gaeweb/wiki/index.php/ICCP_Conference_OSLO.

Part of my time has been spent on editing an article for CISE (<http://www.computer.org/portal/site/cise>) regarding the network as actively managed resource. This is part of the UltraLight project.

I also prepared some tests for web services wrappers around pubdb and phedex db. that are deployed at

Recently two Pakistani students have arrived at Caltech. One of them (Ahsan) is working with me on providing a webstart based GUI to Clarens services. This will enhance the functionality compared to the existing Javascript based GUI. The idea is to have a "plug in" framework such as JAS (but than for Clarens Web Services GUIs) and the ability to connect to multiple servers at once, and (cross) use results of distributed services. For example, being able to find a file on several clarens servers or querying multiple catalogs at the same time.

5.1.4 D0

The D0 SAMGrid project worked making available to DZero the resources provided by the LHC Computing Grid (LCG) through the SAM-Grid system. A project plan was defined, a prototype implementation completed through a visit to IN2P3 in June. A report of the work is available at <http://www-d0.fnal.gov/computing/grid/doc/SAMGrid-LCG-integration-Lyon-report.pdf>

Working proceeded on getting the VOMS/VOMRS available for production use by D0. The the users and services (taken from SAM database) are registered with the Dzero VO at <https://fermigrid2.fnal.gov:8443/vomrs/vo-dzero/vomrs>. Work proceeded on a mechanism for the SAM-Grid job management to regulate the flow of jobs to execution sites through throttling policies. This is documented at <http://home.fnal.gov/~garzogli/JobSubmissionPolicies.pdf> . D0 has started to explore how to provide resource selection of OSG, and how specify a common abstract to concrete job description as an extension to the existing SAM capabilities.

5.1.5 STAR

5.1.5.1 General

This quarterly was marked by several interesting emerging interests in Grid technologies and connectivity from within our experiment. We are first proud to have our STAR collaborators from the Instituto de Fisica da Universidade de São Paulo showing not only interest but bringing in the STAR Grid the first fully functional South American site participating in a Nuclear Physics experiment. We were then blessed by learning that some of our European users have tried (and succeeded) running their user analysis jobs on our STAR Grid. This was truly encouraging as it was spontaneous (just a curious user).

A possible concern everyone has is that Grid is its infancy. We are now in STAR in a position however to "unleash" or users on the Grid for simple analysis ... Several questions could be raised:

- Technology wise, what will happen if all users try?? Will the infrastructure scale with the number of jobs (STAR users submits several 10k jobs A DAY, which is order of magnitude higher than the average grid exercise)?? Will the network allow for large/massive amount of output to be moved around on the grid?? Several components are missing to make the Grid a production system (error recovery and event correlation, consistent job tracking and control, no global ticket system exists so difficult to trace if something goes wrong). How to support the needed activities necessary to complete this promising sign and make it a scalable reality when the future funding seem to be more on the "artistic blur" side than a tangible reality?

- What to expect in terms of geo-politics when our most remote institutions will want to try to use and exercise accessing resources via a Grid gatekeeper??

As many questions we will see answered under fire soon (it is our intent to understand those issues sooner than later). An extract of one of our user, Thorsten Kolleger, follows as a testimony:

The idea behind was to have the ability to submit jobs to a STAR Grid site where the data resides and get the output (in my case some histograms/trees in a root file) back to Frankfurt. Remote interactive analysis is a pain even with the high-speed/high-bandwidth connection between Frankfurt and BNL, so we are doing the final interactive analysis in Frankfurt anyhow. We have enough disc space here [at our institute] ... Since the jobs copy the output directly back they don't use the data disks at PDSF/RCAS [leaving more space to other users], and you now how many jobs crash because [the space at the major facility] fills up [whenever we are close to a major conference]..

Thank you Thorsten, what a better way to express how important this activity is for a running experiment than a user who spontaneously volunteers such comment ...

5.1.5.2 Infrastructure

Discussions were held with the RCF as per the difficulties we are having in terms of support. It appeared BNL had three different ticket systems making the follow-up difficult and causing unusual delays. A solution was provided: although imperfect (RHIC/STAR cannot Authenticate to the ticket system migrated to the Atlas computing side), Jason Smith has been paying particular attention to our problems and the improvement has been noted.

We proceeded as well with our initial plans to drive into the Grid effort more internal manpower. Wayne Betts, our computer support specialist has been spending some time understanding the details of deploying the software stack (VDT, OSG) necessary for making a gatekeeper functional. Wayne brought up a fully functional OSG-ITB gatekeeper; this is very good news for the STAR team at BNL as per our past experience and learning curve.

Gabriele Carcassi from the RACF has been providing an exemplary support for VOMS, VOMRS and GUMS. Several issues came up with user certificate not being propagated in tie (or at all) and Gabriele always promptly acted upon our report.

Our deployment of the OSG stack at LBNL (using the SGE jobmanager) appeared to be functional. The later deployment of a similar stack at our Sao Paulo site (also using SGE) revealed that the provided SGE jobmanager is incomplete (at best) and non-working as distributed. This issue is beyond the MIS-CI patch Iwona and Bockjoo discussed. The module was just not functioning at all and would abort with a status 256 (i.e. 1 with no error string) which I traced as reproducible whenever the SGE dispatcher is called with improper environment variable set. Alex Suaide patched his install and so did Iwona later. We will document this and send it back to the community.

5.1.5.3 SRM and production demonstrator

This activity is mainly supported by Eric Hjort (LBNL/STAR) with help from Alex Sim (LBNL/SDM). As the experiment has entered into a production operation-driven mode (data mining is constant without interruption), Eric continued the data transfers this quarter, using SRM+RRS: Apr 4.0 TB, May 5.5 TB, June 5.8 TB to a total 15.3 TB and 120k files. The data rates were not very impressive and this is a concern for the long term as our demand in throughput would definitely benefit from rates greater than 40 MB/sec (we can barely do 10 MB/sec to LBNL average). At this rate, data mining at Tier1 centers is not possible as we produce from 150 TB now (with planned for up to 1 PB in the outer years). We therefore continue to limit ourselves to transferring pass2 files and use our Tier1 center for user-end data analysis.

In the last quarter, we had a meeting at BNL where several use case and Grid prototypes were discussed. Especially, we believe a job output should be delegated to a local SRM, this SRM taking care of transferring the data. Not only this approach makes the meaning of "loosely coupled components" have a meaning, but by proceeding by delegation, one segregate SE and CE in a clear fashion, leaving to the expert tools aimed to manage Storage the task to do so. We discussed with the SRM developers of the case

we would like to prepare: running simulation jobs and handing the output to SRM. Later, this use case would be extended for data mining on Grid to sites offering sufficient network bandwidth and/or local buffering or storage.

To achieve this goal, two main enhancements to SRM were needed and one major issue came up:

- Logical filenames needed to be supported: this enables a file transferred from local storage on a worker node to the gatekeeper to be referenced by a second SRM transfer for transfer to final destination. In other words, the chain of SRM to the final destination will use a LFN.
- Support of secondary locations for proxies and certificate. This enables authentication on worker nodes. Testing and development led to SRM release of v1.2.8 in late. Individual pieces of the new release have been tested but it is yet to function for STAR grid jobs.
- A final issue showed itself during our testing and has to do with the way SRM works. Invoking the client would usually initiate a transfer to a remote server but that server would then use a callback mechanism for acknowledgement. This would work only if the firewall permits inbound connections. New versions of the SRM client were issued disabling the callback and use instead a polling mechanism (outbound) of the remote SRM status for the given operation. This mechanism is best suited for firewalls.

We believe we now have all the components necessary to resume our testing and demonstrator. We hope to come with a fully functional use case and put it into operation and large scale testing on OSG. We discussed an approach of “edge service” where an SRM service would be running on a gatekeeper and performing the handshake with a remote SRM. This would be suited to clusters isolating their worker nodes from the network (no outbound, only connections through a single or few nodes allowed).

5.1.5.4 STAR Unified Meta-Scheduler (SUMS)

This activity is mainly supported by Levente Hajdu. Levente has prepared and distributed local configuration files to our sites running the STAR Unified Meta-Scheduler (SUMS) and wanting to submit to OSG gatekeepers. Especially, our most recent site adopting the scheduler is São Paulo and the initial help in setting up the client submission is really appreciated by local site administrators.

Levente also made several extensions to SUMS.

- Because a given centralized storage solution may have a different resource requirement than another centralized storage solution (for examples AFS cannot be stressed like a a NFS file system and none of the two are comparable to a PanFS scalable solution in terms of simultaneous IO operation), the batch system dispatchers was modified to accept between zero to an infinite number of Resource Requirement rules (java objects). Additionally, each rule may apply based on a regular expression pattern matching of the targeted storage element. For example, a default rule may apply to all storage element, another apply to certain path or one to certain device class.
- Because of the different ways of copying and/or moving files between sites on the grid (that is, depending on which service is available at which site: SRM, globus-url-copy, etc ...) it was necessary to implement a method for selecting the copy back command from a particular site. The U-JDL already provided hooks for these (which would later be part of a service discovery mechanism) but never fully implemented. This is now functional and tested and we will be looking into automatically discovering services without the use of a hardwired list of possibilities.

5.1.5.5 Outreach and planning

We made a presentation at the ICFA Workshop in Daegu/Korea in May 2005. While Jerome Lauret was invited for a presentation, the timing did not allow for physical presence. The talk was prepared and Doug Olson presented it on his behalf over the internet while in Europe. This inspiring event stresses the importance of a strong communication network (Access-Grid, VRVS or similar). Such mean greatly enhance our ability to reach faraway corners of the World and maintain presence at a low cost.

STAR participated in

- the OpenScience Grid council meeting in Madison in May which led to the elaboration of the governance procedures amongst other action item and the formation of a committee designed to define the scope and a framework for pre-proposal submission. The committee was composed of Vicky White, Paul Avery, Lothar Bauerdick, Jim Shank, Miron Livny, Jerome Lauret, Ruth Pordes, and Mark Green. The pre-proposal framework document was presented to the council early June.
- The OpenScience Grid all-hand and inauguration meeting in Milwaukee. The council met on the last day and pre-proposals were reviewed. Several committee were formed to review the consistency of the diverse proposals in several areas. Especially, Jerome Lauret agreed to participate in the discussions of the TeraGrid (along with Raymond Bair, Ruth Pordes, Dane Skow, Ian Foster and Charlie) and on Storage (along with Ruth Pordes, Don Person and Ian Foster)

Our São Paulo site joining the OSG was advertised on [Science Grid This Week](#) at www.interactions.org.

5.1.5.6 Other activities

Lidia Didenko (STAR production coordinator) continued her participation in Grid exercises and productions. In fact, most of our simulation has been running on the Grid in recent times and besides the sometimes intriguing technical difficulties, the success rates remained high.

With the venue of a new site, the question of a global file catalog was raised again and will come as a needed component. The users would like to take advantage of the already placed data sets on the Grid, i.e. following the STAR Data-Grid oriented structure (job moves where the data is) rather than trying to achieve a computational grid (where data is moved to the CPU). Our Catalog was designed and modeled after a “federation” model but little incentive existed before with only two sites. This activity will be resumed in the next quarter.

We continued our activities on the Xrootd front. Pavel Jakl, a student from the Czech Republic came to work with us during the summer and continue his evaluation and study of Xrootd. We met with Andrew Hanushevsky, Arie Shoshanie and Alex Sim at Berkeley. The diverse design issue of interfacing Xrootd with SRM were discussed (there is work to do, it is not immediate) and we believe we all understand well now the roadmap to this project completion. In the mean time, we discussed some technical immediate details relating to Xrootd: its functionalities as per self-managing disk space (cache) conflicts with pre-populated disks (like any community would have if transitioning from rootd to Xrootd). Pavel modified Xrootd to re-conciliate both world and the home-made version is under testing. We will as usual pass on our modifications.

Valeri Fine started to evaluate PACMAN. Our hopes being that PACMAN would greatly help with the automation of the packaging and deployment of our code and layered libraries at a remote site. Valeri encountered several difficulties (cache could not be removed, command got stuck etc ...) and got in touch with Saul Youssef who has provided outstanding support and attention. Some problems were found and fixed (some remains and brought to Saul's attention). We gave created a PACMAN cache for STAR, have it populated with our “most commonly used” programs and testing its integration with SUMS. We hope to have this brought to conclusion by the end of next month. If so, it would resolve the issue of deployment of code at remote sites NOT being STAR reserved and therefore make new resources at reach with little to no maintenance.

Valeri has completed the Logger project as well. This project is aimed to provide a consistent login interface to ANY STAR programs, scripts etc ... from job submission to job completion. This component will serve as a powerful job tracing mechanism. Late discussion involved implementing/supporting redirect to other service information. For example, an output file would be handed to an SRM service, the job would hand and the logger would add in its entries a redirect reference to an SRM service for inquiring the SRM request ID relating to the job. The Logger is under testing in our development area and will be fully integrated in the next release of the STAR software stack.

5.1.6 ALICE

A meeting was held at LBNL on June 10 to discuss issues about ALICE computing and OSG (<http://dtd.lbl.gov/~olson/alice-10jun05/>). ALICE has a model of deploying a gateway from the AliEn system to interface to grid systems already in place for the LCG. It was decided that setting up such a gateway to OSG is the appropriate plan and Joseph Ghobrial at U.H. has started working on this. In effect, it amounts to having an AliEn service submitting jobs through Condor-G to OSG sites in much the same way that an AliEn service submits jobs to a local batch scheduler on a cluster.

5.2 Facilities

5.2.1 Jlab

5.2.2 BNL RCF/ACF

Gabriele Carcassi

Gabriele Carcassi has participated in the OSG Integration activity, providing support for GUMS and other components in the Privilege infrastructure. In particular, he collaborated with the VDT team to package and distribute GUMS, and worked with Operations for the requirements and default configuration of the privilege infrastructure.

5.2.3 FNAL

The Fermilab central services established VOs for all the local user communities and now host a central VO Management Service (VOMS) and Site Account Mapping Service (GUMS) for those VOs to use the local Fermi intra-grid FermiGrid as well as OSG. The Central Mass Storage systems are being readied for presentation to OSG in the next quarter.

5.2.4 NERSC/PDSF

A gatekeeper for PDSF was deployed for the OSG production grid and is now functioning as seen in the OSG monitoring services. A gatekeeper for the OSG-ITB is planned and should be operating in the next quarter. The access policies at NERSC at present do not allow grid access to all members of all VOs participating in OSG. At present the procedures to enable grid access even for those VO members that can be allowed access are somewhat problematic and it will be a work item for a while to make the procedures authorizing new users more efficient.

5.2.5 SLAC

Matteo Melani

During the months from April to June more than 80% of my time was dedicated to finishing the requirements for the OSG accounting system and completing a preliminary design of the system. The work was presented at the OSG Consortium Meeting in Milwaukee. The complete list of requirements can be found in the OSG document 130: "OSG Accounting Service Requirements". A document that describes the architecture is in preparation at the time of this writing.

The remaining 20% of my time was dedicated to help the operations of the OSG here at SLAC.

5.2.6 Collaboration with IEPM, Network Performance Monitoring

5.2.6.1 Bandwidth/Throughput Monitoring

The [DataGrid Wide Area Network Monitoring Infrastructure](#) (DWMI) now has [IEPM-BW](#) monitoring successfully installed, making measurements, collecting, analyzing and reporting results at: [BNL](#), [Caltech](#), [CERN](#), [FNAL](#), and [SLAC](#).

The measurements include traceroutes and pings at 10 minute intervals, and capacity/throughput, plus we have just added support for pathload. Analysis, and visualization includes time series, the [traceroute visualization](#), and a beta version of the bandwidth change analysis. Based on our own observations/studies and [tests made at CAIDA](#) and reported at PAM2005, we are replacing the [ABwE/abing](#) lightweight packet pair dispersion bandwidth estimation with [pathchirp](#). This should provide more accurate results at a cost of more network utilization (factor of 10) and time to make a measurement (10 seconds vs. one second). Pathload requires another factor of 100 (above pathchirp) and provides even more accuracy. For monitoring between the monitoring sites we are also using the more intensive [iperf](#) and have introduced support for [thruRay](#) a new tool from Stanislav Shalunov of Internet2. Both iperf and thruRay measure achievable throughput. Iperf provides multi-stream measurements. ThruRay provides the [Round Trip Time](#) and also appears to be easier to control than iperf. [Initial comparisons](#) indicate that the two methods give similar results.

We worked with Bill Allcock of ANL to try and [integrate GridFTP into IEPM-BW](#). The main challenge was with the certificates and running the tests unattended (without requiring manual renewing of the certificate). For the minute we have dropped this activity, and will use bbcp or bbftp if we need file transfer rates to be measured.

The BNL IEPM-BW monitoring host has been upgraded to a more powerful (3GHz) cpu with a 1Gbits/s interface. We installed the latest version of IEPM-BW, configuring this host, getting the appropriate ports enabled, optimizing the TCP window sizes, and getting the measurements to run stably. We then beta tested the ESnet On-demand Secure Circuits and Advance Reservation System ([OSCARs](#)) to reserve circuits with dedicated bandwidth (between SLAC and CERN, and SLAC and BNL) and investigated [the impact on achievable throughput and jitter](#).

For SC2005 and iGrid2005 demos, we have added [maps of the IEPM-BW deployment](#). We are also working with Iosif Legrand of CERN/Caltech to decide how to provide [MonALISA](#) access to the data via a graphical web interface. With the new version of IEPM-BW (version 3), the web services access to the data no longer works. We will be reviving this to enable application access to the data.

5.2.6.2 Passive Monitoring

To potentially reduce traffic on the network, provide monitoring at 10Gbps, a realm where packet pair dispersion techniques are likely to fail due to Network Interface Card offloading function offloading, and also the lack of sufficient granularity for the Unix system, we are looking at collecting Netflow records for big flows (> 100Kbits/s) and seeing if we can use these for forecasting between sites. This required getting approval from the SLAC security folks to make filtered Netflow records available to NIIT, mapping the data from multiple weeks onto a single week for a given site pair, followed by regularizing and interpolating the data, followed by a forecasting technique such as [Holt-Winters](#). Initial results look promising. There appear to be about 15 sites seen from SLAC where there are sufficient Netflow flows to be able to make forecasts even in the presence of seasonal variations. Further the Netflow throughputs are strongly correlated with the packet pair available bandwidth measurements.

5.2.6.3 PingER and Developing Region Monitoring

Two graduate students from NIIT are now at SLAC for a year as part of the MAGGIE-NS collaboration working on the IEPM project and PingER.

We put together and made a presentations for the International Workshop on HEP Networking (in Korea) and for the Internet2 Members Meeting (in Washington), on [Quantifying the Digital Divide from Within and Without](#) including new measurements made from the new monitoring sites in Pakistan and India. We added new sites in Uganda, and are working to get a monitoring site in South Africa (the first on the African continent).

5.2.6.4 Problem Case Reports

We put together two [case reports](#) in this period, both for the Budker Institute of Nuclear Physics at Novosibirsk. Both had to do with major routing changes. The [first](#) resulted in loss of connectivity due to security configurations, the [latter](#) used [GLORIAD](#) to increase the achievable throughput from 500kbits/s to 7Mbits/s and reduce the RTT from 350ms to 250ms.

5.2.6.5 Testbeds

The 10Gbps wide area network testbed at Sunnyvale is still in place with a connection to UltraLight, and an imminent connection to UltraScienceNet. Due to lack of fundings its future is uncertain. We have been assisting UltraScienceNet with their connection at Sunnyvale.

ESnet has recently connected SLAC at 10Gbits/s to the BAMAN. We assisted ESnet to demonstrate utilization of the BAMAN between SLAC and NERSC for Ray Orbach's visit to LBNL on June 24th, 2005.

With Caltech, Manchester, FNAL and others we plan to participate at iGrid2005 (in San Diego) and SC2005 (in Seattle). We have put together a [web site](#) to publicize our efforts.

We gave an invited talk at the Sun Microsystems SuperG meeting in Washington on [Characterization and Evaluation of TCP and UDP-based Transport on Real Networks](#).

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

We hosted: a one day site visit by our DoE/MICS project manager, Thomas Ndousse; a one week visit by the MAGGIE-NS PI from Pakistan, Prof. Dr. Arshad Ali; a visit by Kars Ohrenberg of DESY.

With FNAL we submitted a paper to the 14th IEEE Workshop on Local and Metropolitan Area Networks, entitled "Anomalous Event Detection for Internet End-to-end Performance Measurements".

With Caltech and others we had published: [FAST TCP: From Theory to Experiments](#), Cheng Jin et. al. IEEE Network, Jan/Feb 2005, vol 19, No.1, also SLAC-PUB-10869.

We made the following presentations:

- [Forecasting Network Perfomance](#) presented by Les Cottrell at the Grid Performance Workshop, Edinburgh June 2005.
- [Diagnostic Steps](#) presented by Les Cottrell at the Networking for Non-Networkers second Workshop in Edinburgh, June 2004.
- [Quantifying the Digital Divide from Within and Without](#) presented by Les Cottrell at the International ICFA Workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, Daegu, Korea, May 23-27, 2005.
- [Internet Monitoring and Tools](#) presented by Les Cottrell at the International ICFA Workshop on HEP Networking, Grid and Digital Divide Issues for Global e-Science, Daegu, Korea, May 23-27, 2005.
- Presentations for site visit of Thomas Ndousse IEPM funding manager , DoE, April 26, 2005:
 - [End-to-end Anomalous Event Detection in Production Networks](#),
 - [Internet End-to-end Monitoring Project at SLAC](#),
 - [Characterization and Evaluation of TCP and UDP-based Transport on Real Networks](#),
 - [Testbeds](#).
- [Characterization and Evaluation of TCP and UDP-based Transport on Real Networks](#) presentation by Les Cottrell at the Sun SuperG meeting, Washington DC, April 20, 2005.

We submitted the following proposals:

- *eMAP: end-to-end Measurement Architecture for PERN* submitted with NIIT to USAID.
- *Socrates* submitted with GATech, U. Virginia, and Colorado State, to the NSF DDDAS solicitation.

We are also working on a proposal to Cisco for network topology discovery.

5.3 Computer Science & Middleware

5.3.1 Condor

Work continued on the implementation and deployment of Condor-C for the EGEE project. Condor-C allows jobs in one machine's job queue to be moved to another machine's job queue. These machines may be far removed from each other, providing powerful grid computation mechanisms, while requiring only Condor software and its configuration. Condor-C is highly resistant to network disconnections and machine failures on both the submission and remote sides. An expected usage sets up Personal Condor on a laptop, submits some jobs that are sent to a Condor pool, waits until the jobs are staged on the pool, then turns off the laptop. When the laptop reconnects at a later time, any results can be pulled back. Condor-C scales gracefully when compared with Condor's flocking mechanism. The machine upon which jobs are submitted maintains a single process and network connection to a remote machine, without regard to the number of jobs queued or running. Submitting jobs to run under the grid universe has not yet been perfected. GSI and KERBEROS authentication mechanisms are under test.

The INFN VOMS, JClarens, Privilege Project GUMS and PRIMA modules, updated Generic Information Providers (GIP) and the gLITE CE Monitor were added to the Virtual Data Toolkit (VDT).

5.3.2 Globus – ANL

5.3.2.1 Coordination and Support

Primary coordination vehicle is via the Open Science Grid blueprint activity. Stu Martin has taken the lead on the OSG / TeraGrid Interoperability effort and will be leading sessions at the upcoming face-to-face meeting in Milwaukee.

Support through the discuss lists and Bugzilla are available to all experiments.

5.3.2.2 Globus Toolkit 4.0 released

GT4.0 was released April 29th and is the current supported stable release of the Globus Toolkit. It contains the complete pre-Web Services (ala GT2.x) code base as well as the Web Services Resource Framework (WSRF) version of the services. Note that GridFTP is a 100% new implementation, but is 100% wire protocol compatible. These new WSRF based services represent dramatic improvements in robustness, stability, and performance. 1.1.7 Training, Presentations and Papers

“GT4 GRAM” presented by Stu Martin at the Open Science Grid Consortium Meeting, Stanford, CA

5.3.2.3 Security

The primary thrust of our security efforts have revolved around flexibility with regard to the authorization mechanism and standardization of interfaces to those mechanisms. The GridShib effort is progressing. Shib 1.3 should be released soon, and a beta of GridShib will be released as soon as possible thereafter. The OGSA-authz documents have all passed last call and moving to the GGF editor for public comment. Finally, two Globus activities (but not ANL) that are of interest to the PPDG community include GUMS integration work going on at ISI and the MyProxy 2.0 release with support for PAMS (pluggable authorization modules).

5.3.2.4 Monitoring and MDS work

MDS4 was release as part of the GT 4.0 release in April. Basic functionality includes an Index Service, which functions much as a GT2 GRIS or GIIS as a central aggregation point or registry that caches the last value; a Trigger service, which can monitor service data and perform simple actions when a user-defined threshold is reached; basic information providers to interface to cluster systems (Ganglia, Hawkeye), queuing systems (PBS, LSF), as well as WSRF services (GRAM, RLS, RFT and CAS), and a simple visualization tool, WebMDS. In addition, several security patches were made available to resolve bug 3367.

The MDS4 scalability analysis work continued, and a paper was submitted to SuperComputing 2005. The most noted result in the past quarter was that an MDS4 index server stayed up over 80 days, serving over 500 Million requests, and still running.

Discussions have begun with individual groups on transitioning from MDS2 to MDS4, and we are beginning conversation with OSG about use of MDS4 as a common interface tool for the various OSG data sources.

5.3.2.5 Data Management

The data management components in Globus that are of interest to PPDG are GridFTP, the Reliable File Transfer (RFT) Service, and the Replica Location Service (RLS). Work over the last quarter for all components has focused on completing the debugging, hardening, and documentation of GT4, release activities, and intensive support of any issues arising from the new WSRF implementation, porting, configuration, etc..

The new GridFTP server deployment plan for OSG is in place, test deployments have begun and initial testing indicates no compatibility issues and marked stability improvements. The Reliable File Transfer (RFT) service has been deployed on the TeraGrid for testing and integration. On the Open Science Grid, RFT is used for staging by GRAM and this will drive its deployment there.

One concern within the data management area is that of inter-operability. We still have the long standing issue of the CERN implementation of a replica system and the Globus Replica Location System (RLS) being incompatible. We are also in danger of having a similar issue with the Reliable File Transfer Service and the CERN File Transfer Service (FTS). The latter could be indicative of a significant and broad problem. RFT is WSRF based and the CERN FTS is a pure web service. This means that RFT exposes its internal state via resource properties, notifications, and the GetResourceProperties methods. Pure web services have no standard way of exposing state and so the FTS exposes several non-standard and service unique methods for accessing that state. In general, services following the WSRF family of specifications will not be inter-operable with services that do not.

5.3.2.6 Execution Management / Edge Services

GRAM represents the most dramatic change in GT4. The new WSRF implementation resolves the most severe issues with the pre-WS GRAM implementation, particularly the load on the head node. This is accomplished by switching to a model where a single process monitors the scheduler logs for status on jobs rather than having a monitor process for every job.

Stu Martin has been working with Mike Wilde and other OSG members to develop the GRAM service readiness plan. It has been submitted for review and is awaiting comment. Discussions have begun with the VDT team for adding GT4 GRAM to VDT. For testing prior to GRAMs addition to VDT, a GT4 web services container has been set up at Purdue and Condor-g will be used to submit CMS jobs to it for early feedback.

Stu Martin, Kate Keahey, and Bill Allcock have also been carrying on extensive discussions with members of OSG, CERN, the TeraGrid, and other communities regarding how Edge Services (services dynamically deployed by the VO on edge nodes to start up required VO services) can be deployed quickly, easily, but in a controlled manner so that a mis-behaving edge service can not render the head node, and potentially the entire resource it virtualizes useless. Kate Keahey also continues to work on the development of virtual machine based "workspaces" to further extend the edge services concept.

5.3.3 SRM –

Participants: Alex Sim, Junmin Gu, Viji Natarajan, Arie Shoshani, Kurt Stockinger

DRM and HRM enhancements

- Ensured that DRM works with GT2 and GT3 proxy formats
 - DRMs failed when the newer GT3 proxy was provided. We identified the problem, and made sure that they can work with either version by installing a patch from Globus for GSI.
- A new SRM client was developed

Since we completed DRM v2.1, we needed a client for it. We have developed a client that support the 2.1 functionality including space reservation and directory functions in addition to srmGet, srmPut, and srmCopy. This client works with srm v1.1 as well. A flag is used to select which version is desired.

Support activities

- Fixing existing clients exit codes

Client programs including srm-copy, srm-get, srm-ls, srm-ping, have been enhanced to return error codes in order to the calling programs to work more efficiently.

- client configuration options are now more compact; specifically, the following 3 options are now available

```
client*ClientLogFile=log_path
```

```
client*EnableLogging=true
```

```
client*SRMName=DRM_obj_name
```

New Features

- Support for Posix I/O

One of OSG requirements is to open, close, and read files that were put into DRM cache. However, immediately after it is written it should have the write permission removed and only read allowed. The DRM was enhanced to include this feature to ensure the integrity of files written into DRMs (as well as HRMs). This capability was added to the DRM component that supports prepareToPut.

- Avoiding copying local files when archiving to HPSS

This capability was added to HRM as a requirement from a Wayne State user (Elisabeth Atems) from the STAR experiment. Given that the file is already in user space, and prepareToPut is invoked to archive it to HPSS, the HRM previously copied the file into its space first. This is unnecessary and inefficient for large files. A change was made so that when the user specifies file:// as source, the HRM skips copying the file into cache and will go ahead and archive directly.

OSG support

We provided frequent support to Jorge Rodriguez (U of Florida), who has continued to test the installation and running requests with DRM, as well as using the SRM-tester we have provided. Also, we started a new OSG analysis scenario activity with the STAR project. In this scenario, files are requested through DRM at NERSC to get files from BNL's HRM; the files are analyzed at NERSC, and the output files go to the DRM cache, and then "pushed" to BNL's HPSS through HRM. We are collaborating with Eric Hjort on this activity.

Publication

A paper on Replica Registration Service (RRS) was accepted to the VLDB Workshop on Data Management in Grids (http://liris.cnrs.fr/~jpierson/DMG_VLDB05/). The paper is entitled: RRS: Replica Registration Service for Data Grids, by Arie Shoshani, Alex Sim, Kurt Stockinger.

5.3.4 Caltech

5.3.4.1 Clarens/JClarens

Conrad Steenberg:

A major release, numbered 0.7.0 was made during this time that included the improved browser support and updated web interfaces described in the previous quarter.

An updated version of the Clarens manual was posted on the project web pages that describes the new server version, including a new chapter to aid developers in writing browser-based interfaces.

During the past quarter the majority of my time was spent on development and support of the Clarens web services framework. Major developments include: - The implementation of a session garbage collector to ensure that server session databases do not grow indefinitely. Part of the benefit of this development is a major reduction in the session establishment latency experienced by clients that connect to the server multiple tens of times per second using the same certificate, as is done by the JobMon implementation. - Updates to the shell command execution service requested by the CMS Monte Carlo Processing Service (MCPS) developers, to allow for more flexible server-side sandbox management.

- The addition of the ability to retrieve large numbers of small files as a tar archive using a single request, greatly improving the download speed of data for the CMS experiment where this situation is common.
- More efficient publication of service discovery information to the MonALISA monitoring system.

On the client side, new versions of the stand-alone Python language version was released. Full programming API documentation for this client was written and placed on the Clarens web pages.

Michael Thomas:

JClarens is a grid service framework that provides a platform for developing and deploying Web Services in a grid environment. JClarens includes a core set of services in the base distribution. A handful of 'add-on' services that provide additional functionality are also available.

With help from Tahir Azim, improvements have been made to the JClarens core to further simplify the developer experience. JClarens version 0.6.0, containing these improvements, was released on May 25. Improvements include:

- * A general-use xmlrpc binding class provides a generic xmlrpc adapter for almost all add-on services. It is no longer necessary to write specific xmlrpc adapters for each new service.
- * The server's logging configuration can be modified while the server is running. Changes take effect within 5 seconds of changes being made without the need to restart the server.
- * Refactored authentication code to make it easier to plug in new authentication systems that implement a specific authentication interface.
- * Upgrade to the latest version of the hsqldb database. This includes many bug fixes and improvements over the older hsqldb database that was part of JClarens.
- * An improved configuration API makes it easier for add-on services to read configuration settings from disk. This improved configuration API allows type-specific configuration properties, such as integers, floating point numbers, and lists.
- * File upload support added to the file service. This can be optionally disabled for security-paranoid sites.
- * A new JSON-RPC encoding was included to add support for more browsers in the Clarens browser-based user interface. In particular, Mac users can now interact with the browser-based interface, a feat that was not possible with the previous browser interface based on xmlrpc. browser interface. I became a participant in the jsonrpc community to provide feedback and get support with this work.

Files for the new release are available at SourceForge:

https://sourceforge.net/project/showfiles.php?group_id=53073&package_id=114981&release_id=329971

The Discovery Service is one of the add-on services that are available for JClarens. The Discovery Service is a Web Service registry, similar to UDDI, but with a more dynamic and distributed approach that is much better suited for a Grid environment. In early 2005 the Discovery Service became an optional component of the OSG software stack. A set of scripts are now available that make it possible to publish to and query from the discovery service from the command line. Modifications to the original scripts were made that made it possible to publish the existence of OSG voms-admin services to the Discovery Service. Additionally, the SRM/DRM development team added Discovery Service client components to the upcoming DRM release so that DRM services are published to the Discovery Service.

Two days of meetings were held with Iosif Legrand to discuss improvements to the way that the Discovery Service makes use of MonALISA. Work was started to implement the agreed-upon changes. This work will continue into the next quarter. Notes from these meetings can be found at: <http://ultralight.caltech.edu/gaeweb/wiki/index.php/ClarensMonalisaMeetings>.

5.3.4.2 MonALISA

Iosif Legrand:

We developed new VO modules for Jobs and Grid FTP to provide much better and detailed monitoring and accounting information. The new modules are deployed on OSG-ITB and we are testing and validating them.

The VO_Jobs module provides interfaces to work with Condor, SGE, LSF and PBS. It collects information for all the individual jobs running in the system and the resources used by individual jobs. For each type of batch queuing system a dedicated module is used to extract specific information. Based on the mapping of users in different VOs, it generates aggregate information per VO. The accounting specific information (CPU consumption, used memory, IO performed) is reported as rates and in this way allows generating accounting reports on any time interval.

The new VO_IO module collects information from Grid FTP log files. It is significantly more CPU effective (perform up to 10 times faster than the old modules). It collects transfer rates and also provides details for transfers between all the sites involved in OSG. Accounting information for global network traffic is generated from the information collected by this module.

The VO storage module developed by Matteo Melani from SLAC was interfaced and included in the MonALISA distribution.

A complete new version for the MonALISA repository was developed for OSG. We installed and configure it at: <http://grid02.uits.indiana.edu:8080/>. The same new version is deployed and use for the OSG-ITB: <http://gocmon.uits.iupui.edu:8888/>

The new repository provides a new map view that allows to present different functions (load, job, jobs per VO) on any region. It is based on the MonALISA interactive client and allows generating detailed maps for any region with arbitrary resolution. Currently we use more than 2GB of maps to cover the world. The repository engine can generate very efficiently any view and serve many clients concurrently. The new repository provides a set of new global views to present activities and resources per VO or per site.

A working repository for OSG is installed at <http://pccil.cern.ch:8889/> to developed dedicated accounting panels for Jobs and data transfer activates specific for OSG. They are based on the new VO modules currently deployed on OSG-ITB sites.

In OSG distribution we included the WAN topology agents. These dedicated agents are used to construct in real-time the WAN topology graph. From each site a tracepath like measurement (using a multithreaded requests with different TTL) are periodically performed to all the other sites in a group or virtual organization. The topology graph is constructed from all these measurements, identifying all the routers, the network IDs and corresponding AS. The system evaluates the mean latency on each on each segment and can easily detect asymmetric routing. Agents are dynamically configured from servlets that use the discovery mechanism in MonALISA and they also detect the IP of the requesting agent.

5.3.5 SRB

SRB 3.3.1 was released on April 6. This release includes the interfaces between GridFTP and SRB developed last quarter, support for GT3 GSI (in addition to the long-existing SRB support of previous versions of GSI), an interface from MDSPlus to SRB, and many other new features and bug fixes. Release notes are available at <http://www.sdsc.edu/srb/CurrentSRB/ReleaseNotes3.3.1.html> and <http://www.sdsc.edu/srb/CurrentSRB/ReleaseNotes3.3.html>

A major current effort is consulting support for users of SRB data grids, testing of the software, and development of bug fixes.

The SRB data grid is being used to support data management for the BaBar experiment, and the transport and replication of data between SLAC and Lyon, France. An independent data grid has been installed at each site. This means each site manages a separate metadata catalog and separate name spaces for storage resources, users, files, and metadata. Federation is the process of cross-registration of selected name spaces to support sharing of any combination of storage resources, files, users, and metadata.

W. Kroger established a zone at SLAC and Jean-Yves Neif established a zone at CCIN2P3. Tests are underway to demonstrate that the required access permissions and remote operations work correctly between the two data grids for a data grid federation.

In the SLAC zone there are two storage resources:

SlacHPSS: "HPSS_NOCHK file system"

bbr-xfer02: "unix file system"

The "unix file system" consists of a logical resource SlacXFER which contains the three physical resources bbr-xfer01, bbr-xfer02 and bbr-xfer03. This allows SLAC to do some load balancing. Files are registered into SlacHPSS and the permissions are set so that the group bfactory@bbslac (in the slac zone) has 'all' permissions to these files (bbrlyon@babar, bbusr@bbslac are members of bfactory@groups). Note that "babar" denotes the data grid in Lyon, and "bbslac" denotes the data grid in SLAC. Thus users from two data grids are being authorized to access files in the SlacHPSS storage resource.

When one issues a SRB list command (Sls -C), one can list the access controls of the specified collection (InclK0s_6060.01.root).

```
> Sls -C InclK0s_6060.01.root
```

```
InclK0s_6060.01.root  bbdatt  all  bbdatt@bbslac
InclK0s_6060.01.root  bbdatt  all  bbrlyon@babar
InclK0s_6060.01.root  bbdatt  all  bbusr@bbslac
InclK0s_6060.01.root  bbdatt  all  bfactory@groups
```

The SRB command Ssfstage is used to stage data from an archive to a disk system.

Then the command:

```
> Ssfstage -r -S bbr-xfer02 /slacBabar/store/PRskims/R18/18.0.4/InclK0s/60/InclK0s_6060.01.root \
  /ddist1/srbVault/slacBabar/InclK0s_6060.01.root
```

can be used to copy the file from SlacHPSS to a local disk on bbr-xfer02. The Ssfstage command is now working except for the '-r' option. So far CCIN2P3 is using a slac-zone user to do the staging but as soon as the registration problem is understood and fixed, BaBar will switch to a remote bbrlyon@babar user.

6 Appendix

6.1 Milestones

The complete milestone list for PPDG is not updated for this report. However, the major milestone of deploying the OSG connection for the DOE Lab facilities was achieved in time for the OSG grand opening on July 20, 2005, for BNL, FNAL, NERSC/PDSF and SLAC.

Target Date	Completed	Responsible Task	Milestone Deliverable

Target Date	Completed	Responsible Task	Milestone Deliverable
May-04	√	PPDG-CMS	Enable clients to access the Pool metadata catalog via the Clarens web service environment
Jun-04	√	PPDG-SAMGrid	Test JIM job scheduling at CDF
Jun-04	√	PPDG-SAMGrid	JIM job scheduling on remote resources for next round of D0 reprocessing
Jun-04	√	PPDG-SRM	SRM Specification V3
Jun-04	√	PPDG-BaBAR	SRB based production system for data distribution
Jun-04	√	PPDG-CS-11	Implementation of the CS11 Dataset Query Service by at least Caltech and SLAC
Jun-04	[1]	PPDG-CS-11	Complete specification of 3 more of the standard interfaces for Analysis Services
Jul-04	√	PPDG-OSG-Authorization	Specification and Requirements for Authorization Version 1
Jul-04	[2]	PPDG-OSG-Monitoring	Specify Requirements V1
Jul-04	√	PPDG-STAR	Grid-based Monte-Carlo production start to migrate from centralized Tier-0 center to a Tier-1 distributed model.
Jul-04	[1]	PPDG-STAR	Batch oriented user analysis will start.
Sep-04	partial	PPDG-CMS	Deployment of second version of lookup service for dynamic discovery. Multiple instances of service will be deployed.
Sep-04	[1]	PPDG-OSG-Troubleshooting	Specify Baseline Capabilities
Oct-04	√	PPDG-OSG-Accounting	Specification and Requirements for Accounting Version 1
Oct-04	[1]	Oct-04 PPDG-OSG-WMGT	Process Wrapper (ProRap) for remote execution of HENP applications V1
Nov-04	√	PPDG-SRM	SRM Implementation V2 Interfaces
Dec-04	[3]	PPDG-OSG-Accounting	Implementation of Accounting V1.0
Dec-04	[4]	PPDG-OSG-Monitoring	Monitoring of the system infrastructure across all the Laboratory facilities
Dec-04		PPDG-ATLAS	Deploy prototype sufficient to assess POOL-based ATLAS applications
Dec-04	√	PPDG-JLAB	A prototype implementation of SRM, based on specification 2.x,
Dec-04	√	PPDG-OSG-RRS	Replica Registration Service Specification V1
Dec-04	Jul 05 – Condor-C deployed on EGEE	PPDG-OSG-WMGT	Enhance the scalability of Condor-G (support 10^5 jobs)
Dec-04	√	PPDG-SAMGrid	JIM job scheduling as part of the D0 RAC infrastructure
Dec-04	√	PPDG-SAMGrid	Prototype services which use more general Grid resources, in particular run tests for CDF and D0 on OSG resources
Dec-04		PPDG-CS-11	Multiple Implementations of the three interfaces selected
Dec-04	√Mar-05 Testing	PPDG-DMGT	Use of Enhanced GridFTP
Dec-04		PPDG-OSG-Monitoring	Instrumentation of selected data transport and replication applications
Jan-05	√Mar-05 OSG deployment	PPDG-OSG-Authorization	Implementation V1.0
Jan-05	Partial: >2000	PPDG-OSG	Multi-experiment use of 2000 CPUs, 100TBytes.

Target Date	Completed	Responsible Task	Milestone Deliverable
	CPU's on OSG ITB		July 2005: Achieved as part of the OSG Deployment.
Jan-05		PPDG-BaBAR	Applications that publish in the RLS the SRB MCAT knowledge on the data location.
Apr-05		PPDG-OSG-WMGT	Process Wrapper (ProRap) for remote execution of HENP applications V2
May-05	√ Partial:	PPDG-OSG-Monitoring	Monitoring and instrumentation roadmap
May-05		PPDG-CS-11	Complete specification of remaining ~15 interfaces for CS-11

[1] We identify this as a slipped milestone and will review this task during the next quarter.

[2] We identify this work as being done elsewhere in Open Science Grid.

[3] Storage accounting infrastructure added to MonaLisa. In test and considered for deployment on OSG ITB.

[4] Extensions to MonaLisa for monitoring of common grid infrastructure.